Diagnostic Accuracy of Upper Cervical Spine Instability Tests: A Systematic Review

Nathan Hutting, Gwendolijn G.M. Scholten-Peeters, Veerle Vijverman, Martin D.M. Keesenberg, Arianne P. Verhagen

Background. Patients with neck pain, headache, torticollis, or neurological signs should be screened carefully for upper cervical spine instability, as these conditions are “red flags” for applying physical therapy interventions. However, little is known about the diagnostic accuracy of upper cervical spine instability tests.

Purpose. The purpose of this study was to evaluate the diagnostic accuracy of upper cervical spine instability screening tests in patients or people who are healthy.

Data Sources. PubMed, CINAHL, EMBASE, and RECAL Legacy databases were searched from their inception through October 2012.

Study Selection. Studies were included that assessed the diagnostic accuracy of upper cervical instability screening tests in patients or people who are healthy and in which sensitivity and specificity were reported or could be calculated using a 2 × 2 table.

Data Extraction and Quality Assessment. Two reviewers independently performed data extraction and the methodological quality assessment using the QUADAS-2.

Data Synthesis. Depending on heterogeneity, statistical pooling was performed. All diagnostic parameters (sensitivity, specificity, predictive values, and likelihood ratios) were recalculated, if possible.

Results. Five studies were included in this systematic review. Statistical pooling was not possible due to clinical and statistical heterogeneity. Specificity of 7 tests was sufficient, but sensitivity varied. Predictive values were variable. Likelihood ratios also were variable, and, in most cases, the confidence intervals were large.

Limitations. The included studies suffered from several biases. None of the studies evaluated upper cervical spine instability tests in patients receiving primary care.

Conclusions. The membranes tests had the best diagnostic accuracy, but their applicability as a test for diagnosing upper cervical spine instability in primary care has yet to be confirmed.
The prevalence of upper cervical spine instability varies among different types of patients. For patients seeking chiropractic care, a prevalence rate of 0.6% was reported. Upper cervical spine instability is associated with inflammatory conditions such as rheumatoid arthritis and ankylosing spondylitis. Trauma and congenital deviation (e.g., Down syndrome) also can cause upper cervical spine instability. Symptoms of upper cervical spine instability are variable, and reported consequences of instability include neck pain, limited mobility, torticollis, and neurological symptoms.

In clinical practice, neck pain, headache, and limited cervical mobility are common reasons to apply cervical manual therapy (physical therapy) interventions. However, when suspected of having cervical instability, these patients should be referred back to the clinician instead of receiving any treatment. According to the guidelines for cervical manual therapy intervention, to minimize the risk of complications, it is recommended to screen patients for upper cervical instability, especially in those at high risk for complications.

Several clinical screening tests are considered able to detect hypermobility and instability of the craniocervical ligaments. In clinical practice, the Sharp-Purser test (SPT), side-bending test, passive upper cervical flexion test, and lateral stability test are commonly used for assessment of upper cervical spine instability. However, these tests are useful only if the reliability and validity are deemed sufficient. Apparently, there is no consensus regarding the use of upper cervical instability testing among practitioners. For example, 34.3% of the members of Musculoskeletal Physiotherapy Australia rarely performed instability screening, 23% never performed instability screening, and 12.3% performed screening tests prior to upper cervical spine manipulative therapy (SMT).

Little is known about the diagnostic accuracy of the upper cervical spine instability tests, and to our knowledge no systematic reviews are available on this topic. Therefore, the aim of this systematic review was to evaluate the diagnostic accuracy in terms of sensitivity, specificity, predictive values, and likelihood ratios of the upper cervical spine instability screening tests in patients or people who are healthy.

Method

Data Sources and Searches
Searches were made in PubMed, CINAHL, EMBASE, and RECAL Legacy databases from their date of inception until the end of October 2012 using MeSH terms (PubMed), thesaurus (EMBASE, CINAHL), and free-text words. Two authors (N.H., M.D.M.K.) independently performed the search. Search terms were related to the diagnostic parameters, SMT, and upper cervical spine instability tests (for details, see Appendix 1).

Study Selection
Studies were included that assessed diagnostic accuracy of upper cervical instability screening tests in patients or people who were healthy and in which sensitivity and specificity were reported or could be calculated using a 2 × 2 table. Two authors (N.H., M.D.M.K.) independently screened the titles and abstracts, followed by a screening of the possibly relevant full-text articles. No restrictions were applied to the year of publication or language. Disagreements were resolved by discussion or arbitration by a third author (G.G.M.S-P.). The references of the included studies also were manually checked for relevant studies possibly missed in the electronic databases.

Data Extraction
Two authors (N.H., G.G.M.S-P.) independently performed data extraction using a standardized form. The following data were extracted: author, year, characteristics of the study population, index test, and reference standard. Data on sensitivity, specificity, predictive values, and likelihood ratios also were extracted. In cases where raw data were provided in the article, we contacted the primary author by e-mail.

Quality Assessment
The methodological quality of the included studies was evaluated with QUADAS-2. This tool is designed to assess the quality of primary diagnostic accuracy studies. Appendix 2 summarizes QUADAS-2 and lists all signaling, risk of bias, and applicability rating items. The QUADAS-2 tool consists of 4 key domains that discuss bias associated with patient selection, index test, reference standard, flow of patients through the study, timing of the index test, and reference standard (flow and timing).

Two authors (N.H., M.D.M.K.) independently scored the items as low, high, or unclear. Any differences in assessment were discussed and, when disagreement persisted, were solved by a third author (G.G.M.S-P.).

Data Synthesis and Analysis
Agreement of methodological quality between reviewers was calculated and quantified by kappa (κ). Kappa was categorized as poor (≤0.00), slight (0.00–0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80) or almost perfect agreement (0.81–1.00). Prior to statistical pooling, clinical and statistical sources of heterogeneity were assessed. In case of heterogeneity (I² > 40%), a descriptive analysis was performed.
performed. In case the authors did not report diagnostic accuracy, the raw data were used to calculate diagnostic accuracy data using a 2 x 2 table. We added 0.5 to all fields in case of an empty field in the 2 x 2 table.15

Sensitivity, specificity, and predictive values of at least 80% were considered to be sufficient. Also, a positive likelihood ratio of >10 and a negative likelihood ratio of <0.1 were considered to be sufficient.16

**Results**

**Study Selection**

The search identified 773 potential citations. Figure 1 presents the flowchart of the study selection process. After removal of double citations and excluding articles not fulfilling the inclusion criteria based on title and abstract, 4 studies17–20 were retrieved for full-text assessment, and 3 of them were included.18–20 One study was excluded because it did not provide sensitivity and specificity values or raw data to calculate these values.17 Reference checking provided 2 additional studies.21,22 Finally, 5 studies were found eligible for inclusion in this review.18–22 For 1 study, we contacted the primary author to provide raw data or sensitivity and specificity data.19 We received only the likelihood ratios of that study.

**Description of the Studies**

Details on the 5 included studies are presented in Table 1; of these studies, 4 were published between 1969 and 1999.18–22 The average number of people included in the studies was 115.5 (range = 31–123). Only 3 studies reported the average age of the participants.19,20,22 Four studies included patients with rheumatoid arthritis.18,19,21,22 In 3 studies, the diagnosis of rheumatoid arthritis was based on the criteria of the American Rheumatism Association,18,21,22 and in 1 study, the criteria were unclear.19 One study included patients with whiplash-associated disorders (WAD) diagnosed by a physician.20 Two studies also included a healthy population.20,21 Two studies included patients visiting a hospital,18,22 and 1 study included patients who visited the physical therapy department of the hospital.21 In another study, the setting was unclear.19 Prevalence of instability in the included studies varied between 0.07 and 0.44.

**Index Tests**

A total of 7 tests were evaluated: SPT,18,19,21,22 clunking test,21 palate sign,21 alar ligament test,20 transverse ligament test,20 tectorial membrane test,20 and posterior atlanto-occipital membrane test20 (Tab. 1). Kappa values of the interobserver reliability of the SPT varied between .06 and .67, and the SPT was not considered to be reliable.19,23 Of the other tests, no information about the reliability was available.

**Reference Tests**

Four studies compared 1 or more instability tests with roentgenradiation (x-ray films) as the reference test.18,19,21,22 On x-ray films, the atlas dens interval (ADI) was measured on a lateral scan. In 1 study, the ADI was measured in cervical flexion and extension,18 and in 3 other studies, the ADI was measured only in flexion.19,21,22 In 4 studies, an ADI ≥3 mm was classified as abnormal.18,19,21,22 One study also calculated sensitivity and specificity values when the ADI was ≥4 mm.18 One study compared the alar ligament test, transverse ligament test, tectorial membrane test, and poste-
rior membrane test with magnetic resonance imaging (MRI).

**Methodological Quality**
The interobserver reliability of the methodological quality assessment with QUADAS-2 was kappa = 0.58 (confidence interval = 0.32–0.83) and is considered to be moderate agreement. The overall agreement was 80%. Disagreements appeared in 4 items and were mainly due to reading errors or differences in interpretation of the items. All disagreements were solved during a consensus meeting.

Overall, the studies suffered from various types of potential bias, and about 50% of the items were scored as unclear. This was particularly the case regarding patient selection, the index test, and the reference standard. Applicability concerns also were present regarding the reference standard. The results of the methodological quality assessment using QUADAS-2 are presented in Table 2.

**Diagnostic Accuracy**
Figure 2 presents forest plots of the sensitivity and specificity of the included studies. Data on the diagnostic accuracy in terms of sensitivity, specificity, predictive values, and likelihood ratios are presented in Table 3. We calculated diagnostic accuracy data from the raw data presented in 3 studies. Statistical pooling of 3 studies investigating the SPT in which raw data were provided was not possible due to statistical heterogeneity (sensitivity I² = 87.8%, specificity I² = 91.6%).

Sensitivity ranged from 0.19 to 0.96, and specificity ranged from 0.71 to 1.00. Positive predictive values ranged from 0.11 to 1.00, and negative predictive values ranged from 0.56 to 0.99. The positive predictive values of the clunking test and the palate sign were considered to be insufficient, and the positive predictive values of the SPT varied. Negative predictive values of the clunking test and palate sign were considered sufficient. Negative predictive values of the SPT were vari-

### Table 1.
Information on the 5 Included Studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Population (n)</th>
<th>Mean Age (Range) (y)</th>
<th>Index Test</th>
<th>Reference Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forrester and Barlas, 1999</td>
<td>RA (31)</td>
<td>57.5 (29–74)</td>
<td>Sharp-Purser test</td>
<td>Radiographs (lateral, flexion)</td>
</tr>
<tr>
<td>Kaale et al, 2008</td>
<td>WAD (92)</td>
<td>39.1</td>
<td>Alar ligament test</td>
<td>MRI</td>
</tr>
<tr>
<td></td>
<td>HC (30)</td>
<td>45.3</td>
<td>Transverse ligament test</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Tectorial membrane test</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Atlanto-occipital membrane test</td>
<td></td>
</tr>
<tr>
<td>Mathews, 1969</td>
<td>RA (76)</td>
<td>54.2 (21–75)</td>
<td>Sharp-Purser test</td>
<td>Radiographs (lateral, flexion)</td>
</tr>
<tr>
<td>Stevens et al, 1971</td>
<td>RA (100)</td>
<td></td>
<td>Clunking test</td>
<td>Radiographs (lateral, flexion)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Palate sign</td>
<td></td>
</tr>
<tr>
<td>Uitvlugt and Indenbaum, 1988</td>
<td>RA (123)</td>
<td></td>
<td>Sharp-Purser test</td>
<td>Radiographs (lateral, flexion, extension)</td>
</tr>
</tbody>
</table>

RA = rheumatoid arthritis, WAD = whiplash-associated disorder, HC = healthy controls, MRI = magnetic resonance imaging.

### Table 2.
Assessment of Methodological Quality With QUADAS-2

<table>
<thead>
<tr>
<th>Study</th>
<th>Risk of Bias</th>
<th>Applicability Concerns</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Patient Selection</td>
<td>Index Test</td>
</tr>
<tr>
<td>Forrester and Barlas, 1999</td>
<td>↑</td>
<td>↑</td>
</tr>
<tr>
<td>Kaale et al, 2008</td>
<td>↑</td>
<td>↑</td>
</tr>
<tr>
<td>Mathews, 1969</td>
<td>↓</td>
<td>↑</td>
</tr>
<tr>
<td>Stevens, 1971</td>
<td>↑</td>
<td>↑</td>
</tr>
<tr>
<td>Uitvlugt and Indenbaum, 1988</td>
<td>↑</td>
<td>↑</td>
</tr>
</tbody>
</table>

*↑ − Low risk, ↓ − high risk, ? − unclear risk.*
able. Generally, the predictive value of the alar end transverse ligament test and the tectorial and atlanto-occipital membrane tests was sufficient.

Positive likelihood ratios ranged from 0.67 to 185.6, and, in most cases, the confidence intervals were very large. Negative likelihood ratios ranged from 0.04 to 1.13. The positive likelihood ratios of the alar ligament test and the transverse ligament test were considered sufficient. The tectorial membrane test and atlanto-occipital membrane test showed sufficient positive likelihood ratios as well as sufficient negative likelihood ratios.

Discussion

Main Findings

Diagnostic accuracy data of the SPT, the only test that was evaluated more

Figure 2.

Sensitivity and specificity of the included studies. TP = true-positive findings, FP = false-positive findings, FN = false-negative findings, TN = true-negative findings, 95% CI = 95% confidence interval.
Diagnostic Accuracy of Upper Cervical Spine Instability Tests

than once, were generally not sufficient. Specificity of almost all the tests was sufficient, which means that the tests can be used to rule in patients with upper cervical spine instability. The confidence intervals of the likelihood ratios were generally extremely wide, indicating low precision. In most studies, the methodological quality was poor to moderate, mainly because most of the items were scored “unclear” due to lack of information.

Clinical Implications
In pretreatment screening procedures, we aim to identify patients with upper cervical spine instability to omit treatment interventions and to refer patients back to the appropriate medical professionals. It is important to prevent false-negative results because, in that case, patients incorrectly will receive treatment of the upper cervical spine. Therefore, the sensitivity of these tests needs to be high.24 Sensitivity of most tests is insufficient for detecting upper cervical spine instability; therefore, the clinical value of these tests is low.

Sufficient specificity indicates the ability of a test to prevent false-positive results. Specificity is less important than sensitivity, as a false-positive result of the test is not potentially harmful for the patient.24 However, a potential effective treatment might be withheld from the patient.

Likelihood ratios are alternative statistics to express diagnostic accuracy.25 Likelihood ratios ≥10 and ≤0.1 are assumed to provide strong evidence to rule in or rule out a diagnosis.16 The positive likelihood ratios of the tectorial membrane test and atlanto-occipital membrane test were sufficient. Also, the negative likelihood ratios of the tectorial membrane test and atlanto-occipital membrane test were 0.04 and 0.06, respectively, indicating that these tests are able to rule out cervical spinal instability. However, the confidence intervals were wide (possibly due to small sample sizes), indicating a lack of precision. Moreover, in the studies included in this review, these tests were evaluated only once. Nevertheless, these tests show most promise and should be evaluated in future studies to establish their clinical value.

To be of value for clinical practice, besides high sensitivity and specificity, diagnostic tests should have acceptable reliability. The upper cervical flexion test showed acceptable intrarater and interrater reliability in children with Down syndrome: in 3 of 4 investigators, the intrarater reliability was significant, whereas in 4 of 6 pairs of investigators, the interobserver reliability was significant.25 The SPT and the lateral displacement test were not considered to be reliable.19,25

Strengths and Weaknesses of the Study
This is the first systematic review to assess the diagnostic accuracy of upper cervical instability tests. Unfortunately, most tests were evaluated only once, and none of the studies evaluated the diagnostic accuracy of the upper cervical spine instability tests as a pretreatment screening test for applying in primary care. These findings imply limited generalizability of our results to clinical practice. Also, 4 of the 5 studies included patients with rheumatoid arthritis in a hospital setting, limiting generalizability to primary care; in hospitalized populations, the prevalence of upper cervical spine instability is higher than in patients usually seen in primary care.20,26 We also found wide confidence intervals for likelihood ratios, indicating a lack of precision,27 mainly due to the absence of false-positive findings.

The prevalence of the condition (upper cervical spine instability), as indicated by the reference test in the included studies, was variable. Moreover, it was a variable in the same study investigating several tests.18,20,21 Because the same population was used, it is remarkable that the prevalence varied.

Overall, most of the included studies suffered from a lack of adequate information, leading to the possibility of various types of bias, which hampers the possibility to draw firm conclusions. A reason for this lack of information might be that most of the studies are relatively old. Agreement on interobserver reliability for scoring with QUADAS-213 was moderate, although the 95% confidence interval was large, with the lower limit categorized as “fair” and the upper limit categorized as “almost perfect.” Lack of precision in the methodological quality assessment was due to problems with interpretation of the items, reading errors, and the low number of articles included in the review. The reviewers had difficulties in scoring the QUADAS-213 because of unclear reporting. In the future, reporting of studies may be improved by using the STARD guidelines.28

Four studies used the ADI on radiographs to identify instability. Although radiographs are often used in the screening for upper cervical spine instability, this method has some limitations.29 Because the test-retest reliability of radiographs is reported to be unsatisfactory,29 this is not the perfect reference standard. According to the American Association of Radiologists, the reference value for diagnosing instability is an ADI >2.5 to 3 mm for adults and an ADI >4.5 to 5 mm for children.25 About 95% of people who are healthy have an ADI of 0.3 to 1.8 mm in flexion, of 0.4 to 0.2 mm in neutral position, and of 0.3 to 2.2 mm in
Because the ADI is small, assessors need to be aware of possible errors in measurement. Some authors have suggested that the diameter of the spinal canal is a better diagnostic criterion. Other authors suggest that computed tomography and MRI are preferable reference standards for diagnosing upper cervical spine instability. Kaale et al. compared loss of collagen integrity as judged by MRI (grade 0–3) with the degree of increased mobility as judged by clinical examination (category 0–3); they explored whether lesions of a specific neck structure affected the passive mobility of that structure. Until now, MRI has been regarded as the best tool to visualize the collagen integrity of soft tissue structures and is considered to represent the “gold standard.” The prevalence of grade 2 to 3 ligament high-signal intensity (on alar and transverse ligaments) in patients with WAD was similar to the prevalence in non-injured patients with chronic neck pain. These findings indicate possible physiologic ligament variants with loose connective tissue.

**Conclusions**

Overall, the studies suffered from various types of potential bias, and the sensitivity varied. Therefore, we cannot make overall conclusions. The anterior atlanto-axial membrane test is a reasonable test to use in the clinical setting. The specificity of the anterior atlanto-axial membrane test is known to be lower than that of the posterior atlanto-occipital membrane test. However, the specificity of the anterior atlanto-axial membrane test is still high enough to be considered a valid test in this clinical setting.
Diagnostic Accuracy of Upper Cervical Spine Instability Tests

and the tectorial membrane test showed the best diagnostic accuracy in patients with WAD; however, the role of these tests in diagnosing upper cervical spine instability in pretreatment procedures has yet to be confirmed.

Mr Hutting, Ms Vijverman, and Dr Scholten-Peeters provided concept/idea/research design and project management. Mr Hutting and Mr Keessenig provided data collection. Mr Hutting, Dr Scholten-Peeters, Ms Vijverman, and Dr Verhagen provided writing and data analysis. All authors have read and approved the final manuscript.


References

10 Hing WA, Reid DA, Monaghan M. Manipulation of the cervical spine. Man Ther. 2003;8:2–9.
Appendix 1.

Search Terms

**PubMed**
(atlanto-axial OR atlantoaxial OR atlanto axial OR craniovertebral) AND (instability OR sub-luxation OR subluxation) AND (validity OR “sensitivity and specificity”[MeSH] OR test OR tests) (upper cervical instability OR upper cervical subluxation OR cervical instability OR cervical subluxation) AND (validity OR “sensitivity and specificity”[MeSH] AND (test OR tests)) (validity OR “sensitivity and specificity”[MeSH]) AND “Manipulation, Spinal”[MeSH] AND (test OR tests) sharp purser OR sharp purser

**CINAHL**
(atlanto-axial or atlantoaxial or atlanto axial or craniovertebral) and (instability or subluxation or sub-luxation) and (validity or “sensitivity-and-specificity” or test or tests) (premanipulative or pre-manipulative) and (test or tests) and (validity or “sensitivity-and-specificity”) (upper cervical instability or upper cervical subluxation or cervical instability or cervical subluxation) and (validity or “sensitivity-and-specificity” or test or tests) (validity or “sensitivity-and-specificity”) and cervical manipulation and (test or tests) sharp purser OR sharp purser

**EMBASE**
(atlanto-axial or atlantoaxial or atlanto axial or craniovertebral) and (instability or subluxation or sub-luxation) and (validity or “sensitivity-and-specificity” or test or tests) (premanipulative or pre-manipulative) and (test or tests) and (validity or “sensitivity-and-specificity”) (upper cervical instability or upper cervical subluxation or cervical instability or cervical subluxation) and (validity or “sensitivity-and-specificity” or test or tests) (validity or “sensitivity-and-specificity”) and manipulation and (test or tests) sharp purser or sharp-purser

**RECAL**
atlantoaxial instability
atlantoaxial subluxation
atlanto-axial instability
atlanto-axial subluxation
craniovertebral instability
craniovertebral subluxation
upper cervical subluxation
upper cervical instability
sharp purser
sharp-purser
Appendix 2.
Overview of QUADAS-2 and Lists of All Signaling, Risk of Bias, and Applicability Rating Questions

<table>
<thead>
<tr>
<th>Domain</th>
<th>Patient Selection</th>
<th>Index Test</th>
<th>Reference Standard</th>
<th>Flow and Timing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>Describe methods of patient selection.</td>
<td>Describe the index test and how it was conducted and interpreted.</td>
<td>Describe the reference standard and how it was conducted and interpreted.</td>
<td>Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the (\times) table (refer to flow diagram). Describe the time interval and any interventions between index test(s) and reference standard.</td>
</tr>
<tr>
<td>Signaling questions</td>
<td>Was a consecutive or random sample of patients enrolled?</td>
<td>Were the index test results interpreted without knowledge of the results of the reference standard?</td>
<td>Is the reference standard likely to correctly classify the target condition?</td>
<td>Was there an appropriate interval between index test(s) and reference standard?</td>
</tr>
<tr>
<td>(yes/no/unclear)</td>
<td>Was a case-control design avoided?</td>
<td>If a threshold was used, was it prespecified?</td>
<td>Were the reference standard results interpreted without knowledge of the results of the index test?</td>
<td>Did all patients receive a reference standard?</td>
</tr>
<tr>
<td></td>
<td>Did the study avoid inappropriate exclusions?</td>
<td></td>
<td></td>
<td>Did all patients receive the same reference standard?</td>
</tr>
<tr>
<td>Risk of bias: high/low/unclear</td>
<td>Could the selection of patients have introduced bias?</td>
<td>Could the conduct or interpretation of the index test have introduced bias?</td>
<td>Could the reference standard, its conduct, or its interpretation have introduced bias?</td>
<td>Were all patients included in the analysis?</td>
</tr>
<tr>
<td>Concerns regarding applicability: high/low/unclear</td>
<td>Are there concerns that the included patients do not match the review question?</td>
<td>Are there concerns that the index test, its conduct, or interpretation differs from the review question?</td>
<td>Are there concerns that the target condition as defined by the reference standard does not match the review question?</td>
<td>Could the patient flow have introduced bias?</td>
</tr>
</tbody>
</table>

\* Reproduced with permission from: http://www.bris.ac.uk/quadas/.
Diagnostic Accuracy of Upper Cervical Spine Instability Tests: A Systematic Review
Nathan Hutting, Gwendolijne G.M. Scholten-Peeters, Veerle Vijverman, Martin D.M. Keesenberg and Arianne P. Verhagen

PHYS THER. 2013; 93:1686-1695.
Originally published online July 25, 2013

References
This article cites 29 articles, 10 of which you can access for free at:
http://ptjournal.apta.org/content/93/12/1686#BIBL

Subscription Information
http://ptjournal.apta.org/subscriptions/

Permissions and Reprints
http://ptjournal.apta.org/site/misc/terms.xhtml

Information for Authors
http://ptjournal.apta.org/site/misc/ifora.xhtml