

## Assessment of Neck Pain and Its Associated Disorders

### Results of the Bone and Joint Decade 2000–2010 Task Force on Neck Pain and Its Associated Disorders

Margareta Nordin, PT, Dr Med Sc,\*§ Eugene J. Carragee, MD, FACS,¶||  
Sheilah Hogg-Johnson, PhD,\*\*†† Shira Schechter Weiner, PT, PhD (Candidate),†‡  
Eric L. Hurwitz, DC, PhD,‡‡ Paul M. Peloso, MD, MSc, FRCP(C),§§  
Jaime Guzman, MD, MSc, FRCP(C),¶¶||| Gabrielle van der Velde, DC,\*\*\*\*†††††§§§  
Linda J. Carroll, PhD,¶¶¶ Lena W. Holm, Dr Med Sc,||||  
Pierre Côté, DC, PhD,\*\*\*§§§\*\*\*\*††† J. David Cassidy, PhD, Dr Med Sc,\*\*\*§§§††††  
and Scott Haldeman, DC, MD, PhD§§§¶¶¶¶¶

#### Study Design. Best evidence synthesis.

**Objective.** To critically appraise and synthesize the literature on assessment of neck pain.

**Summary of Background Data.** The published literature on assessment of neck pain is large and of variable quality. There have been no prior systematic reviews of this literature.

**Methods.** The Bone and Joint Decade 2000–2010 Task Force on Neck Pain and Its Associated Disorders conducted a critical review of the literature (published 1980–

2006) on assessment tools and screening protocols for traumatic and nontraumatic neck pain.

**Results.** We found 359 articles on assessment of neck pain. After critical review, 95 (35%) were judged scientifically admissible. Screening protocols have high predictive values to detect cervical spine fracture in alert, low-risk patients seeking emergency care after blunt neck trauma. Computerized tomography (CT) scans had better validity (in adults and elderly) than radiographs in assessing high-risk and/or multi-injured blunt trauma neck patients. In the absence of serious pathology, clinical physical examinations are more predictive at excluding than confirming structural lesions causing neurologic compression. One exception is the manual provocation test for cervical radiculopathy, which has high positive predictive value. There was no evidence that specific MRI findings are associated with neck pain, cervicogenic headache, or whiplash exposure. No evidence supports using cervical provocative discography, anesthetic facet, or medial branch blocks in evaluating neck pain. Reliable and valid self-report questionnaires are useful in assessing pain, function, disability, and psychosocial status in individuals with neck pain.

**Conclusion.** The scientific evidence supports screening protocols in emergency care for low-risk patients; and CT-scans for high-risk patients with blunt trauma to the neck. In nonemergency neck pain without radiculopathy, the validity of most commonly used objective tests is lacking. There is support for subjective self-report assessment in monitoring patients' course, response to treatment, and in clinical research.

**Key words:** best evidence synthesis, cervical spine, neck pain, whiplash-associated disorder, assessment, diagnosis. **Spine 2008;33:S101–S122**

From the \*Departments of Orthopaedics and Environmental Medicine and Program of Ergonomics and Biomechanics, School of Medicine and Graduate School of Arts and Science, New York University, NY; †Occupational and Industrial Orthopaedic Center (OIOC), New York University-Hospital for Joint Diseases, New York University Medical Center, New York; ‡Program of Ergonomics and Biomechanics, Graduate School of Arts and Science and School of Medicine, New York University, New York; §Occupational and Industrial Orthopaedic Center (OIOC), New York University Medical Center, NY; ¶Department of Orthopaedic Surgery, Stanford University School of Medicine, Stanford, CA; ||Orthopaedic Spine Center and Spinal Surgery Service, Stanford University Hospital and Clinics, Stanford, CA; \*\*Institute for Work and Health, Toronto, Canada; ††Department of Public Health Sciences, University of Toronto, Canada; ‡‡Department of Public Health Sciences, John A. Burns School of Medicine, University of Hawaii at Mānoa, Honolulu, HI; §§Endocrinology, Analgesia and Inflammation, Merck & Co. Rahway, NJ; ¶¶Department of Medicine, University of British Columbia, BC, Canada; |||Occupational Health and Safety Agency for Healthcare in BC, Canada; \*\*\*\*Departments of Public Health Sciences and Health Policy, Management and Evaluation, University of Toronto, Canada; †††Institute for Work & Health, Toronto, Canada; ‡‡‡University Health Network Rehabilitation Solutions, Toronto Western Hospital, Canada; §§§Division of Health Care and Outcomes Research, Toronto Western Research Institute, Toronto, Canada; ¶¶¶Department of Public Health Sciences, and the Alberta Centre for Injury Control and Research, School of Public Health, University of Alberta, Canada; ||||Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden; \*\*\*\*Institute for Work & Health, Toronto, Canada; ††††Department of Health Policy, Management and Evaluation, University of Toronto, Canada; ‡‡‡‡Centre of Research Expertise in Improved Disability Outcomes (CREIDO), University Health Network Rehabilitation Solutions, Toronto Western Hospital, Toronto, Canada; §§§§Department of Neurology, University of California, Irvine, CA; ¶¶¶¶Department of Epidemiology, School of Public Health, University of California, Los Angeles, CA.

The device(s)/drug(s) is/are FDA-approved or approved by corresponding national agency for this indication.

Corporate/Industry, Foundation, and Professional Organizational funds were received in support of this work. No benefits in any form have been or will be received from a commercial party related directly or indirectly to the subject of this manuscript.

Address correspondence and reprint requests to Margareta Nordin, Dr Sci, OIOC-NYU HJD, 63 Downing Street, New York, NY 10014. E-mail: margareta.nordin@nyu.edu

From the conceptual model presented in Guzman et al,<sup>1</sup> people with neck pain may or may not seek care for their symptoms. For those who do, once they enter the clinical setting, the diagnostic process begins.

Diagnostics is the process of identifying a medical condition or disease by its signs and symptoms from the results of a clinical examination and other evaluative procedures. The conclusion reached through this process is called a diagnosis. Diagnostics may be used to either “rule in” or, to “rule out” a condition, disease, or disorder. The term “diagnostic criteria” designates the combination of findings which allows the clinician to ascertain the diagnosis of the respective disease.

Typically, someone with abnormal symptoms will consult a physician, who will then obtain a history of the patient's illness and examine the individual for signs of disease. The clinician will formulate a hypothesis of likely diagnoses and in many cases will obtain further testing to confirm or clarify the diagnosis, before suggesting definitive treatment.

In modern Western medicine the diagnoses of illness, along with the diagnostic accuracy of individual or combined diagnostic tests, serves as the basis for decisions on treatment strategies, referrals, disability assessments, reimbursement, and more.

This article presents the main results of a systematic review looking at the evidence regarding the validity and utility of diagnostic tests and self-reported disability assessment in people with neck pain. It is hoped that our best evidence synthesis approach will serve to inform clinicians on how best to confirm or refute a diagnosis or confirm a diagnosis. (Note: The literature search and critical review strategy are outlined in detail in Carroll *et al.*<sup>2</sup>)

## ■ Methods

We conducted a systematic search and critical review of the literature using a best evidence synthesis. The search and review strategies are outlined in detail elsewhere.<sup>2</sup> In brief, we systematically searched the electronic library database Medline for literature published from 1980 through 2005 on neck pain and its associated disorders, we systematically checked the reference lists of relevant articles and we updated our search to include key articles from 2006 and early 2007. Details of our electronic search strategy are outlined in Carroll *et al.*<sup>2</sup> and online through Article Plus.

We excluded studies on neck pain that was associated with serious local pathology or systemic disease, such as neck pain from infections, myelopathy, rheumatoid arthritis, and other inflammatory joint diseases, or tumors. We also excluded neck pain from fractures or dislocations, except for diagnostic and assessment studies relating to ruling out fractures and dislocations in neck pain, which were included in the critical review. Screening criteria are outlined in more detail in Carroll *et al.*<sup>2</sup>

### **Type of Studies Needed to Validate Diagnostic Tests**

Three primary features of a diagnostic test are key to understand the accuracy of any test, they are: reliability (or reproducibility), validity (or accuracy), and predictive value in different populations. The validity of a diagnostic test refers to its ability to correctly identify people as diseased (positive for disease or at risk for disease) or nondiseased (negative for disease or not at risk for disease).

**Reliability.** For a test to be valid, it must first be shown to be *reliable*. That is, a test should consistently give the same result when it is repeated on the same person under the same conditions in a set time frame. Differences in results on repetition of a test, even under the same conditions, can arise for several reasons. The commonest are normal biologic variations in the test subject, individual observer inconsistencies (intraobserver variability), differences across observers (interobserver variability) as well as level of experience in applying the test, and differences in the underlying technology of the test equipment.

**Validity.** The validity or accuracy of a diagnostic test is typically demonstrated by comparing it to a "gold standard." A gold standard is a well-accepted and commonly applied method of identifying the disease or clinical entity of interest. There are standard processes and statistics used to understand a diagnostic test. Sensitivity of a test is the proportion of people with the disease who will have a positive test result. Specificity is the proportion of people without the disease who will have a negative test result.<sup>3</sup>

**Predictive Value.** Often, clinicians are more interested in other attributes of the test that is the predictive values. The positive predictive value is the probability that a person has the disease of interest given a positive test result. Similarly, the negative predictive value is the probability that someone with a negative test result does not have the disease. Sensitivity and specificity are generally thought of as properties of the test. Sensitivity and specificity are largely conditional on the disease state. However, positive and negative predictive values are related to both the accuracy of the test (sensitivity and specificity) and the general prevalence of the disease within the population of interest.<sup>4</sup> Although all 4 statistics (sensitivity, specificity, positive, and negative predictive values) are indicators of test accuracy, some of these may be more important in particular clinical contexts. For instance, if trying to rule out a serious underlying cause of disease, it is most important to have a very high sensitivity, to ensure no cases of serious disease are missed. Likewise, a high negative predictive value is essential so clinicians can be assured that once they accept that the disease is not present (because the test result is negative), no harm is caused to the patient as result of this conclusion.

### **Evaluating Diagnostic Studies**

All diagnostic tests undergo a normal scientific evolution to prove their clinical value. Sackett and Haynes<sup>5</sup> proposed a system to classify various developmental stages of a diagnostic test (phase I-IV studies). This scheme can be used to classify studies of diagnostic tests based on what kind of research question is being addressed in the each study. Early studies of novel tests suggest these tests might be useful, but preliminary studies are not *bona fide* proof that these novel tests are valid, useful or should be widely adopted clinically. Clinicians should understand where novel tests are along this evolutionary scientific continuum, to make the best judgments about whether they should adopt these tests in the clinic.

#### **Phase I Studies of Diagnostic Tests**

These studies are designed to answer the following question: *Do test results in affected patients differ from those in normal individuals?* Such studies are typically conducted among patients known to have the disease and a group of individuals definitely known not to have the disease. If a test is found to be very rarely positive in healthy, normal controls with no suggestion of the disease, this is a good first step and the test will need to be investigated in more clinically relevant settings (phase II-IV) (Table 1). This is the most basic assessment of a test's value. An encouraging phase I study cannot confirm diagnostic validity. However, if the test does not pass this first phase of investigation (*i.e.*, many healthy subjects without the disease have positive test results) the test is very unlikely to have further diagnostic value.

#### **Phase II Studies of Diagnostic Tests**

These studies are designed to answer the following question: *Are patients with certain test results more likely to have the target disorder?* Phase II studies compare the range of test results of groups of patients who already have the established

**Table 1. Research Characteristics of a Diagnostic Study Adapted From Sackett and Haynes<sup>5</sup>**

Disease Present/Absent	Disease +/-	Gold Standard Yes/No	Blinding Yes/No	Utility Yes/No
Phase I	+/-	No	No	No
Phase II	+	No	No	No
Phase III	+/-	Yes	Yes	No
Phase IV	+	No	Yes	Yes

\*Phase I-IV.

diagnosis. The fundamental question here is whether certain values of test results are able to predict the presence of the disease than are other values. This testing strategy only includes patients for whom the clinician already has diagnostic certainty and the clinician is performing the test to categorize the range of results seen in this condition (Table 1). As such, phase II diagnostic tests do not confirm validity and require evaluation in phase III and IV designs before they can be recommended for widespread clinical adoption.

### Phase III Studies of Diagnostic Test

These studies are designed to answer the following question: *Do test results distinguish patients with and without the target disorder among those in whom it is clinically sensible to suspect the disorder?* Given promising results in phase I and phase II studies, it is necessary to determine the outcome of the diagnostic test among patients clinically suspected to have the disease (with the signs and symptoms suggesting the disease but where it is unclear whether the patient definitively does or does not have the disease). That is, clinicians rarely order tests when they are certain of the diagnosis. More typically, clinicians order a test in patients where they have some diagnostic uncertainty and want the test result to reduce that uncertainty. Well-conducted phase III studies are necessary to establish diagnostic validity, and are a prerequisite of widespread clinical adoption of a test (Table 1). A diagnostic test may perform well in completely normal subjects (almost always negative) but may be positive in an unacceptable proportion of subjects without the disease who have similar symptoms. In that case, the test, despite good performance in phase I and II studies, would have poor validity when tested in a clinically relevant setting.

There are key features of phase III diagnostic tests that distinguish them from earlier phases. It is important for clinicians to appreciate that phase III testing is a test of the test, not only a test of the clinical population. The first key feature is that the test must be conducted in a clinical study population in which the disease status is uncertain. The second key feature refers to blinding, that is the results of the test must be independently interpreted from a recognized gold standard. Clinicians are used to ordering tests in sequence, with one test result informing the other. For example, if pneumonia is suspected, a physical examination is done looking for cough, sputum, and abnormal breath sounds. If these are positive, then a chest radiograph (gold standard) is done. To *test a test*, using this example, the interpretation of the chest radiograph must be done independent (blinded) of the results on cough, sputum, and abnormal breath sounds and *vice versa*. Results from phase III testing are hypothesis confirming and can form the basis for widespread adoption of a test.

### Phase IV Studies of Diagnostic Test

These studies are designed to answer the following question: *Do patients undergoing a specific diagnostic test fare better in their health outcomes than similar patients who have not been exposed to the test?* This is a study of test utility, *i.e.*, a test may be valid but have no impact on outcomes (*e.g.*, if there is no effective treatment available) or even adversely affect the patient who has the test done (*e.g.*, if particularly morbid tests are commonly applied but treatments are ineffective). This study design requires follow up of cohorts of patients who have used the experimental test in their evaluation and those who have not (Table 1). A diagnostic test with high utility will show much better health outcomes when the test is used compared to when it is not used. By default, this requires that prospective design, random-balanced allocation of test administration and disclosure, standardized protocol, and blinded interpretation, mentioned earlier, are also important.

Phase I and II evaluations when positive for a diagnostic test are promising and require further studies by phase III and IV studies (Table 1). When the phase I and II studies are negative, *i.e.*, the test does not discriminate between healthy and diseased, there is little value to study the test in a more rigorous design. Phase III and IV tests are necessary to recommend for the use in clinical practice. Positive phase III and IV studies are prerequisite before a diagnostic test can be recommended as clinical routine for implementation. Negative phase III and IV studies for a test prove that the test is not useful in clinical practice and should not be implemented.

## Results

Most studies related to diagnosis in this systematic review were phase I, II, or III studies. Of a total of 95 scientifically admissible studies related to diagnosis, there was 1 phase IV study,<sup>6</sup> and 3 systematic reviews; 1 related to whiplash associated disorders (WAD),<sup>7</sup> another to neck pain with radiculopathy, and manual provocation tests,<sup>8</sup> and the last one, related to intersegmental cervical motion in patients with neck pain.<sup>9</sup> There was 1 meta-analysis related to imaging and emergency care of neck pain.<sup>10</sup>

### Results-Diagnostic Tools and Protocols

The hierarchy on judging the scientific evidence of diagnostic test research has been outlined earlier. To place it in the proper clinical context, some clinically related comments are necessary.

The approach to the musculoskeletal system clinical evaluation includes inspection, range of motion, strength, palpation, and additional tests.<sup>11</sup> Following physical examination, laboratory, and radiologic tests often follow in the clinical setting. We will use this sequential approach to present the diagnostic tests uncovered in our systematic review of the literature. To help readers understand how material in this chapter is organized, we present a description of current clinical practice as applied to neck pain.

### The Clinical Approach to Neck Pain

When a clinician sees a patient with neck pain, the first thought will probably be: "is there an underlying sinister cause of this patient's neck pain?" In the emergency care

setting, the serious underlying concern is fracture, dislocation, or other structural injury requiring special care and/or surgical correction. Fracture or instability may also be of concern in a patient without acute traumatic onset, but within association with other conditions, such as cancer, infection, systemic diseases, inflammatory arthritis, and neurologic compromise.

In the context of the Neck Pain Task Force mandate evaluating neck pain without serious underlying structural disease, the next clinical question is “what is the problem and what is this patient’s likely prognosis?” At this point, the clinician may or may not order further diagnostic testing and discuss treatment options with the patient.<sup>12</sup>

In most cases, clinicians will offer fewer tests to patients whose neck pain is low burden and who are at low risk for disability. Factors that influence subsequent decisions on diagnostic testing for neck pain include: demographics, past experience with the health care system, patient’s past therapies, setting (*i.e.*, workplace *vs.* nonworkplace), patient’s compensation status, the nature of the surrounding health care system, and legal systems.

Although physical examination and other tests can inform about prognosis and treatment options, patient-completed questionnaires have an important role in understanding a patient’s current perceived disability and prognosis. As such, patient questionnaires are also reviewed here as a “diagnostic tests” for status and prognosis.

We have divided the literature review into 3 sections (the order of presentation is based on clinical practice as described earlier):

Section 1 includes all scientifically admissible studies on ruling out serious underlying pathology in neck trauma.

Section 2 includes all scientifically admissible studies related to patients seeking nonemergency care for neck pain with or without arm pain (radiculopathy) and/or headache.

Section 3 includes all scientifically admissible studies related to neck pain and self-assessment questionnaires.

Our systematic review on assessment of tests used to diagnose neck pain concludes with a series of Evidence Statements from the Neck Pain Task Force, which summarize these findings and which may be used as a guide to users of this systematic review.

### Section 1

#### Clinical Emergency Assessment

**Screening for Serious Neck Injury in Patients With Blunt Trauma to the Neck.** Twenty-one studies evaluated screening for possible serious cervical spine injury. The case definition for patients in all admissible studies in this section is “patients seeking care in an emergency room for neck pain after blunt trauma to the neck.” Serious injury to the neck includes fracture, dislocation, sublux-

ation, and/or evidence of spinal cord injury. The studies accepted compared diagnostic accuracy in several ways: emergency clinical screening *versus* radiograph; CT-scan *versus* radiograph; 3-view standard radiograph *versus* 5-view radiograph; F/E radiograph *versus* CT scan and finally thin scan tomography *versus* radiograph.

**Alert Low-Risk Patients With Blunt Trauma to the Neck.** Eleven studies showed excellent performance for 2 screening instruments that were studied in large population-based studies (~40,000 individuals) in emergency care for alert low-risk patients with blunt trauma to the neck (Figure 1).

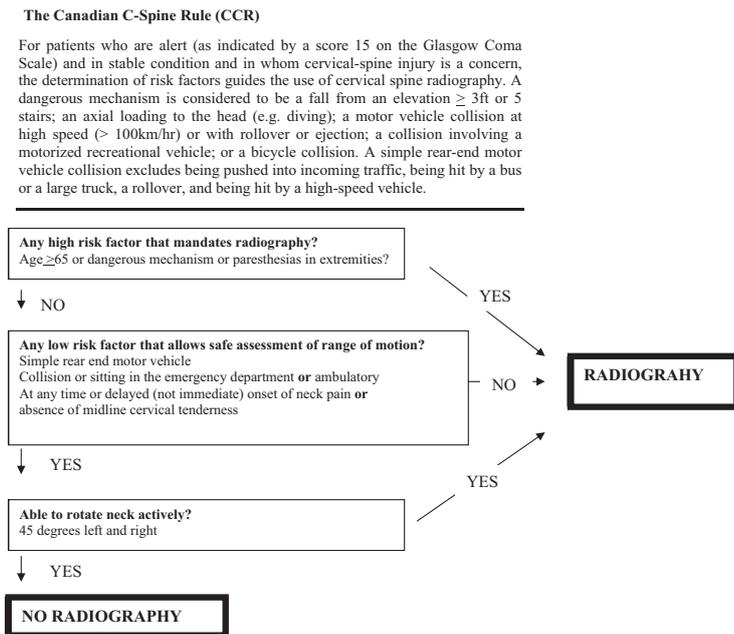
- The Canadian C-Spine Rule (CCR).<sup>13–15</sup>
- The Nexus Low-Risk Criteria (NLC)<sup>16–22</sup>

Tested against a gold standard of radiography (standard 3-view radiograph including lateral, anteroposterior, and open mouth views), both the CCR and the NLC instruments performed well with a high sensitivity and excellent negative predictive value for ruling out serious injury in alert patients with “low risk” neck trauma. Thus, they effectively inform clinicians on optimal test ordering in patients presenting with low-risk neck trauma (Table 2). The NLC is suitable for use in patients over 65 years of age, and it is important to note that there is a relative risk of up to 3 times more fractures in elderly people than younger adults seeking care in the emergency room (Table 2).<sup>23</sup>

**High-Risk Patients With Blunt Trauma to the Neck (Glasgow Coma Scale  $\leq 14$ ).** In high-risk patients with blunt trauma to the neck, CT scan outperforms standard radiograph (3 views), achieving higher predictability and accuracy. Eight studies suggest that CT scan outperformed plain radiograph in patients with cervical trauma and recommended CT scan as first imaging for obtunded, high-risk, and/or multi-injured blunt trauma patients.<sup>10,24–30</sup> These criteria include elements of inspection (alertness, intoxication, and movement), active range of motion (rotation), passive range of motion, palpation (midline tenderness), and additional screening (Glasgow Coma Scale). Two other studies using other criteria for radiography screening for high-risk cervical spine injury were scientifically admissible but had lower accuracy and predictability of serious cervical spine injury in adults.<sup>31,32</sup>

**Screening of Children With Blunt Trauma to the Neck.** No validated screening instrument has been developed for children with blunt trauma to the neck. However, suggested indicators for injury in children with neck trauma are neck pain, altered mental state, and abnormal peripheral neurologic examination (sensation, reflexes, and strength). Risk factors suggestive of significant injury are amount of force, neck tenderness, limitation of neck motion, and major distraction injury.<sup>33,34</sup>

**Other Studies and Blunt Trauma to the Neck.** Flexion-extension radiographs and 5-view radiographs (cross ta-



- The NEXUS Low Risk Criteria (NLC) Algorithm for screening of neck injuries**
- No posterior midline cervical spine tenderness - Midline posterior bony cervical-spine tenderness is present if the patient reports pain on palpation of the posterior midline neck from the nuchal ridge to the prominence of the first thoracic vertebrae, or if the patient evinces pain with direct palpation of any cervical spinous process.
  - No evidence of intoxication - Patients should be considered intoxicated if they either of the following: a recent history provided by the patient or an observer of intoxication or intoxicating ingestion, or evidence of intoxication on physical examination such as an odor of alcohol, slurred speech, ataxia, dysmetria, or other cerebellar findings, or any behavior consistent with intoxication. Patients may also be considered to be intoxicated if tests of bodily secretions are positive for alcohol or drugs that affect level of alertness.
  - A normal level of alertness - An altered level of alertness can include the following: a Glasgow Coma Scale score of 14 or less; disorientation to person, place time, or events; an inability to remember three objects at five minutes; a delayed or inappropriate response to external stimuli; or other findings.
  - No focal neurological deficit - A focal neurological deficit is any focal neurological finding on motor or sensory examination
  - No painful distracting injuries - No precise definition of painful distracting injury is possible. This category includes any condition thought by the clinician to be producing pain sufficient to distract the patient from a second (neck) injury. Such injuries may include, but are not limited to, any long-bone fracture; a visceral injury requiring surgical consultation; a large laceration, degloving injury, or crush injury; large burns; or any other injury causing acute functional impairment. Physicians may also classify any injury as distracting if it is thought to have the potential to impair the patient's ability to appreciate other injuries.

Figure 1. The Canadian C-Spine Rule (CCR<sup>13-15,17</sup>) and the Nexus Low Risk Criteria (NLC) for screening of low risk injuries of blunt trauma to the neck in an emergency setting.<sup>14,16,19,20-22</sup>

ble lateral, anterior-posterior and odontoid views) in the acute stage of blunt neck trauma in adults or children added little to static radiography in predictability and accuracy.<sup>22,25,35</sup> There are additional risks with the performance of flexion and extension radiographs in subjects with uncertain cervical stability after trauma, especially in those with altered mental status, e.g., lack of accurate pain response and muscle stabilization.

Interpretation of standard radiographs of the neck by clinicians and radiologists show high variability in emergency care, but training and experience seem to reduce the variability of interpretation.<sup>36</sup> This is a potentially modifiable source of variability that should be taken into account in the clinical setting.

**Emergency Medical Services Protocol for Spine Immobilization.** One study used a specific protocol to immobilize the cervical spine for transportation of trauma pa-

**Table 2. Performance Criteria of CCR and NLC in Ruling in or Ruling Out Cervical Spine Injuries in Patients With Low Risk Blunt Trauma to the Cervical Spine Seeking Emergency Care**

	CCR* (All Patients)	NLC† (All Patients)	NLC‡ (Patients $\geq$ 65 yr)
Characteristics	Value% (95% CI)	Value% (95% CI)	Value% (95% CI)
Sensitivity	100 (98-100)	99.0 (98.0-99.6)	98.5 (94.8-99.7)
Specificity	42.5 (40-44)	12.9 (12.8-13.0)	14.6 (14.5-14.8)
PPV <sup>2</sup>	2.9 (2.5-3.4)	2.7 (2.6-2.8)	5.3 (5.2-5.3)
NPV <sup>1</sup>	100 (99.9-100)	99.8 (99.6-100)	99.5 (98.3-99.9)

\*Canadian C-Spine Rule, n = 8924.<sup>13-15</sup>  
 †Nexus Low Risk Criteria, n = 34,069.<sup>16-22</sup>  
 ‡Patients  $\geq$ 65 yr Include (n = 2943).<sup>23</sup>  
 NPV indicates negative predictive value; PPV, positive predictive value.

tients (n = 13,483) to the emergency room.<sup>37</sup> Both spine injury assessment by the emergency medical services and spine immobilization were evaluated for predictive values for serious spine injury, i.e., fracture or instability. Spine injury assessment had a sensitivity of 91% (95% CI, 88.3-93.8) and specificity of 40% (95% CI, 39.2-40.9). Spine immobilization had a sensitivity of 92% (95% CI, 89.4-94.6) and specificity of 40% (95%, CI 38.9-40.5). About 8% of injuries to the neck injuries were missed, i.e., 33 of 415 fractures; none of these missed fractures involved spinal cord injury.

**Section 2**

**Nonemergency Clinical Assessment**

**Clinical Evaluation of Patients With Neck Pain With or Without Arm Pain and/or Headache.** This section includes all admissible studies we found related to clinical assessment and diagnostic tools for patients seeking care for neck pain in a nonemergency situation. The case ascertainment in this section includes patients with neck pain, neck pain and headache, and neck pain and radiculopathy at various stages of disease duration (acute, subacute, or chronic). The majority of diagnostic tests reviewed are studies concerning clinical physical evaluation or imaging.

**Patient History.** Since the Québec Task Force published its findings on WAD,<sup>7</sup> no scientifically admissible studies were found evaluating patient history as a diagnostic tool for patients with neck pain. Therefore, the Neck Pain Task Force carefully evaluated existing recommendations for ruling out serious conditions affecting the lumbar spine. We recommend a system of "Red Flags" (similar to the one now used in assessing patients with

**Table 3. Suggested “Red Flags” for Triage of Patients Seeking Nonemergency Care for Neck Pain**

Suggested “Red Flags”	Definition
Trauma	Minor or no trauma but decreased bone loss due to osteoporosis or corticosteroid treatment
Tumor/Cancer/Malignancy	Previous history of cancer, unexplained weight loss, failure to improve with a month of therapy
Spinal cord compromise	Cervical myelopathy (where about half of patients with cervical myelopathy have pain in their neck or arms; most have symptoms of arm, leg or, uncommonly bowel and bladder dysfunction)
Systemic diseases	Ankylosing spondylitis, inflammatory arthritis or other
Infections	Intravenous drug abuse, urinary tract infection or skin infection
Pain	Intractable pain, tenderness over vertebral body
Prior medical history	Previous neck surgery

low back pain), which would allow clinicians to rule out serious pathology in patients seeking care for neck pain with no exposure to blunt trauma (Table 3).<sup>38–41</sup> Important serious diseases to consider include pathologic fractures (following minor trauma or spontaneous), neoplasm (previous history of cancer, unexplained weight loss, constitutional symptoms, failure to improve with a month of therapy), systemic inflammatory diseases (e.g. ankylosing spondylitis and inflammatory arthritis), infections, cervical myelopathy, and/or previous cervical spine or neck surgery or open injury.

**Clinical Assessment for Patients With Neck Pain.** Sixty-three scientifically admissible studies were found. These are presented according to a sequential basic clinical examination: inspection, range of motion, strength, palpation, neurologic examination, and additional tests. Additional tests in this patient population include blood

tests, electro diagnostics, functional tests, tests of symptom amplification, diagnostic anesthetic injections, provocative discography, and imaging studies. Figure 2 indicates the number of scientifically admissible studies in each assessment category.

**Reliability of Clinical Examination of the Neck.** Clinical tests used in a neck examination as a group are not standardized, and their predictive values are quite variable. One study using inspection, range of motion, palpation, and provocation tests on volunteers with and without neck pain and experienced clinicians reported reliability coefficients ranging from inverse poor to moderate ( $\kappa$  coefficient =  $-0.18$  to  $0.52$ ).<sup>42</sup>

**Visual Inspection of Neck and Upper Extremity.** Inter-examiner reliability for visual inspection for abnormal

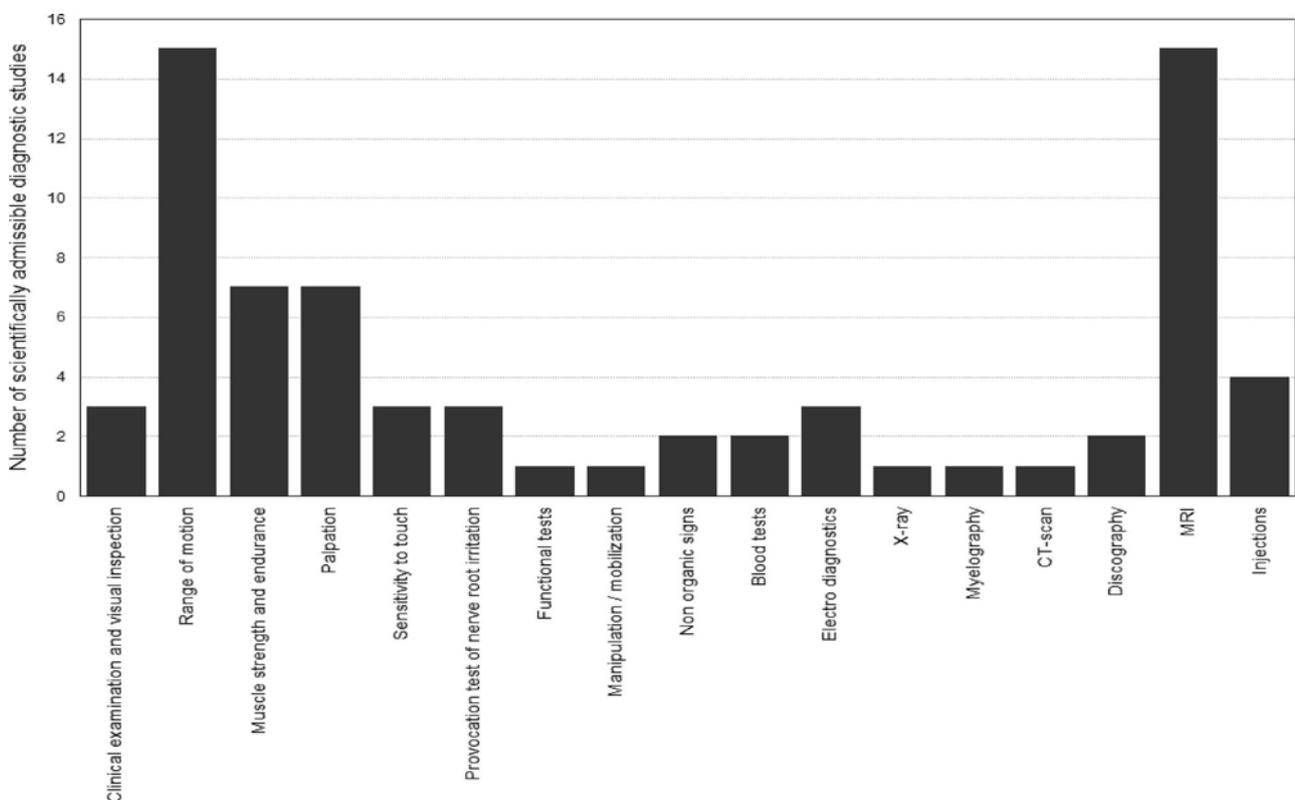


Figure 2. Number of scientifically admissible diagnostic studies in each assessment topic for patients with neck pain with or without radiculopathy and/or headache seeking nonemergency care. A study may have been cited more than once if findings described were valid for more than one topic.

signs (muscle wasting, swelling, tenderness, redness, warmth, scars, nodules, and ganglions) of the neck and upper extremity in patients with neck pain and radiculopathy and nonpatients ranged from fair to excellent, *e.g.*, kappa = 0.32–0.81.<sup>43,44</sup> Interexaminer reliability increased as level of disease prevalence decreased, *i.e.*, agreement and reliability of visual inspection increased to kappa = 0.96–1.00 in healthy controls.<sup>43</sup>

**Range of Motion of the Neck.** Fifteen studies reported range of motion for diagnostic purposes. These studies included intersegmental range of motion of the cervical spine, passive, and active range of motion of the neck, measured with and without devices, in controls and patients with neck pain (with and without radiculopathy).

**Reliability Range of Motion of the Cervical Spine.** Intersegmental cervical spine motion, tested by physical therapists, had slight to moderate inter-rater reliability (Kappa = 0.05–0.61) in 2 small studies in patients with neck pain<sup>45,46</sup> and 1 systematic review.<sup>9</sup> Inter-rater examination reliability for passive cervical range of motion has also been shown as slight to moderate; however, these data should be interpreted with caution as only 2 small studies were found using 2 experienced therapists in each study.<sup>46,47</sup> Active range of motion of the cervical spine can be visually estimated by clinicians or measured with external devices.<sup>44,48–55</sup> Only 1 of the studies (phase III) for active range of motion of the neck used a gold standard (radiograph in asymptomatic subjects) as a comparison.<sup>54</sup> Active range of motion of the neck visually estimated by clinicians was as reliable as using an external device for measuring neck range of motion, with moderate (Kappa  $\leq$  0.60) intrarater and inter-rater reliability. The variations in ratings of motion in the cervical spine were about 10° for intrarater agreements and about 20° for inter-rater agreements, irrespectively of method used.<sup>49</sup> Measurements of protraction and retraction of the head showed less reliability compared to flexion, extension, side bending, or rotation of the head.<sup>48,51</sup>

**Range of Motion in Patients Versus Nonpatients.** Patients with neck pain with or without radiculopathy on average had slightly less volitional motion (phase I and II studies) compared to individuals with no neck pain, but there is a large degree of overlap between groups.<sup>48,50,52,54</sup> Patients reporting acute benign WAD moved more slowly through the range of motion and have decreased volitional range of motion of the neck compared with asymptomatic controls.<sup>48,56</sup>

Chronic WAD patients recruited for examination by an insurance company had significantly lower volitional range of motion in the cervical spine compared to controls.<sup>54</sup> Patients with neck pain and nonpatients were equally accurate in estimating normal range of motion

but less accurate in estimating reduced range of motion of the cervical spine in 2 studies.<sup>57,58</sup> In 1 population study (phase III) in which subjects performed a self-assessment of their neck range of motion, and a clinical examination was used as gold standard (physician assessment following a strict protocol) for comparison, sensitivity ranged from 0.20 to 0.44 and specificity ranged from 0.95 to 0.98.<sup>58</sup>

**Accuracy of Neck Movement Pattern in Women With and Without Exposure to Whiplash Trauma.** A cross-sectional phase I study investigated the reliability and discriminant validity of a new test, “The Fly,” to detect accuracy of neck movement patterns in women.<sup>59</sup> The Fly test is a computerized test in which the subject follows a slow-moving object on the screen. The object has an unpredictable movement path that the subjects follow by moving their heads. The head movements are traced by software. Twenty women, reporting chronic (>6 months pain) pain complaints after exposure to whiplash trauma (WAD Grades I and II), were tested and compared to an aged matched control group. Intraclass correlation measured for reliability and neck movement pattern ranged from 60% to 77% for controls and from 79% to 86% for WAD patients when tested on 2 consecutive days. There was a significant difference ( $P < 0.05$ ) between groups for each study movement pattern traced and tested. The test seems to have construct validity.

**Muscle Strength and Endurance.** We accepted 7 studies (6 phase I and II studies and 1 phase III study) dealing with neck muscle strength. For diagnostic purposes, muscle testing of the neck and upper extremity had consistent slight to moderate interexaminer reliability (kappa  $\leq$  0.60) in patients with neck pain with or without radiculopathy.<sup>43,44,55</sup> Interexaminer reliability increases as level of disease prevalence decreases.<sup>43</sup> The coefficient of variation in patients having repeat testing within days of an index test was about 8%.<sup>50</sup> There is some evidence that patients with chronic neck pain or “myalgia” have slightly lower neck muscle strength compared with control subjects. In subjects with neck pain, self-reported pain and disability ratings showed no correlation with strength measurements.<sup>47,50,60</sup>

One study (phase I) evaluated neck muscle endurance in patients ( $n = 71$ ) with WAD II compared with aged matched healthy controls ( $n = 71$ ).<sup>61</sup> The Gold standard was the Neck Disability Index.<sup>62</sup> Cervical flexor endurance tested in a supine position could distinguish well between WAD II patients and healthy controls ( $P = 0.00$ ). Muscle endurance measurement by EMG for repeated forward flexion of the arm tested in cleaners ( $n = 25$ ) with neck pain and myalgia compared with symptom-free subjects ( $n = 46$ ) was significantly lower.<sup>60</sup>

**Palpation-Trigger Points and Tender Points.** There were 7 studies related to assessment by palpation in patients

with neck pain with and without radiculopathy including patients with whiplash exposure.

**Reliability of Palpation for Tender Points and Trigger Points.** Assessments of trigger points around the neck by clinicians have fair to moderate inter-reliability ( $\kappa = 0.24-0.56$ ) in patients with acute neck pain with or without arm pain or chronic neck pain ( $n = 52$ ).<sup>44</sup> In 1 study (60 chronic neck patients), using an algometer increased inter-reliability for trigger point examination from moderate to excellent.<sup>51</sup>

**Assessment of Trigger Points/Tender Points Against Gold Standard.** When palpation around the neck in patients and nonpatients was tested against a gold standard (pain elicitation on physical examination), the sensitivity and specificity for trigger points was about 80% for both.<sup>63,64</sup> Trigger point distribution was not found to discriminate between subjects with neck pain alone, neck pain with radiculopathy, or neck pain and MRI disc "bulging."<sup>65</sup>

**Patient Self-Assessment of Tender Points.** In one phase I population study, individuals with or without neck pain could determine some presence and good absence of pain with self-palpation of predefined trigger points around the neck compared with a strict protocolized physician examination as gold standard (PPV 0.16–0.39, NPV 0.92–0.96).<sup>58</sup>

**Sensitivity to Touch According to Cervical Dermatomes.** Sensitivity to touch (light touch and pin prick) has been evaluated on patients with neck pain with radiculopathy in 2 clinical studies<sup>44,55</sup> and in 1 population-based study.<sup>58</sup> Inter-rater reliability for sensation was slight to substantial ( $\kappa = 0.16-0.67$ ) with higher reliability for increased sensation compared with decreased sensation in patients with radiculopathy.<sup>44,55</sup> Compared with a gold standard test (physician assessment following a strict protocol), subjects' self-assessment (using a predefined protocol of the ulnar and median nerve dermatomes) demonstrated large variability in sensitivity and high specificity, *i.e.*, the subjects showed large variability to rule in the decreased sensation and high predictability to rule out decreased sensation.<sup>58</sup>

**Provocation Tests for Neck Pain with Radicular Involvement.** There is good evidence from 3 studies<sup>55,64,66</sup> and 1 systematic review<sup>8</sup> that clinical provocation tests for nerve root compression have high predictive values when compared with gold standards (MRI, nerve conduction/EMG, and myelography).<sup>8,55,64,66</sup> Tests including neck contralateral rotation of the head and extension of the arm and fingers extended yielded high accuracy for pain elicitation radiating in the arm associated with cervical root irritation, with sensitivity ranging from substantial to excellent (0.77–0.90) and specificity ranged from fair to excellent (0.22–0.94).<sup>55,64</sup>

**Functional Tests: Lifting, Stepping, and Walking Tests.** There is some evidence from a construct validity study that patients with chronic neck pain and high neck pain intensity during functional testing (lifting, stepping, and walking) have low-test performance.<sup>67</sup>

**Manipulation, Mobilization for Diagnostic Purpose.** There is evidence from 1 double-blind randomized trial (phase IV) that tested the utility of low-amplitude manipulation and endplay assessment of the cervical spine and showed it to be low.<sup>6</sup> Employing endplay assessment in the evaluation did not improve utility, *i.e.*, the primary outcome (same-day relief of pain and stiffness) observed in neck pain patients ( $n = 104$ ) receiving the tests.

**Nonorganic Signs Manual Test.** There is evidence from 2 small studies that clinical testing for nonorganic signs in patients with chronic neck pain with and without radiculopathy had high inter-rater variability ranging from slight to excellent ( $\kappa = 0.08-1.00$ ).<sup>55,68</sup> Inter-rater variability increased somewhat for patients with chronic neck pain without radiculopathy.<sup>68</sup>

**Blood Testing.** Two phase I studies examined routine serology testing in patients with neck pain. One study showed slightly elevated blood markers (MNC and T-cells, CCR5) in patients with WAD 3 days after exposure to whiplash injury compared with blood markers in healthy controls; however, these markers normalized within 2 weeks.<sup>69</sup> No differences were found in routine serology for rheumatic and thyroid diseases administered to 2 groups of subjects: people exposed to repetitive work with high prevalence of chronic neck and shoulder complaints, and a group of controls.<sup>63</sup>

**Electro Diagnostics.** A number of electrodiagnostic tests have been considered or used with the intent of documenting cervical radiculopathy. These include needle electromyography, F-responses, and nerve conduction studies, mixed nerve or dermatomal somatosensory-evoked responses, and quantitative sensory testing. We found no accepted scientific studies that supported the use of these tests in the diagnosis of cervical radiculopathy and no studies that can be used to determine the sensitivity or specificity of these tests in the clinical assessment of radiculopathy and neck pain.

Needle EMG, however, is an established test for the detection of acute and chronic muscle denervation that is often the hallmark of motor radiculopathy. Needle EMG has been used in 3 studies as the gold standard for manual clinical tests that have been studied for the detection of radiculopathy.<sup>55,64,66</sup> It should be noted that reports from the American Academy of Neurology's Therapeutics and Technology Subcommittee concludes that the current evidence is insufficient to determine the appropriateness of dermatomal somatosensory-evoked responses for any condition and that this test should be regarded as investigational (American Academy of Neu-

rology 1997, reaffirmed on November 9, 2006). The same Subcommittee of the American Academy of Neurology assessed the literature on quantitative sensory testing and reached the conclusion that there were no adequate studies to consider quantitative sensory testing useful for any neurologic diagnosis.<sup>70</sup>

Surface electromyography (sEMG) has been proposed as a test to distinguish patients with neck pain from those without neck pain. We accepted 3 studies (phase I), which involved sEMG. One study recorded sEMG of the upper trapezius over the workday in subjects (22 shopping center employees and 44 health care workers) with and without neck pain.<sup>71</sup> The 2 groups of workers represented high (health care workers) *versus* low (shopping center employees) biomechanical work exposure for the neck. There was no difference in average sEMG measurements between those who reported pain and those who did not.

Another study compared subjects ( $n = 20$ ) with persistent neck pain with pain-free matched controls.<sup>72</sup> sEMG activity was continuously recorded from the upper trapezius muscle of the dominant side and the results averaged over 3 days. The gold standard in this study was self-reported persistent pain in the trapezius area, and the disease being diagnosed was persistent cervical muscle pain. The 2 groups did not show a significant difference in average sEMG activity over 3 days despite the difference in reported pain and sensory scores.

In another phase I study purported, "brain stem-mediated antinociceptive reflexes" of the temporalis muscle were measured 5 days after injury in patients ( $n = 82$ ) with benign whiplash injury and acute posttraumatic headache.<sup>73</sup> A control group ( $n = 43$ ) of age and sex-matched volunteers was recruited in this cross-sectional study. sEMG was recorded from the right temporalis muscle. The data were analyzed for duration of ES2 for inhibitory temporalis reflex pattern (duration in ms), and ES1 (latency and duration) of interposed EMG burst. ES2 was significantly reduced in WAD patients from 36.5 (SD 9.4) ms compared with controls 49.0 (SD 7.1) ms ( $<0.001$ ).

The meaning and clinical usefulness of these findings for sEMG remain to be determined. The American Academy of Neurology similarly found that studies of sEMG for low back pain were inconclusive or inadequate, and this test was considered unacceptable as a clinical tool in the evaluation of patients with low back pain.<sup>74</sup> This statement was reaffirmed on November 9, 2006.

**Diagnostic Imaging (Radiograph, Myelography, CT Scan, Discography, and MRI).** Eighteen studies were scientifically admissible concerning imaging in symptomatic and asymptomatic individuals and patients seeking care for neck pain with and without radiculopathy or spinal stenosis in a nonemergency situation. These studies are presented in the following order: radiograph, myelography, CT scan, discography, and MRI.

**Radiograph in Neck Pain.** In 1 cross-sectional study (phase I study), cervical curvature was analyzed in a group of patients ( $n = 488$ ) with WAD who sought care 2 weeks after injury; the test was also administered to a group of healthy controls ( $n = 495$ ).<sup>75</sup> Both groups underwent an anteroposterior and lateral radiograph in sitting position. Lordosis and kyphosis standardized measurements were done from radiograph images. No significant difference between measurements was found between groups for cervical lordosis *versus* nonlordosis or kyphosis. The study did find that age and gender were significantly associated with both increased lordosis and kyphosis, especially among women.

**Myelography in the Assessment of Neck Pain.** One scientifically admissible study found that patients referred for cervical myelography who received water-soluble contrast agent iohexal ( $n = 368$ ) had significantly less morbidity compared with patients given metrizamide ( $n = 90$ ). Both agents produced similar visualization characteristics during imaging.<sup>76</sup>

**Computerized Tomography in the Assessment of Neck Pain.** We accepted 1 study that looked at the use of computerized tomography (CT) scans in patients with neck pain ( $n = 38$ ) with and without radiculopathy including patients with spinal stenosis.<sup>77</sup> Inter-rater reliability readings of cervical spine CT scans by 6 neuroradiologists in patients with spinal stenosis yielded fair to moderate agreement ( $\kappa = 0.26-0.50$ ). When test findings from CT scans and MRI readings were compared, agreements were slight to fair ( $\kappa = 0.15-0.37$ ).<sup>77</sup>

**Discography in the Assessment of Neck Pain.** Provocative cervical discography injections to determine if an injection reproduces a neck-pain patient's usual symptoms are purported to be useful to identify primary cervical discogenic pain illness and to guide treatment. One phase I and 1 phase II diagnostic studies of discography were scientifically admissible.<sup>78,79</sup>

One phase I study tested provocative discography injections (iohexol) in neck/head chronic pain patients and asymptomatic subjects ( $n = 20$ , 10 in each group).<sup>78</sup> This study found 7 of 10 (70%) asymptomatic subjects had a painful response to a disc injection of 4 or 5 (on a 0-10 scale of increasing pain intensity), and 2 subjects had a pain response of 6/10, *i.e.*, all false positive for pain intensity. The patient group, as a whole, reported greater mean pain responses with disc injection. The production of pain with disc injection does not seem to confirm the presence of discogenic pain as the primary cause of a serious neck pain illness in chronic neck/head pain patients compared with asymptomatic subjects. The proportion of healthy asymptomatic subjects, reporting pain values  $\geq 4$  of 10 with cervical disc injection, seems substantially higher in this study of cervical injections than similar experimental injections in the lumbar spine in subjects asymptomatic for low back problems.<sup>80,81</sup>

One descriptive study of patients referred for provocative discography ( $n = 41$ ) showed that injections at each disc level elicited pain in a broad area about the head, neck and chest with considerable overlap between levels.<sup>79</sup> Unfortunately, it was not clear that the cervical disc degeneration *per se* elicits specific axial or other pain symptoms because the study did not use a gold standard (for example surgical removal of the purported pain generator) to confirm primary symptomatic disc disruption.

We found no scientifically admissible phase III or IV diagnostic studies that tested the validity of discography as demonstrating primary discogenic pain in the cervical spine. There have been no gold standard comparisons of cervical discography results with proven discogenic pain by histopathologic or outcomes standards. Given the high proportion (70%) of healthy asymptomatic subjects reporting a painful response to disc injection in the Schellhas study,<sup>78</sup> it is not clear that the underlying premise of the test is valid. Furthermore, there have been no studies that show discography will improve clinical outcomes (phase IV) in patients considering surgery for neck pain symptoms and degenerative changes in the cervical spine.

**Magnetic Resonance Imaging in the Assessment of Neck Pain.** Magnetic resonance imaging (MRI) is frequently recommended for the evaluation of neck pain syndromes. The MRI is often definitively used to determine the presence of serious underlying disease beyond the mandate of the Neck Pain Task Force (*e.g.*, tumor, malignancy, rheumatoid panus extension, syrinx development, and others). When MRI is used in conjunction with physical examination, needle EMG, and provocative tests, MRI is helpful in confirming the site and level of root compression. We did examine the evidence for the use of MRI in patients reporting neck pain (with and without radiculopathy and/or headache) for whom these serious diseases have been excluded. We accepted 15 scientifically admissible studies examined MRI in asymptomatic individuals and in patients with neck pain (with and without radiculopathy or cervical spinal stenosis).

**Reliability of MRI.** The reliability of repeated readings of MRI for common findings for was generally fair to moderate in 6 studies. Interobserver reliability coefficients of determining anterior disc protrusion, disc degeneration, and foraminal stenosis in the cervical spine was moderate ( $kappa = 0.51-0.60$ ) in 1 study.<sup>82</sup> In another study, the rating of severity of cervical stenosis ( $kappa = 0.37$ ) and the determination of the cause of stenosis (bone *vs.* disc or combination,  $kappa = 0.40$ ) were found to have fair reliability.<sup>77</sup> The use of a digitizer compared to a ruler did not seem to improve reliability.<sup>83</sup>

In patients with chronic WAD ( $n = 92$ ) and matched controls ( $n = 30$ ) (phase I studies), examiners looked specifically for MRI bright signal abnormalities in the alar ligaments in the upper cervical spine (special se-

quence studies). It has been suggested that these signal abnormalities might indicate ligament injury.<sup>84</sup> The study found fair to moderate inter-reliability ( $kappa = 0.31-0.57$ ) in identifying signal abnormality of the cervical alar ligament.<sup>84</sup> MRI signal intensity readings were less reliable (slight to fair  $kappa = 0.17-0.39$ ) for the transverse, tectorial, and posterior atlanto-occipital membranes.<sup>85,86</sup>

In summary, the reliability of MRI readings for common degenerative or other pathologic findings in the cervical spine is moderate at best.

**Standard MRI Versus Enhanced MRI.** Standard MRI as gold standard *versus* enhanced MRI was studied in patients ( $n = 61$ ) evaluated for degenerative diseases (cervical discs and or osteophytes) in the cervical spine.<sup>87</sup> Clear advantages for the enhanced MRI were not seen, and the standard MRI appeared to perform better than enhanced MRI in this patient population.

**MRI Versus Surgery as Gold Standard.** One phase III study looked at MRI *versus* surgical observation and palpation (gold standard) in patients ( $n = 54$ ) with cervical disc herniation who were referred to surgical intervention.<sup>88</sup> The aim of the study was to determine the accuracy of MRI in predicting the presence of disc material posterior to the posterior longitudinal ligament (PLL). Surgery confirmed 26 of 54 levels of disc material posterior of the PLL. MRI had a sensitivity of 42% and a specificity of 93% for disc material posterior to the PLL. The generalizability of this study is not clear as the same unblinded surgeon read the presurgical MRI and reported the intraoperative findings.

A retrospective study in 41 patients with neck pain including cervical radiculopathy ( $n = 15$ ) or myelopathy ( $n = 19$ ) compared surgery, as the gold standard, *versus* MRI in determining the presence or absence of a hard disc protruding into the cervical spinal canal.<sup>89</sup> Three independent observers (2 neurosurgeons and 1 neuroradiologist) participated; 1 surgeon did all surgery. Sensitivity (ruling in) ranged between 75% and 96%, specificity (ruling out) from 27% to 60%, PPV 68% to 75%, and NPV from 60% to 80% between independent raters' MRI and surgical findings.

**MRI Findings in Asymptomatic and Symptomatic Individuals With or Without Neck Pain.** Three studies evaluated MRI the cervical spine in asymptomatic individuals (number of subjects in all 3 studies combined  $n = 649$ ).<sup>82,90,91</sup> All studies found that positive MRI changes in the cervical spine were common in asymptomatic subjects and increase significantly with age ( $P$  ranged from 0.05 to 0.0001). Two studies showed a linear relationship between degenerative changes and age.<sup>82,91</sup> Still, in no age-group tested (including subjects <25 years of age) were degenerative findings rare (<5%). Positive findings and number of cervical discs assessed for disc degeneration in asymptomatic individuals are shown in Table 4.

**Table 4. MRI Positive Findings (%) Related to Disc Degeneration in the Cervical Spine and Age and Disc Levels (n) Assessed in Asymptomatic Individuals**

Reference/No. of Cervical Disc Assessed	Age <30 yr	Age <40 yr	Age >40 yr	Age >60 yr
Lehto <i>et al</i> <sup>91</sup> n = 533	14%	25%	57%	
Matsumoto <i>et al</i> <sup>75</sup> n = 2480	M 17%, W 12%			M 86%, W 89%
Boden <i>et al</i> <sup>90</sup> n = 264		8%	37%	
Siivola <i>et al</i> <sup>92</sup> n = 186	<25 yr 5%			

M indicates Men; W, Women.

In asymptomatic subjects, disc degeneration was often accompanied (up to 78%) by other positive MRI findings such as disc bulging, posterior or anterior disc protrusion, narrowing of the disc space, foraminal stenosis, and/or abnormal spinal cord contour.<sup>82,90</sup> The high prevalence of these positive findings in MRI of the cervical spine in asymptomatic individuals emphasizes that common degenerative findings on MRI cannot be assumed to be the primary cause of symptoms in adult patients with neck pain.

One population-based study of young adults (n = 547) was aimed at determining whether subjects with persistent or recurrent neck and shoulder pain were more likely to have abnormal MRI findings of the cervical spine than those without neck and shoulder pain.<sup>92</sup> The first survey was performed on 17-years-old subjects; a follow-up survey was done 7 years later. Those subjects (n = 26) who responded with no symptoms at both surveys comprised the “no symptom” group and those subjects (n = 40) who responded with weekly symptoms in neck and shoulder at both surveys made up the “symptom” group. Thirty-one of the subjects from both groups had an MRI: 15 in the symptom-free group and 16 in the group reporting weekly symptoms over 7 years. A trend was found in the proportion of disc herniation seen on MRI in the symptomatic group ( $P = 0.10$ ); however, the sample size is small, there was some apparent work-up bias (not all identified subjects were scanned). Thus, the findings should be interpreted with caution.

One study (phase I) evaluated MRI of the cervical spine of fighter pilots (n = 12) and age-matched controls (n = 12). The study showed premature cervical disc degeneration among senior fighter pilots exposed frequently to extremely high +Gz forces.<sup>93</sup> Fighter pilots had significantly more degenerative disc findings at C3–C4 (88%) versus controls (36%).

In summary, the evidence reviewed indicates that neck pain without clear radiculopathy is not reasonably ascribed to specific common degenerative changes seen on MRI.

**MRI Findings in Patients With Whiplash Associated Disorders.** Patients (n = 40) with benign WAD were exposed to cervical and cerebral MRI 2 days after injury;

findings were compared with controls (n = 20) not exposed to whiplash trauma.<sup>94</sup> The study failed to demonstrate unique or specific soft tissue lesions by MRI following acute whiplash exposure.

The possible presence of demonstrable ligamentous injury to the upper cervical spine after whiplash exposure has been investigated with special sequence MRI. A phase I study (n = 30) showed that bright signals in the alar, transverse ligaments, and other structures have been observed more frequently in subjects with whiplash trauma exposure after 6 years (range, 2–9 years) than control subjects.<sup>84–86</sup> However, the reliability of different observers in classifying the presence or degree of ligamentous injury, as shown by the MRI signal change showed high variability. Validation of this finding as diagnosing *bona fide*, and clinically relevant ligamentous injury has not been demonstrated in these patients with WAD (grades I–III).

**MRI in Asymptomatic Individuals and Patients With Cervicogenic Headache.** One cross-sectional phase 1 study compared MRI images of the cervical spine in patients with cervicogenic headache (n = 22) to MRIs of asymptomatic controls (n = 20).<sup>95</sup> Findings of bulging cervical disc were found in 45% both in patients and controls. Other common degenerative findings were no more frequently found in patients compared with controls ( $P > 0.05$ ). The study suggests that MRI may not be an adequate method to detect pathologic findings for cervicogenic headache (if such specific pathologic findings exist).

**Injections.** Anesthetic blockade or provocative injections of the cervical facet (zygapophysial) joints have been purported to diagnose neck pain due to primary facet joint pain in the absence of clear serious facet pathology (e.g., fracture, tumor, and isolated arthrosis). Various injections schemes have been advocated including

- injection of saline into the facet joint to “reproduce” the patients usual pain,
- injection of anesthetic agents into the facet,
- injection of anesthetic agents above or below the joint to anesthetize the medial branches (MB) innervating the facet,
- comparison of different anesthetic agents (long vs. short-acting agents) injected a facet joint,
- comparison of anesthetic agents versus placebo (saline) injections to a facet joint.

We accepted 4 studies on injections into the cervical spine for diagnostic purposes<sup>96–99</sup> in patients with chronic neck pain (with or without WAD). Although accepted in our systematic review, these studies were all performed in specialty-clinic settings, and all included significant work-up bias, *i.e.*, either not testing all facet joints or testing all study subjects in the same manner. Three studies found that a high proportion of subjects in

a specialty research clinic responded with some pain relief to an anesthetic injection to the facet joints or medial branch nerves. Pain relief response was found in 71%,<sup>98</sup> 96%,<sup>100</sup> and 97%,<sup>97</sup> respectively, of patients in the accepted studies. Whether any of those subjects actually had primary facet joint pain, as the cause of their illness is unknown, because gold standard comparisons were not made.

However, when single positive injection with 1 anesthetic agent were challenged with a second injection (n = 45) using an anesthetic agent of different expected duration only 51% (95% CI 44%–58%) of these subjects had the expected “appropriate” pain relief with the second injection. That is, using the “comparative” block criteria, after an initial “positive” injection, the joint was found to be “negative” for pain relief about half the time with the second injection.<sup>100</sup> Another study using a different definition of a positive injection, found 73% (95% CI 62%–85%) of subjects responding to the first anesthetic injection with appropriate pain relief.<sup>97</sup>

Furthermore, when subjects (n = 27) who had been found to have pain relief with both short- and long-acting anesthetic blocks were challenged by a placebo (saline) injection, the study found that 26% (95% CI 16%–35%) responded with complete relief to the saline injection.<sup>99</sup> Conversely, of subjects responding with complete relief after a single anesthetic injection and having no response to a saline injection, only 55% had an appropriate response to another anesthetic injection.

In summary, diagnostic facet injections have not been validated to identify facet joint pain as the primary cause of pain in patients with serious neck pain illness. Nor do these injections seem to have acceptable reliability. Anesthetic blockade by a series of alternative methods does not seem to consistently identify the same subjects as having presumed primary facet joint pain. Finally, in no study has the utility of these injections been established based on improved outcomes.

### Section 3

**Self-Assessment Questionnaires.** Self-administered questionnaires are commonly used in clinical practice. These instruments primarily deal with perceived pain, perceived disability, inability, or ability to cope with neck pain, inability, or ability to function and/or healthcare utilization. Although subjective measures may provide useful evaluative information for determining the etiology of the pain, questionnaires alone cannot establish or confirm a diagnosis; however, these questionnaires can:

- provide valuable insight into the impact of neck pain;

- be used to monitor change of the condition over time;
- be very helpful in establishing the patient’s perceived functional deficit and/or psychosomatic status; and
- be useful for choice of appropriate and effective treatment for both clinicians and patients.

To determine the real value of a questionnaire, one should know about its reliability and validity (content, construct, and predictive), its responsiveness to change, the ease of administration, and how acceptable the tool is to both patients and clinicians. Based on this systematic review and scientifically admissible articles, not all criteria are available for any of the questionnaire cited.

There were 19 scientifically admissible studies describing 13 self-administered instruments used for the clinical evaluation of patients with neck pain (with or without radiculopathy or WAD) in a nonemergency situation. Most of the questionnaires have been designed specifically to evaluate patients with neck pain, several questionnaires have been designed to evaluate disorders of the spine in general, and yet others are generic (*i.e.*, not intended to assess any single health condition). Most of the questionnaires are also potentially useful outside the clinical setting (*e.g.*, for research studies and/or for screening).

Only those studies that involved patients seeking care for neck pain, using these self-assessment questionnaires, are reported here. This section identifies the assessment tools as specific to the neck pain population or generic. We also report results, where available, which are related to the questionnaire items, scaling, scoring, and/or psychometric properties. Psychometric properties include test-retest and inter-rater reliability, internal consistency, content, construct and predictive validity, and responsiveness to change.

The questionnaires are presented in groups according to the specific focus of the instrument: pain and self-assessment, function/disability and self-assessment, psychosocial items and self-assessment, and finally health care utilization and self-assessment.

**Pain and Self-Assessment.** Questionnaires that incorporate assessments of pain include the extended Aberdeen Spine Pain Scale (APS),<sup>101</sup> Bournemouth Questionnaire (BQ),<sup>102</sup> Cervical Spine Outcome Questionnaire (CSOQ),<sup>103</sup> Current Perceived Health 42 Profile (CPH42),<sup>104</sup> Neck Disability Index (NDI),<sup>62,105–108</sup> Problem Elicitation Technique (PET),<sup>105</sup> Sickness Impact Profile (SIP),<sup>51</sup> Visual Analog Scale (VAS),<sup>47,55,108–110</sup> and the Whiplash Disability Questionnaire (WDQ).<sup>111</sup>

The VAS is the most cited pain measure, largely because it is simple to use, has good psychometric properties and is often cited as the gold standard against which other questionnaires are judged.<sup>51,55,109,110</sup> Because we found no scientifically admissible studies specifically ad-

**Table 5. Self-Assessment Questionnaires (Specific and Generic) Designed for Patients With Neck Pain**

Questionnaire/Reference/Acronym (Alphabetical by title)	Constructs Measured (Pain; Function/Disability; Psychosocial Status; Health Care Utilization)	Classification (Specific or Generic)
Aberdeen spine pain scale (extended) (Williams <i>et al</i> <sup>101</sup> ) (APS)	Pain	Generic
Bournemouth questionnaire (Hurst and Bolton <sup>102</sup> ) (BQ)	Pain	Generic
Cervical spine outcome questionnaire (Bendebba <i>et al</i> <sup>103</sup> ) (CSOQ)	Pain; Function/Disability; Psychosocial Status; Health Care Utilization	Specific
Copenhagen neck functional disability scale (Jordan <i>et al</i> <sup>124</sup> ) (CNFDS)	Function/Disability	Specific
Current perceived health 42 profile (Chiu <i>et al</i> <sup>104</sup> ) (CPH42)	Pain; Function/Disability; Psychosocial Status	Generic
Global assessment of neck pain (Fejer <i>et al</i> <sup>114</sup> ) (GANP)	Function/Disability	Specific
Neck disability index (Hains <i>et al</i> <sup>62</sup> ) (NDI)	Pain; Function/Disability	Specific
Neck pain and disability scale (Ylinen <i>et al</i> <sup>47</sup> ) (NPDS)	Function/Disability	Specific
Northwick park neck pain questionnaire (Hoving <i>et al</i> <sup>105</sup> ) (NPQ)	Function/Disability	Specific
Problem elicitation technique (Hoving <i>et al</i> <sup>105</sup> ) (PET)	Pain; Function/Disability; Psychosocial Status	Generic
Sickness impact profile (Olson <i>et al</i> <sup>51</sup> ) (SIP)	Function/Disability	Generic
Visual analog scale (Wlodyka-Demaille <i>et al</i> <sup>108</sup> ) (VAS)	Pain; Function/Disability	Generic
Whiplash disability questionnaire (Pinfold <i>et al</i> <sup>111</sup> ) (WDQ)	Pain; Function/Disability; Psychosocial Status	Specific whiplash

Any questionnaire not designed specific for neck pain is considered generic. Reliability, validity, responsiveness and acceptability for each questionnaire is available where evidence exists on the NPTF website.

addressing the properties of neck pain, we decided that non-neck pain articles could be cited and were appropriate to include in this systematic review given. VAS is a generic pain instrument.<sup>112,113</sup>

The VAS is best at detecting change in patients who improve. The VAS has been used to show a weak association between pain and disability<sup>51</sup> and a negative correlation between neck strength output and pain.<sup>47</sup> The NDI, a neck-specific questionnaire, which overlaps with other measures, showed moderate to good agreement with the SF-36, a generic questionnaire, and the NDI is the most valid of the tools reported. The APS, CSOQ, CPH42, and NDI were all responsive to change, with some variation.<sup>101,103,104</sup> The BQ showed high sensitivity and specificity in distinguishing neck patients who had clinically significant improvement compared with those who did not improve, based on a 34% raw score change.<sup>102</sup> The APS was most responsive when health improved; this is typical for a questionnaire and responsiveness.<sup>101</sup> Most self-assessment questionnaires are more responsive in capturing health improvement than deterioration. The NDI discriminates between those who improved or deteriorated, but did not detect change in score in those who remained stable.<sup>108</sup> As expected, there is no change to detect if people remained stable; therefore, the instrument behaved appropriately.

**Function/Disability and Self-Assessment.** Eleven questionnaires were reviewed for evaluating function and disability in patients with cervical pain, with and without arm pain. These include the CSOQ,<sup>103</sup> Copenhagen Neck Functional Disability Scale (CNFDS),<sup>114,115</sup> Current Perceived Health 42 Profile (CPH42),<sup>104</sup> Global Assessment of Neck Pain (GANP),<sup>114</sup> Neck Disability Index (NDI),<sup>62,106,107</sup> Neck Pain and Disability Scale (and the modified scale) (NPDS),<sup>47,108,109</sup> Northwick Park Neck Pain Questionnaire (NPQ),<sup>105,108,116</sup> Problem Elicitation Technique (PET),<sup>105</sup> Sickness Impact Profile (SIP),<sup>51</sup> Visual Analog

Scale (VAS),<sup>47,51,108</sup> and Whiplash Disability Questionnaire (WDQ).<sup>111</sup> These questionnaires are denoted as generic or body specific for the neck by the authors of the cited studies (Table 5).

The CNFDS was tested on chronic neck pain patients and showed moderate to good validity.<sup>115</sup> The CSOQ and CNFDS both showed good reliability.<sup>103,115</sup> The NDI and VAS have been cited in the literature as the gold standard for other questionnaires.<sup>51,55,62,105,109,110</sup> Responsiveness to change was high for the CSOQ, NPDS, NPQ (between patients who improved or remained stable),<sup>103,108</sup> and VAS (in patients who improve).<sup>51</sup> The WDQ was shown to be a valid and reliable tool for assessing disability in WAD patients.<sup>111</sup> In WAD patients with several abnormal structures on MRI, significant increases in NDI scores were noted.<sup>106</sup> Acceptable cut points and predicting cut-points that differentiate level of severity have been described for the CNFDS and GANP.<sup>114</sup> The NPQ has been translated to Chinese and Spanish,<sup>110,116</sup> and the NPDS to Turkish.<sup>109</sup>

**Psychosocial Items and Self-Assessment.** The CSOQ,<sup>103</sup> PET,<sup>105</sup> and CPH42<sup>104</sup> all addressed the psychosocial well-being of neck patients. The CSOQ, despite its performance within other constructs, showed low to moderate responsiveness for psychological distress.<sup>103</sup> The CPH42, a 42-question scale, showed good reliability, moderate validity, and good responsiveness to change.<sup>104</sup> Although considerable overlap exists among the various questionnaires, the PET identifies emotional and social problems common in this population.<sup>105</sup>

**Health Care Utilization and Self-Assessment.** Only the CSOQ attempted to describe health care utilization. However, despite its performance within other constructs, it showed low to moderate responsiveness for health care utilization.<sup>103</sup>

**Table 6. Recommendations of Best Evidence Synthesis Based on Study Design and Consistency of Findings**

Consistency of Findings	Study Design		
	Strong (Phase III, IV) (Patient Relevant Outcome)	Moderate (Phase II, III) (Intermediary Patient Outcome)	Weak (Phase I, II)
<b>Emergency studies</b>			
Consistent	<ul style="list-style-type: none"> <li>● Use screening protocols for alert adults with low risk with blunt trauma to the neck</li> <li>● Use CT scan for high risk adults with blunt trauma to the neck</li> <li>● Do not use flexion-extension or 5 view X-ray in emergency neck patients</li> </ul>	<ul style="list-style-type: none"> <li>● CT scan for low risk adults with cervical blunt trauma to the neck</li> <li>● Screening protocol for children with blunt trauma to the neck</li> </ul>	
Inconsistent	<ul style="list-style-type: none"> <li>● Training radiologist to interpret X-ray film on neck trauma</li> </ul>		Training of EMT's
<b>Nonemergency studies</b>			
Consistent	<ul style="list-style-type: none"> <li>● Strong evidence for manual provocation test of the neck and upper extremity for radiculopathy</li> </ul>	<ul style="list-style-type: none"> <li>● Moderate evidence for physical exam (musculoskeletal and neurological exam) for ruling out condition</li> <li>● Moderate evidence for sensitivity to touch</li> </ul>	<ul style="list-style-type: none"> <li>● Weak consistent evidence against MRI</li> <li>● Weak consistent evidence for range of motion</li> <li>● Weak consistent evidence for palpation</li> <li>● Weak consistent evidence against use of routine blood tests</li> </ul>
Inconsistent	<ul style="list-style-type: none"> <li>● Strong evidence against manipulation as a diagnostic examination although findings should be replicated</li> </ul>	<ul style="list-style-type: none"> <li>● Moderate evidence against injections in chronic cervical pain</li> <li>● Moderate evidence for non organic tests</li> </ul>	<ul style="list-style-type: none"> <li>● Weak inconsistent evidence against discography</li> <li>● Weak evidence for muscle strength</li> <li>● Weak evidence for functional test</li> <li>● Weak evidence against nerve conduction</li> <li>● Weak evidence against cervical lordosis/kyphosis</li> <li>● Weak evidence against CT scan for nerve involvement</li> </ul>

■ Discussion

**Emergency Screening for Serious Neck Injuries in Patients With Blunt Trauma to the Neck**

There is strong evidence from several high quality phase III studies to suggest that practitioners can reliably employ either the Canadian C-spine rules (CCR)<sup>15</sup> or the Nexus low-risk criteria (NLC)<sup>20</sup> to rule out the need for further imaging in adult patients at low risk of neck pain injury seeking emergency care (Figure 1 and Table 6).<sup>13-22</sup>

There is strong evidence to suggest that use of routine cervical spine radiographs alone, (compared to CT scans) may miss important injuries in the evaluation of patients with traumatic high-risk neck injuries in emergency situations, and that CT scan should be used instead.<sup>10,24-30</sup> Coupled with the fact that there is important variability in reading the plain radiographs and that there is good evidence to suggest that the CT scan has a superior performance, routine radiographs alone may be superseded by CT in the setting of acute neck trauma in high-risk patients. Where CT scan facilities are not available to patients with high-risk injuries and radiographs are inconclusive, patients may need to be stabilized, and transported to facilities with other imaging alternatives. Enthusiasm for CT imaging in cervical trauma must be

tempered by the economic burden if universally applied and the much higher radiation exposure to sensitive tissues, especially in children and younger adults.<sup>124,125</sup>

Our evidence review suggests that there is lack of guidelines for children and neck trauma injuries; developing and testing such guidelines should be a priority for the clinical research community.

**Clinical Assessment of Nonemergency Neck Patients**

There is insufficient available evidence to confirm the utility of conventional “Red Flag Symptom” for triaging nonacute neck patients, although their use has been strongly encouraged<sup>41</sup> (Table 6). Although it is sensible that the same types of presentation (or predisposing risk) of serious structural disease that occurs in the lumbar spine<sup>39</sup> may also occur in the cervical spine, the cervical spine area has special anatomic considerations and risks (e.g., the presence of the spinal cord, specific rheumatoid destructive processes, specific adjacent vascular and visceral diseases). These idiosyncratic processes demands objective evidence and further studies be performed to define those subgroups of neck pain patients at higher risk as a result of these serious structural diseases.

We suggest a new classification for Neck Pain expressed as grade I-IV encompassing all neck pain building on the Québec Task Force Classification<sup>7</sup> as a diag-

nostic classification for the conditions including neck pain with and without trauma not leading to serious injury or diseases. WAD and other neck pain do not differ once serious neck conditions have been ruled out. The classification is based on 5 axes including the source of subjects, the setting and sampling of subjects or patients and the severity, duration, and pattern of neck pain.<sup>1</sup> The new proposed classification for neck pain and its associated disorders has not been validated.

Remarkably, there is little information on the validity or utility of the self-reported history in evaluating neck pain disorders. There is some information that self-reported questionnaires regarding past medical care may not have a high accuracy.<sup>103</sup> Similarly, data from the orthopedic trauma literature (not specifically reviewed for the Neck Pain Task Force) suggests the history received in specialty spine clinics in subjects reporting continued axial pain after MVA may systematically underestimate previous low back and neck pain problems and comorbidities associated with poor recovery.<sup>117</sup>

The current literature indicates that the clinical routine physical examination is more effective in ruling out cervical radiculopathy than confirming its presence. An exception is the manual provocation test for nerve root compromise, which seems to have a high sensitivity and high PPV.

As far as the physical examination of patients seeking care for neck pain with associated disorders, there is some evidence that some features of inspection, range of motion, strength, palpation, and provocation tests can be useful. Inspection of the neck patient for abnormal signs (for example, muscle wasting, swelling, redness, scars, and others) has low to moderate interexaminer reliability. Range of motion is moderately reliable, and it does not seem to matter whether it is assessed by the clinician (assessing active, or passive range of motion with or without a device), or whether it is self-described by the patient.

In addition, the available evidence suggests that subjects with neck pain identify discomfort with palpation for trigger points around the neck had moderate to high predictive value for neck pain with and without radiculopathy. Manual provocation tests designed to elicit nerve root compression in the cervical spine also have high positive predictive value (*i.e.*, ruling in radiculopathy). Beyond the physical examination, there is no good evidence from this systematic review that laboratory studies provide any unique value, or that surface, dermatomal, or quantitative sensory electrophysiological studies provide useful ancillary data. Needle EMG examination, although not specifically studied for cervical radiculopathy, is considered the gold standard test for denervation from any cause.

Several studies examined the role of imaging. There does not seem to be good evidence supporting the utility of plain radiographs in patients seeking nonacute care for neck pain who do not have major structural disease. No CT scan study was accepted for predictive values in the nonacute patient with neck pain with and with out

radiculopathy. Despite the many potential advantages of MRI in detecting major structural disease (*e.g.*, neoplasm, infection, *etc.*), current multiple scientifically admissible studies do not suggest that it has any unique role, independent of the history and clinical examination in detecting the cause of neck pain. Combined with symptoms of radicular complaints, specific findings on examination, and possibly needle EMG findings, the MRI may aid clinicians in determining the site and level of neurologic compression.

Other specialized investigative techniques, such as anesthetic facet joint injections and provocative discography, purported to “definitively” identify primary and specific lesions causing serious neck pain illness, do not seem to be supportable based on the current evidence and cannot be recommended as a routine part of clinical practice.

Patient self-assessment questionnaires seem to deserve greater use in routine clinical practice and research. The instruments cited have demonstrated acceptable reliability; many are suitable to characterize patients clinically, have good content validity and are responsive to changes of the patients self reported status. It is unclear from the current systematic review if specific results from these questionnaires are useful in predicting long-term outcome related to pain, disability, and employment.<sup>118–120</sup>

### **Some Limitations of Our Research**

There are limitations of this chapter that merit some discussion. Like all best evidence syntheses, this chapter is limited by both the quantity and the quality of the available evidence. We were surprised at the limited number of studies in several areas, for example, in special populations (children and elderly), electro diagnostics, functional testing, and the use of imaging in diagnosing patient with neck pain in nonemergency situation. We were also surprised at the limited quality of studies, notably, we found only 1 phase IV study addressing the health care consequences addressing mobilization of the neck,<sup>6</sup> and few phase III studies (including gold standard for assessment) in the nonemergency patient populations.

We realize that some readers, who are unfamiliar with the best evidence synthesis approach to summarizing the literature, may not appreciate its value. However, we feel that limiting our conclusions to studies that are of high methodologic quality is a notable strength. An uncritical mixing of studies of lower and higher quality scientific merit would yield potentially confusing and misleading results.

### **Directions for Future Research**

- There is an urgent need for studies of pediatric populations and neck trauma. It is important to understand if modifications of the CCR or the NLC apply to the pediatric population.
- There is a need to test several, potentially promising techniques or commonly used clinical tests in proper designs that have shown promise in phase I and II

studies, for example, test of nonorganic signs and functional tests.

- There is an immediate and strong need to test almost all commonly used clinical examination tests against gold standards, for predictive values and for utility in patients with non serious neck pain and associated disorders. Only provocation tests for cervical radiculopathy were well tested against gold standards<sup>8,55,64,66</sup> and manipulation was tested for utility.<sup>6</sup>

## ■ Evidence Statements

### ***Clinical Emergency Screening for Serious Neck Injury in Patients With Blunt Trauma to the Neck***

- There is strong consistent evidence from 11 studies (phase II and III) of large cohorts that using screening protocols for alert low-risk patients with blunt trauma to the neck will have high predictive values for detecting a cervical spine fracture. The CCR and the NLC have tested more than 40,000 patients. These protocols were tested against a 3-view radiograph as the gold standard, and appear to have an extremely low risk of missing a serious injury in this group.<sup>13–17,19–23,31,32</sup>

- There is consistent evidence that CT-scan (7 studies phase II and III) is more sensitive for finding significant cervical spine injury than plain 3-view radiograph in patients (adult and elderly) with cervical trauma for high risk and/or multi-injured blunt trauma neck patients seeking care in an emergency room.<sup>10,24,25,27–30</sup>

- There is evidence (1 phase I and 1 phase II study) suggesting indicators for screening for serious injury in children seeking care for neck trauma. Suggested indicators are neck pain, altered mental state, abnormal peripheral neurologic examination (sensation, reflexes, strength).<sup>33,34</sup>

- There is evidence against (1 phase I and 1 phase III study) the use of flexion/extension (F/E) radiographs or 5-view radiograph of the cervical spine in adults and children seeking emergency care for acute blunt trauma to the neck. F/E radiograph or 5-views radiograph did not have higher accuracy than standard 3-view radiograph in these studies.<sup>22,35</sup>

- There is limited evidence (1 phase III study) that specialty training for clinicians in the ability to interpret radiograph films in emergency situations for patients with blunt trauma to the neck improves the reliability of the image interpretation and thereby possibly increasing the diagnostic accuracy.<sup>36</sup>

- There is limited evidence (1 phase I study) of the predictive value using a specific screening protocol by EMT workers for immobilizing and transporting patients with suspected neck trauma to the emergency room.<sup>37</sup>

- There is no evidence (no study) to support the routine use of MRI as a screening tool after acute neck blunt trauma in an emergency setting.

### ***Clinical Assessment in Non-Emergency Care of Patients With Neck Pain (With and Without Arm Pain and/or Headache)***

#### ***Clinical Physical Examination***

- There is consistent evidence that the clinical physical examination is generally more predictive at excluding (“ruling out”) a structural lesion or neurologic compression than at diagnosing (“ruling in”) root compression and radiculopathy.<sup>8,55,64,66</sup>

- There is consistent evidence that measuring normal cervical range of motion (14 phase I–III studies) is equally reliable whether measured by visual estimation or external device. Patients’ estimates of reduced range of motion of the neck are less accurate.<sup>9,44–52,54,55,57,58</sup> There is evidence from 2 studies (phase I) that chronic WAD patients and subjects with neck pain and myalgia have less mobility in the cervical spine compared with controls.<sup>54,60</sup> There is evidence from 1 study that patients reporting acute WAD problems have decreased volitional range of motion of the neck compared to asymptomatic controls.<sup>56</sup>

- There is limited evidence (1 phase I study) that patients with chronic neck pain, on average, have slightly lower neck muscle strength compared with controls.<sup>50</sup>

- There is evidence (2 phase I studies) that cervical flexor endurance or arm flexor endurance can discriminate between subjects reporting chronic WAD II problems or subjects with neck pain and myalgia compared to controls.<sup>60,61</sup>

- There is consistent evidence that trigger-point palpation by a clinician (3 phase I studies) or “patient self-palpation” compared with physician palpation is reliable.<sup>58,63,64</sup> There is limited evidence (1 phase II study) that patients with neck pain and those with suspected radiculopathy have similar trigger point distributions.<sup>65</sup>

- There is consistent evidence in patients with radiculopathy (2 phase II studies) that sensory examinations, which demonstrate increased sensitivity to light touch and pin prick, are more reproducible than examinations demonstrating decreased sensation.<sup>44,55</sup> There is limited evidence (1 phase I study) that when subjects fail to identify a sensory change on self-assessment significant nerve root compression is highly unlikely to be found at physician examination.<sup>58</sup>

- There is limited evidence (1 phase I study) against the use of low-amplitude manipulation and endplay assessment of the cervical spine in patients with neck pain. One randomized phase IV trial showed that this assessment did not improve the primary outcome of same day pain level and stiffness relief observed in neck pain patients. These findings need to be replicated.<sup>6</sup>

- There is consistent evidence (3 phase III studies and 1 systematic review) to support the use of radicular pain provocation tests for neck patients to detect

- There is consistent evidence (3 phase III studies and 1 systematic review) to support the use of radicular pain provocation tests for neck patients to detect

probable nerve root compression findings. The most predictive test included contralateral neck rotation and extension of the arm and the fingers of the affected side.<sup>8,55,64,66</sup>

- There is evidence against the use of routine blood tests to distinguish patients with acute whiplash exposure or chronic neck pain complaints from those subjects without exposure to whiplash or chronic neck troubles (2 phase I studies).<sup>63,69</sup> Routine blood tests could not distinguish patients from nonpatients at late stage of WAD or chronic neck pain.
- There is limited evidence (1 phase II study) that patients with chronic neck pain may perform less well on certain functional test.<sup>67</sup>
- There is consistent evidence that nonorganic sign tests had high inter-rater variability among clinicians testing patients with chronic neck pain.<sup>55,68</sup>
- There is evidence against the use of electrodiagnostic testing in patients with neck pain without suspected radiculopathy. Two studies (phase I and II) found that surface EMG activity of the upper trapezius muscle did not distinguish between subjects with and without neck pain.<sup>71,72</sup>
- There is no evidence that the degree of cervical lordosis or kyphosis can accurately distinguish “cervical muscle spasm” or subjects with whiplash exposure from those with no exposure to whiplash. One study (phase I) found that there is no difference in cervical lordosis or kyphosis in patients with subacute WAD compared with controls as documented by radiograph.<sup>75</sup>
- There is no evidence (no scientifically admissible studies) to support the use of surface electromyography, dermatomal somatosensory-evoked responses or quantitative sensory testing in the diagnosis of radiculopathy.
- There is limited evidence (1 phase II study) that the assessment of root compression or canal stenosis of the cervical spine by CT scan has fair to moderate reliability.<sup>77</sup>
- There is no evidence that pain reproduction on provocative disc injection identifies the injected disc as the cause of primary serious neck pain problems. There is weak evidence against provocative discography of the cervical spine in patients with neck pain. There is evidence (1 phase II study) that pain response to provocative discography cannot accurately distinguish between subjects with and without neck pain.<sup>78</sup> There is no evidence that provocative cervical discography has clear utility in treating patients with neck pain (*i.e.*, improves outcomes).<sup>78,79</sup>
- There is consistent evidence from (4 phase I studies) that the identification of common degenerative changes in the cervical spine, identified by MRI is at best fair to moderately reproducible.<sup>77,82,84,85</sup>
- There is evidence against the use of a digitizer to enhance MRI readings or enhanced MRI (2 phase II studies) to improve reliability in reading MRIs for the cervical spine findings.<sup>83,87</sup>

- There is evidence (2 phase II studies) that cervical MRI findings of a hard disc or extrusion of disc material through the cervical posterior longitudinal ligament are often not in agreement with the surgeon-reported findings at surgery.<sup>88,89</sup>
- There is no evidence that common degenerative changes on cervical MRI are strongly correlated with neck pain symptoms. There is evidence (4 phase I and II studies) that MRI findings of the cervical spine of common degenerative changes are highly prevalent in asymptomatic subjects. Abnormal MRI findings of the cervical spine also found to increase with age.<sup>82,90-92</sup>
- There is evidence (1 phase I study) that frequent exposure to extremely high *g*-forces in senior fighter pilots compared to controls is associated with increased cervical disc degeneration.<sup>93</sup>
- There is no evidence that standard sequence MRI accurately detect specific trauma-related findings in the subaxial cervical spine in the absence of fracture, dislocation or major ligamentous disruption. There is evidence (1 phase II study) that patients with acute WAD do not have soft tissue lesions of the cervical spine demonstrated by MRI.<sup>94</sup>
- The validity of high-intensity signals MRI findings in the upper cervical spine ligaments as representing acute whiplash injury has not been demonstrated. There is evidence (3 phase I studies) that identifying signal changes in the ligaments of the upper cervical spine in late stage of WAD by special sequence MRI had slight to moderate reliability.<sup>84-86</sup> The utility of this finding in diagnosing *bona fide* and clinically relevant ligamentous injury and directing effective treatment has not been demonstrated in WAD patients (grade I-III).
- There is no evidence that common degenerative changes on cervical MRI are associated with pain in patients with supposed cervicogenic headache. One phase I study found similar MRI findings in cervicogenic headache patients and asymptomatic controls.<sup>95</sup>
- There is no evidence supporting the validity of diagnostic facet joint or medial branch blocks as diagnosing cervical facet joint pain as the primary cause of serious neck pain illness. There is evidence against (4 phase II studies) the use of diagnostic facet joint or medial branch injections of the cervical spine; these studies show poor reliability. There is no evidence that the use of diagnostic facet injections improves treatment outcomes (utility) in patients with chronic neck pain.<sup>97-100</sup>

#### **Self-assessment by Questionnaires**

- There is consistent evidence that patient self-assessment questionnaires may have utility in routine clinical practice and research by characterizing patients' clinical presentation, subjective functional impact of neck pain and course over time.
- There is no evidence (no studies) that a self-assessment questionnaire alone can accurately diagnose a structural cause of illness in patients with neck

pain. However, the questionnaires cited in this systematic literature review can provide useful information regarding patient self-assessment for pain, function, and perceived disability and psychosocial status.

- Overall there was evidence for moderate to strong performance of all the questionnaires cited for reliability, validity, and responsiveness to change in this systematic literature review. Not all parameters of performance for an instrument cited have been measured for all questionnaires in the scientifically admissible studies.<sup>47,62,101–103,105–109,111,114–116</sup>
- There is evidence (14 studies) that neck specific questionnaires are more responsive to changes in the neck pain and to differences among various groups of patients with neck pain than generic pain questionnaires.<sup>47,62,101–103,105–109,111,114–116</sup>
- There is evidence that generic questionnaires may be more useful than neck specific questionnaires for comparing individuals with neck pain with other disease groups (see Table 5).<sup>47,51,55,104,108–110</sup>
- There is evidence against (1 study) in patients with neck pain to use self-assessment questionnaires to monitor health care utilization *i.e.*, the study showed that patients had poor recollection of healthcare utilization.<sup>103</sup>

### Research Recommendations

As the preceding evidence statements suggest, there are large areas in the diagnostic testing of neck pain associated disorders that are poorly validated, even at the most elementary levels. Few clinical entities related to neck pain have been systematically investigated except the emergency screening for blunt trauma to the neck. Despite the lack of adequate supporting evidence apparent in our comprehensive review, in clinical practice “diagnoses of convenience” are often made. Diagnoses such as “cervical sprain,” “minor facet subluxation,” “primary discogenic pain,” “internal disc derangement,” “postural neck pain,” “primary zygapophysial pain,” and others have been in common usage for decades, often without confirmation of the entity itself or any means of diagnosis according to accepted scientific methods. Tests claiming to make these diagnoses need to be rigorously tested and clear strategies to do so have been well described.<sup>5,121–123</sup>

Investigators interested in designing appropriate studies are strongly advised to consider established guidelines to ensure study validity (*e.g.*, appropriate subject composition, avoidance of work-up bias, avoidance of review bias, testing for reproducibility, and others). Equally important is the need to clearly establish gold standard comparisons that exist outside the test being evaluated.<sup>117,122</sup> As Greenhalgh<sup>122</sup> points out one must be certain “that the test being validated is not being used to define the gold standard.” This is very common error in spinal diagnostic research.

Most importantly, it must be clear where the burden of proof lies in the study of diagnostic methods. Diag-

nostic tests must be assumed clinically uninterpretable until their validity and limitations are established. More often in our review, we have seen tests advocated for popular use on the reverse premise: the tests must be assumed valid until someone can show they are worthless (*e.g.*, provocative discography). Tragically, the spinal literature of the last century is littered with “definitive diagnoses” of pain syndromes, based on a novel test only to be abandoned as invalid only after many years of inappropriate use (*e.g.*, “definitive axial pain diagnoses” on the basis of radiographs showing bone spurs or minor alignment changes, bone scans showing increased uptake, MRI showing disc signal loss, facet injections giving temporary pain relief, *etc.*).

Clinicians should know what a test’s accuracy and limitations are before using it, clinician–investigators need to appreciate that it is very difficult and sometimes impossible to scientifically disprove an ill-defined theory whether it be “intelligent design,” “cervical sprain,” “joint instability,” or “internal disc derangement.”

### Specific Areas of Inquiry

- There is a need to establish screening criteria for infants, children, and adolescents seeking care in an emergency room for blunt trauma to the neck (phase III and IV studies).
- There is a lack of consistency among emergency physicians to interpret radiograph and other imaging in emergency situations for patients with blunt trauma to the neck. Better-designed studies are needed for the most efficacious training of imaging interpretation of these patients’ films.
- There is a need to validate and establish the relative utility and cost-effectiveness of screening patients seeking treatment for nonemergency/nontraumatic neck pain for serious structural disease (“Neck Pain Red Flags”) (phase I–IV studies).
- There is a need to confirm the validity and utility (phases III–IV study) of the clinical musculoskeletal neck examination in patients with neck pain without radiculopathy.
- There is a need to further establish reliability (phase II) and to establish validity and utility of muscle strength and endurance testing of the neck (phase III and IV studies).
- There is a need to replicate the evidence against the utility of using manipulation of the neck to direct specific treatment in patients with neck pain with or without radiculopathy (phase IV study).
- There is a need to establish reliability, validity, and utility of functional capacity testing in patients with neck pain with and without radiculopathy (phase I–IV studies).
- There is a need to establish validity and utility of the nonorganic-signs test in patients with chronic neck pain (phase III and IV studies).

- There is a need to evaluate dermatomal somatosensory-evoked responses or quantitative testing in the diagnosis of radiculopathy (phase I–IV studies).
- There is a need for more robust studies to validate the utility of CT-scan in the assessment of root compression in patients with neck pain and radiculopathy (phase II–IV studies).
- There is a need to demonstrate validity and utility of MRI for patients with acute and chronic WAD II in well-designed studies (phase III–IV studies).
- There is a need to examine the gold standard criteria for many basic neck pain diagnoses. Of the many unvalidated tests and diagnoses these common purported diagnoses may deserve early attention: “cervical strain,” “spinal malalignment,” “cervical instability,” “zygapophysial pain,” “cervicogenic headache,” “internal disc derangement,” “discogenic neck pain,” or “minor disc protrusion” as a cause of neck pain without radiculopathy.
- There is a need to identify clinical subgroups of patients (with neck pain and radicular pain) who are most likely to respond to standard surgical treatment (phase III and IV studies).
- There is a need to measure all performance parameters simultaneously (reliability, validity, responsiveness to change, and easy of administration) in the neck specific self-assessment questionnaires.
- There is a need to identify or develop questionnaires useful to describe healthcare utilization in patients with neck pain with and without radiculopathy and/or headache.

### ■ Key Points

The scientific evidence strongly supports the use of:

- Screening protocols in emergency care in low-risk patient with blunt trauma to the neck.
- CT-scanning in emergency care for high-risk patients with blunt trauma to the neck.

In patients seeking care for nonemergency neck pain, the scientific evidence supports the use of:

- Manual provocation tests in patients with neck pain and suspected radiculopathy.
- The combination of history, physical examination, modern imaging techniques, and needle EMG to diagnose the cause and site of cervical radiculopathy.
- Self-reported patient assessment to evaluate perceived pain, function, disability, and psychosocial status.

In patients seeking care for nonemergency neck pain, there is no evidence to support the diagnostic validity or utility of:

- Provocative discography.

- Anesthetic facet or medial branch blocks.
- Surface electromyography, dermatomal somatosensory-evoked responses or quantitative sensory testing in the diagnosis of radiculopathy.

 tables

Tables available online through Article Plus.

### Acknowledgments

The authors are deeply grateful to Oksana Colson, Stephen Greenhalgh, and Leah Phillips the wonderful “support group” from Department of Epidemiology, University of Alberta, Edmonton, Canada. This group never said no and was always ready to help, sort out, and find information. The authors express a big thank you to them. They are very grateful to Maria Trujillo Ponte, executive secretary for always being there when help was needed at the Occupational and Industrial Orthopedic Center (OIOC), NYU Hospital for Joint Diseases, NY University Medical Center. They are also indebted to Ms. C. Sam Cheng (MLIS) and Ms. Lori Giles-Smith (MLIS), research librarians, for their assistance in the work of the Neck Pain Task Force. The Bone and Joint Decade 2000–2010 Task Force on Neck Pain and Its Associated Disorders was supported by grants from the following: National Chiropractic Mutual Insurance Company (USA); Canadian Chiropractic Protective Association (Canada); State Farm Insurance Company (USA); Insurance Bureau of Canada; *Länsförsäkringar* (Sweden); *The Swedish Whiplash Commission*; Jalan Pacific Inc. (Brazil); Amgen (USA). All funds received were unrestricted grants. Funders had no control in planning, research activities, analysis or results. The report was not released to grantors prior to publication and no approval was required from funders regarding the final report. Dr. Côté is supported by the Canadian Institutes of Health Research through a New Investigator Award and by the Institute for Work & Health through the Workplace Safety and Insurance Board of Ontario. Dr. van der Velde is supported by the Canadian Institutes of Health Research through a Fellowship Award. Dr. Carroll is supported by a Health Scholar Award from the Alberta Heritage Foundation for Medical Research. Dr. Cassidy is supported by an endowed research chair from the University Health Network.

### References

1. Guzman J, Hurwitz EL, Carroll LJ, et al. A conceptual model for the course and care of neck pain. Results of the bone and joint decade 2000–2010 task force on neck pain and its associated disorders. *Spine* 2008;33(suppl):S14–S23.
2. Carroll LJ, Cassidy JD, Peloso PM, et al. Methods for the best evidence synthesis on neck pain and its associated disorders. The bone and joint decade 2000–2010 task force on neck pain and its associated disorders. *Spine* 2008;33(suppl):S33–S38.
3. Altman DG, Bland JM. Diagnostic tests. I. Sensitivity and specificity. *BMJ* 1994;308:1552.
4. Jaeschke R, Guyatt G, Sackett DL. Users’ guides to the medical literature.

- III. How to use an article about a diagnostic test, part A: are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA* 1994; 271:389–91.
5. Sackett DL, Haynes RB. The architecture of diagnostic research. *BMJ* 2002; 324:539–41.
  6. Haas M, Group E, Panzer D, et al. Efficacy of cervical endplate assessment as an indicator for spinal manipulation. *Spine* 2003;28:1091–6.
  7. Spitzer WO, Skovron ML, Salmi LR, et al. Scientific monograph of the Quebec Task Force on Whiplash-Associated Disorders: redefining “whiplash” and its management. *Spine* 1995;20:15–73S.
  8. Rubinstein S, Pool JJ, van Tulder M, et al. A systematic review of the diagnostic accuracy of provocative tests of the neck for diagnosing cervical radiculopathy. *Eur Spine J* 2007;16:307–19.
  9. van Trijffel E, Anderegg Q, Bossuyt PM, et al. Inter-examiner reliability of passive assessment of intervertebral motion in the cervical and lumbar spine: a systematic review. *Manual Therapy* 2005;10:256–69.
  10. Holmes JF, Akkinepalli R. Computed tomography versus plain radiography to screen for cervical spine injury: a meta-analysis. *J Trauma Injury Infect Crit Care* 2005;58:902–5.
  11. Moskovich R, Petrizzo A. Evaluation of the neck. In: Nordin M, Andersson GBJ, Pope M, eds. *Musculoskeletal Disorders in the Workplace*. 2nd ed. Philadelphia, PA: Mosby Elsevier; 1997;55–72.
  12. Guzman J, Haldeman S, Carroll LJ, et al. Practice implications of the results of the bone and joint decade 2000–2010 task force on neck pain and its associated disorders: from concepts and findings to recommendations. *Spine* 2008;33(suppl):S198–S211.
  13. Kerr D, Bradshaw L, Kelly AM. Implementation of the Canadian C-spine rule reduces cervical spine x-ray rate for alert patients with potential neck injury. *J Emerg Med* 2005;28:127–31.
  14. Stiell I, McKnight R, Schull M, et al. The Canadian C-spine rules versus the NEXUS low-risk criteria in patients with trauma. *N Engl J Med* 2003;349: 2510–8.
  15. Stiell I, Wells GA, Vandemheen K. The Canadian C-spine rule for radiography in alert and stable trauma patients. *JAMA* 2001;286:1841–8.
  16. Dearden C, Hughes D. Does the national emergency x-ray utilization study make a difference? *Eur J Emerg Med* 2005;12:278–81.
  17. Dickinson G, Stiell IG, Schull M, et al. Retrospective application of the NEXUS low-risk criteria for cervical spine radiography in Canadian emergency departments. *Ann Emerg Med* 2004;43:507–14.
  18. Heffernan DS, Schermer CR, Lu SW. What defines a distracting injury in cervical spine assessment? *J Trauma Injury Infect Crit Care* 2005;59: 1396–9.
  19. Hoffman JR, Schriger DL, Mower W, et al. Low-risk criteria for cervical-spine radiography in blunt trauma: a prospective study. *Ann Emerg Med* 1992;21:1454–60.
  20. Hoffman JR, Mower WR, Wolfson AB, et al. Validity of a set of clinical criteria to rule out injury to the cervical spine in patients with blunt trauma. National Emergency X-Radiography Utilization Study Group. *N Engl J Med* 2000;343:94–9.
  21. Panacek EA, Mower WR, Holmes JF, et al. Test performance of the individual NEXUS low-risk clinical screening criteria for cervical spine injury. *Ann Emerg Med* 2001;38:22–5.
  22. Pollack CV Jr, Hendey GW, Martin DR, et al. Use of flexion-extension radiographs of the cervical spine in blunt trauma. *Ann Emerg Med* 2001; 38:8–11.
  23. Touger M, Gennis P, Nathanson N, et al. Validity of a decision rule to reduce cervical spine radiography in elderly patients with blunt trauma. *Ann Emerg Med* 2002;40:287–93.
  24. Blackmore CC, Ramsey SD, Mann FA, et al. Cervical spine screening with CT in trauma patients: a cost-effectiveness analysis. *Radiology* 1999;212: 117–25.
  25. Diaz JJ Jr, Gillman C, Morris JA Jr, et al. Are five-view plain films of the cervical spine unreliable? A prospective evaluation in blunt trauma patients with altered mental status. *J Trauma Injury Infect Crit Care* 2003;55:658–63.
  26. Gale SC, Gracias VH, Reilly PM, et al. The inefficiency of plain radiography to evaluate the cervical spine after blunt trauma. *J Trauma Injury Infect Crit Care* 2005;59:1121–5.
  27. Griffen MM, Frykberg ER, Kerwin AJ, et al. Radiographic clearance of blunt cervical spine injury: plain radiograph or computed tomography scan? *J Trauma Injury Infect Crit Care* 2003;55:222–6.
  28. McCulloch PT, France J, Jones DL, et al. Helical computed tomography alone compared with plain radiographs with adjunct computed tomography to evaluate the cervical spine after high-energy trauma. *J Bone Joint Surg Am Vol* 2005;87:2388–94.
  29. Streitwieser DR, Knopp R, Wales LR, et al. Accuracy of standard radiographic views in detecting cervical spine fractures. *Ann Emerg Med* 1983; 12:538–42.
  30. Suzuki T, Morimura N, Sugiyama M, et al. How often should computed tomographic scans following cross-table lateral cervical films be performed? *J Orthopaedic Surg* 2004;12:40–4.
  31. Neifeld GL, Keene JG, Hevesy G, et al. Cervical injury in head trauma. *J Emerg Med* 1988;6:203–7.
  32. Zabel DD, Tinkoff G, Wittenborn W, et al. Adequacy and efficacy of lateral cervical spine radiography in alert, high-risk blunt trauma patient. *J Trauma Injury Infect Crit Care* 1997;43:952–6.
  33. Browne GJ, Lam LT, Barker RA. The usefulness of a modified adult protocol for the clearance of paediatric cervical spine injury in the emergency department. *Emerg Med* 2003;15:133–42.
  34. Jaffe DM, Binns H, Radkowski MA, et al. Developing a clinical algorithm for early management of cervical spine injury in child trauma victims. *Ann Emerg Med* 1987;16:270–6.
  35. Dwek JR, Chung CB. Radiography of cervical spine injury in children: are flexion-extension radiographs useful for acute trauma? *AJR* 2000;174: 1617–9.
  36. Annis JA, Finlay DB, Allen MJ, et al. A review of cervical-spine radiographs in casualty patients. *Br J Radiol* 1987;60:1059–61.
  37. Domeier RM, Frederiksen SM, Welch K. Prospective performance assessment of an out-of-hospital protocol for selective spine immobilization using clinical spine clearance criteria. *Ann Emerg Med* 2005;46:123–31.
  38. Bigos S, Bowyer O, et al. Acute low back pain in adults. Clinical Practice Guidelines no. 14 AHCPR. 95–0642. 1–12-1194. Rockville MD, Agency for Healthcare Policy and Research, Public Health Service, U.S. Department of Health and Human Services.
  39. Deyo RA, Rainville J, Kent DL. What can the history and physical examination tell us about low back pain? *JAMA* 1992;268:760–5.
  40. European Guidelines for Low Back Pain. 2004. [http://www.backpainurope.org/web/files/WG1\\_Guidelines.pdf](http://www.backpainurope.org/web/files/WG1_Guidelines.pdf).
  41. Binder AI. Cervical spondylosis and neck pain. *BMJ* 2007;334:527–31.
  42. Strender LE, Lundin M, Nell K. Interexaminer reliability in physical examination of the neck. *J Manipulative Physiol Ther* 1997;20:516–20.
  43. Salerno DF, Franzblau A, Werner RA, et al. Reliability of physical examination of the upper extremity among keyboard operators. *Am J Ind Med* 2000;37:423–30.
  44. Viikari-Juntura E. Interexaminer reliability of observations in physical examinations of the neck. *Phys Ther* 1987;67:1526–32.
  45. Pool JJ, Hoving JL, de Vet HC, et al. The interexaminer reproducibility of physical examination of the cervical spine. *J Manipulative Physiol Ther* 2004;27:84–90.
  46. Smedmark V, Wallin M, Arvidsson I. Inter-examiner reliability in assessing passive intervertebral motion of the cervical spine. *Manual Therapy* 2000; 5:97–101.
  47. Ylmen J, Takala EP, Kautiainen H, et al. Association of neck pain, disability and neck pain during maximal effort with neck muscle strength and range of movement in women with chronic non-specific neck pain. *Eur J Pain* 2004; 8:473–8.
  48. Hanten WP, Olson SL, Russell JL, et al. Total head excursion and resting head posture: normal and patient comparisons. *Arch Phys Med Rehabil* 2000;81:62–6.
  49. Hoving JL, Pool JJ, van MH, et al. Reproducibility of cervical range of motion in patients with neck pain. *BMC* 2005;6:59.
  50. Jordan A, Mehlsen J, Ostergaard K. A comparison of physical characteristics between patients seeking treatment for neck pain and age-matched healthy people. *J Manipulative Physiol Ther* 1997;20:468–75.
  51. Olson SL, O'Connor DP, Birmingham G, et al. Tender point sensitivity, range of motion, and perceived disability in subjects with neck pain. *J Orthop Sports Phys Ther* 2000;30:13–20.
  52. Osterbauer PJ, Long K, Ribaud TA, et al. Three-dimensional head kinematics and cervical range of motion in the diagnosis of patients with neck trauma. *J Manipulative Physiol Ther* 1996;19:231–7.
  53. Petersen CM, Johnson RD, Schuit D. Reliability of cervical range of motion using the OSI CA 6000 spine motion analyser on asymptomatic and symptomatic subjects. *Manual Therapy* 2000;5:82–8.
  54. Puglisi F, Ridi R, Cecchi F, et al. Segmental vertebral motion in the assessment of neck range of motion in whiplash patients. *Int J Legal Med* 2004; 118:235–9.
  55. Wainner RS, Fritz JM, Irrgang JJ, et al. Reliability and diagnostic accuracy of the clinical examination and patient self-report measures for cervical radiculopathy. *Spine* 2003;28:52–62.
  56. Hartling L, Brison RJ, Arden C, et al. Prognostic value of the Quebec Classification of Whiplash-Associated Disorders. *Spine* 2001;26:36–41.
  57. Hagen KB, Harms-Ringdahl K, Enger NO, et al. Relationship between

- subjective neck disorders and cervical spine mobility and motion-related pain in male machine operators. *Spine* 1997;22:1501-7.
58. Toomingas A, Nemeth G, Alfredsson L. Self-administered examination versus conventional medical examination of the musculoskeletal system in the neck, shoulders, and upper limbs. *J Clin Epidemiol* 1995;48:1473-83.
  59. Kristjansson E, Hardardottir L, Asmundardottir M, et al. A new clinical test for cervicocephalic kinesthetic sensibility: "the fly." *Arch Phys Med Rehabil* 2004;85:490-5.
  60. Larsson B, Bjork J, Elert J, et al. Mechanical performance and electromyography during repeated maximal isokinetic shoulder forward flexions in female cleaners with and without myalgia of the trapezius muscle and in healthy controls. *Eur J Appl Physiol* 2000;83:257-67.
  61. Kumbhare DA, Balsor B, Parkinson WL, et al. Measurement of cervical flexor endurance following whiplash. *Disabil Rehabil* 2005;27:801-7.
  62. Hains F, Waalen J, Mior S. Psychometric properties of the neck disability index. *J Manipulative Physiol Ther* 1998;21:75-80.
  63. Andersen JH, Gaardboe O. Musculoskeletal disorders of the neck and upper limb among sewing machine operators: a clinical investigation. *Am J Ind Med* 1993;24:689-700.
  64. Sandmark H, Nisell R. Validity of five common manual neck pain provoking tests. *Scand J Rehabil Med* 1995;27:131-6.
  65. Hsueh TC, Yu S, Kuan TS, et al. Association of active myofascial trigger points and cervical disc lesions. *J Formos Med Assoc* 1998;97:174-80.
  66. Viikari-Juntura E, Porras M, Laasonen EM. Validity of clinical tests in the diagnosis of root compression in cervical disc disease. *Spine* 1989;14:253-7.
  67. Ljungquist T, Jensen IB, Nygren A, et al. Physical performance tests for people with long-term spinal pain: aspects of construct validity. *J Rehabil Med* 2003;35:69-75.
  68. Sobel JB, Sollenberger P, Robinson R, et al. Cervical nonorganic signs: a new clinical tool to assess abnormal illness behavior in neck pain patients: a pilot study. *Arch Phys Med Rehabil* 2000;81:170-5.
  69. Kivioja J, Rinaldi L, Ozenci V, et al. Chemokines and their receptors in whiplash injury: elevated RANTES and CCR-5. *J Clin Immunol* 2001;21:272-7.
  70. Shy ME, Frohman EM, So YT, et al. Quantitative sensory testing: report of the therapeutics and technology assessment subcommittee of the American Academy Of Neurology. *Neurology* 2003;60:898-904.
  71. Westgaard RH, Vasseljen O, Holte KA. Trapezius muscle activity as a risk indicator for shoulder and neck pain in female service workers with low biomechanical exposure. *Ergonomics* 2001;44:339-53.
  72. Carlson CR, Wynn KT, Edwards J, et al. Ambulatory electromyogram activity in the upper trapezius region: patients with muscle pain vs. pain-free control subjects. *Spine* 1996;21:595-9.
  73. Keidel M, Rieschke P, Stude P, et al. Antinociceptive reflex alteration in acute posttraumatic headache following whiplash injury. *Pain* 2001;92:319-26.
  74. Pullman SL, Goodin DS, Marquez AI, et al. Clinical utility of surface EMG: report of the therapeutics and technology assessment subcommittee of the American Academy Of Neurology. *Neurology* 2000;55:171-7.
  75. Matsumoto M, Fujimura Y, Suzuki N, et al. Cervical curvature in acute whiplash injuries: prospective comparative study with asymptomatic subjects. *Injury* 1998;29:775-8.
  76. Shaw DD, Bach-Gansmo T, Dahlstrom K. Iohexol: summary of North American and European clinical trials in adult lumbar, thoracic, and cervical myelography with a new nonionic contrast medium. *Invest Radiol* 1985;20:S44-50.
  77. Stafira JS, Sonnad JR, Yuh WT, et al. Qualitative assessment of cervical spinal stenosis: observer variability on CT and MR images. *AJNR Am J Neuroradiol* 2003;24:766-9.
  78. Schellhas KP, Smith MD, Gundry CR, et al. Cervical discogenic pain. Prospective correlation of magnetic resonance imaging and discography in asymptomatic subjects and pain sufferers. *Spine* 1996;21:311-2.
  79. Slipman CW, Plastaras C, Patel R, et al. Provocative cervical discography symptom mapping. *Spine* 2005;5:381-8.
  80. Carragee EJ, Tanner CM, Khurana S, et al. The rates of false-positive lumbar discography in select patients without low back symptoms. *Spine* 2000;25:1373-80.
  81. Walsh TR, Weinstein JN, Spratt KF, et al. Lumbar discography in normal subjects. A controlled, prospective study. *J Bone Joint Surg Am* 1990;72:1081-8.
  82. Matsumoto M, Fujimura Y, Suzuki N, et al. MRI of cervical intervertebral discs in asymptomatic subjects. *J Bone Joint Surg Br Vol* 1998;80:19-24.
  83. Cooley JR, Danielson CD, Schultz GD, et al. Posterior disk displacement: morphologic assessment and measurement reliability—cervical spine. *J Manipulative Physiol Ther* 2001;24:560-8.
  84. Krakenes J, Kaale BR, Moen G, et al. MRI assessment of the alar ligaments in the late stage of whiplash injury—a study of structural abnormalities and observer agreement. *Neuroradiology* 2002;44:617-24.
  85. Krakenes J, Kaale BR, Moen G, et al. MRI of the tectorial and posterior atlanto-occipital membranes in the late stage of whiplash injury. *Neuroradiology* 2003;45:585-91.
  86. Krakenes J, Kaale BR, Nordli H, et al. MR analysis of the transverse ligament in the late stage of whiplash injury. *Acta Radiol* 2003;44:637-44.
  87. Ross JS, Ruggieri PM, Tkach JA, et al. Gd-DTPA-enhanced 3D MR imaging of cervical degenerative disk disease: initial experience. *AJNR Am J Neuroradiol* 1992;13:127-36.
  88. Humphreys SC, Hodges SD, Fisher DL, et al. Reliability of magnetic resonance imaging in predicting disc material posterior to the posterior longitudinal ligament in the cervical spine. A prospective study. *Spine* 1998;23:2468-71.
  89. Sengupta DK, Kirillos R, Findlay GF, et al. The value of MR imaging in differentiating between hard and soft cervical disc disease: a comparison with intraoperative findings. *Eur Spine J* 1999;8:199-204.
  90. Boden SD, McCowin PR, Davis DO, et al. Abnormal magnetic-resonance scans of the cervical spine in asymptomatic subjects. A prospective investigation. *J Bone Joint Surg Am Vol* 1990;72:1178-84.
  91. Lehto IJ, Tertti MO, Komu ME, et al. Age-related MRI changes at 0.1 T in cervical discs in asymptomatic subjects. *Neuroradiology* 1994;36:49-53.
  92. Siivola SM, Levoska S, Tervonen O, et al. MRI changes of cervical spine in asymptomatic and symptomatic young adults. *Eur Spine J* 2002;11:358-63.
  93. Hamalainen O, Vanharanta H, Kuusela T. Degeneration of cervical intervertebral disks in fighter pilots frequently exposed to high +Gz forces. *Aviat Space Environ Med* 1993;64:692-6.
  94. Borchgrevink G, Smevik O, Haave I, et al. MRI of cerebrum and cervical column within two days after whiplash neck sprain injury. *Injury* 1997;28:331-5.
  95. Coskun O, Ucler S, Karakurum B, et al. Magnetic resonance imaging of patients with cervicogenic headache. *Cephalalgia* 2003;23:842-5.
  96. Barnsley L, Lord S, Wallis B, et al. False-positive rates of cervical zygapophysial joint blocks. *Clin J Pain* 1993;9:124-30.
  97. Barnsley L, Lord SM, Wallis BJ, et al. The prevalence of chronic cervical zygapophysial joint pain after whiplash. *Spine* 1995;20:20-5.
  98. Bogduk N, Marsland A. The cervical zygapophysial joints as a source of neck pain. *Spine* 1988;13:610-7.
  99. Lord SM, Barnsley L, Bogduk N. The utility of comparative local anesthetic blocks versus placebo-controlled blocks for the diagnosis of cervical zygapophysial joint pain. *Clin J Pain* 1995;11:208-13.
  100. Barnsley L, Lord S, Bogduk N. Comparative local anaesthetic blocks in the diagnosis of cervical zygapophysial joint pain. *Pain* 1993;55:99-106.
  101. Williams NH, Wilkinson C, Russell IT. Extending the Aberdeen back pain scale to include the whole spine: a set of outcome measures for the neck, upper and lower back. *Pain* 2001;94:261-74.
  102. Hurst H, Bolton J. Assessing the clinical significance of change scores recorded on subjective outcome measures. *J Manipulative Physiol Ther* 2004;27:26-35.
  103. Bendebba M, Heller J, Ducker TB, et al. Cervical spine outcomes questionnaire: its development and psychometric properties. *Spine* 2002;27:2116-23.
  104. Chiu TT, Lam TH, Hedley AJ. Psychometric properties of a generic health measure in patients with neck pain. *Clin Rehabil* 2003;50:5-13.
  105. Hoving JL, O'Leary EF, Niere KR, et al. Validity of the neck disability index, Northwick Park neck pain questionnaire, and problem elicitation technique for measuring disability associated with whiplash-associated disorders. *Pain* 2003;102:273-81.
  106. Kaale BR, Krakenes J, Albrektsen G, et al. Head position and impact direction in whiplash injuries: associations with MRI-verified lesions of ligaments and membranes in the upper cervical spine. *J Neurotrauma* 2005;22:1294-302.
  107. Riddle DL, Stratford PW. Use of generic versus region-specific functional status measures on patients with cervical spine disorders. *Phys Ther* 1998;78:951-63.
  108. Wlodyka-Demaille S, Poiraudou S, Catanzariti JF, et al. The ability to change of three questionnaires for neck pain. *Spine* 2004;71:317-26.
  109. Bicer A, Yazici A, Camdeviren H, et al. Assessment of pain and disability in patients with chronic neck pain: reliability and construct validity of the Turkish version of the neck pain and disability scale. *Disabil Rehabil* 2004;26:959-62.
  110. Gonzalez T, Balsa A, Sainz dM, et al. Spanish version of the Northwick Park Neck Pain Questionnaire: reliability and validity. *Clin Exp Rheumatol* 2001;19:41-6.
  111. Pinfold M, Niere KR, O'Leary EF, et al. Validity and internal consistency of a whiplash-specific disability measure. *Spine* 2004;29:263-8.
  112. Bijur PE, Silver W, Gallagher EJ. Reliability of the visual analog scale for measurement of acute pain. *Acad Emerg Med* 2001;8:1153-7.

113. Price DD, McGrath PA, Rafii A, et al. The validation of visual analogue scales as ratio scale measures for chronic and experimental pain. *Pain* 1983;17:45–56.
114. Fejer R, Jordan A, Hartvigsen J. Categorising the severity of neck pain: establishment of cut-points for use in clinical and epidemiological research. *Pain* 2005;119:176–82.
115. Jordan A, Manniche C, Mosdal C, et al. The Copenhagen neck functional disability scale: a study of reliability and validity. *J Manipulative Physiol Ther* 1998;21:520–7.
116. Chiu TT, Lam TH, Hedley AJ. Subjective health measure used on Chinese patients with neck pain in Hong Kong. *Spine* 2001;26:1884–9.
117. Carragee EJ. Validity of self reported history on patients with back and neck pain after motor-vehicle accidents (MVA). *Spine J* 2007;May 22[Epub ahead of print].
118. Carroll LJ, Hogg-Johnson S, van der Velde G, et al. Course and prognostic factors for neck pain in the general population. Results of the bone and joint decade 2000–2010 task force on neck pain and its associated disorders. *Spine* 2008;33(suppl):S74–S81.
119. Carroll LJ, Holm LW, Hogg-Johnson S, et al. Course and prognostic factors for neck pain in whiplash-associated disorders (WAD). Results of the bone and joint decade 2000–2010 task force on neck pain and its associated disorders. *Spine* 2008;33(suppl):S82–S91.
120. Carroll LJ, Hogg-Johnson S, Côté P, et al. Course and prognostic factors for neck pain in workers. Results of the bone and joint decade 2000–2010 task force on neck pain and its associated disorders. *Spine* 2008;33(suppl):S92–S99.
121. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ* 2003;326:41–4.
122. Greenhalgh T. Education and debate: papers that report diagnostic or screening tests. *BMJ* 1997;315:1–12.
123. Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. Getting better but still not good. *JAMA* 1995;274:645–51.
124. Lee CI, Haims AH, Manico EP, et al. Diagnostic CT scans: assessment of patient, physician, and radiologist awareness of radiation dose and possible risks. *Radiology* 2004;231:393–8.
125. Wood B. Lack of Awareness of CT scan Radiation Dose. 2004:65–66.