

# Physical Therapy

Journal of the American Physical Therapy Association



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

Nathan Hutting, Gwendolijne G.M. Scholten-Peeters,  
Veerle Vijverman, Martin D.M. Keessenberg and Arianne  
P. Verhagen

*PHYS THER.* 2013; 93:1686-1695.

Originally published online July 25, 2013

doi: 10.2522/ptj.20130186

The online version of this article, along with updated information and services, can be found online at: <http://ptjournal.apta.org/content/93/12/1686>

---

### **Collections**

This article, along with others on similar topics, appears in the following collection(s):

[Diagnosis/Prognosis: Other](#)  
[Injuries and Conditions: Neck](#)  
[Systematic Reviews/Meta-analyses](#)  
[Tests and Measurements](#)

### **e-Letters**

To submit an e-Letter on this article, click [here](#) or click on "Submit a response" in the right-hand menu under "Responses" in the online version of this article.

### **E-mail alerts**

Sign up [here](#) to receive free e-mail alerts

---

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

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

N. Hutting, PT, MT, MSc, Department of Manual Therapy, Faculty of Medicine and Pharmacology, Vrije Universiteit Brussel, Brussels, Belgium, and HAN University of Applied Sciences, Nijmegen, the Netherlands. Mailing address: Pieter Breughelhof 2, 5121 EW Rijen, the Netherlands. Address all correspondence to Mr Hutting at: [nathan@nathanhutting.com](mailto:nathan@nathanhutting.com).

G.G.M. Scholten-Peeters, PT, MT, PhD, Department of Manual Therapy, Faculty of Medicine and Pharmacology, Vrije Universiteit Brussel.

V. Vijverman, PT, MT, MSc, Department of Manual Therapy, Faculty of Medicine and Pharmacology, Vrije Universiteit Brussel.

M.D.M. Keessenberg, PT, MSc, Center for Physical Therapy & Science, Corpus Mentis, Leiden, the Netherlands.

A.P. Verhagen, PT, PhD, Department of General Practice, Erasmus University Medical Center, Rotterdam, the Netherlands.

[Hutting N, Scholten-Peeters GGM, Vijverman V, et al. Diagnostic accuracy of upper cervical spine instability tests: a systematic review. *Phys Ther*. 2013;93:1686–1695.]

© 2013 American Physical Therapy Association

Published Ahead of Print:  
July 25, 2013

Accepted: July 19, 2013

Submitted: January 7, 2013

**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  $2 \times 2$  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.



Post a Rapid Response to  
this article at:  
[ptjournal.apta.org](http://ptjournal.apta.org)

The prevalence of upper cervical spine instability varies among different types of patients.<sup>1</sup> For patients seeking chiropractic care, a prevalence rate of 0.6% was reported.<sup>1</sup> Upper cervical spine instability is associated with inflammatory conditions such as rheumatoid arthritis and ankylosing spondylitis.<sup>2,3</sup> Trauma and congenital deviation (eg, Down syndrome) also can cause upper cervical spine instability.<sup>2</sup> Symptoms of upper cervical spine instability are variable, and reported consequences of instability include neck pain, limited mobility, torticollis, and neurological symptoms.<sup>4</sup>

In clinical practice, neck pain, headache, and limited cervical mobility are common reasons to apply cervical manual therapy (physical therapy) interventions.<sup>5,6</sup> 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,<sup>5,7-11</sup> especially in those at high risk for complications.<sup>8</sup>

Several clinical screening tests are considered able to detect hypermobility and instability of the cranio-cervical ligaments.<sup>3</sup> 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.<sup>12</sup> 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).<sup>12</sup>

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  $2 \times 2$  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 stud-

ies 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 case no raw data were provided in the article, we contacted the primary author by e-mail.

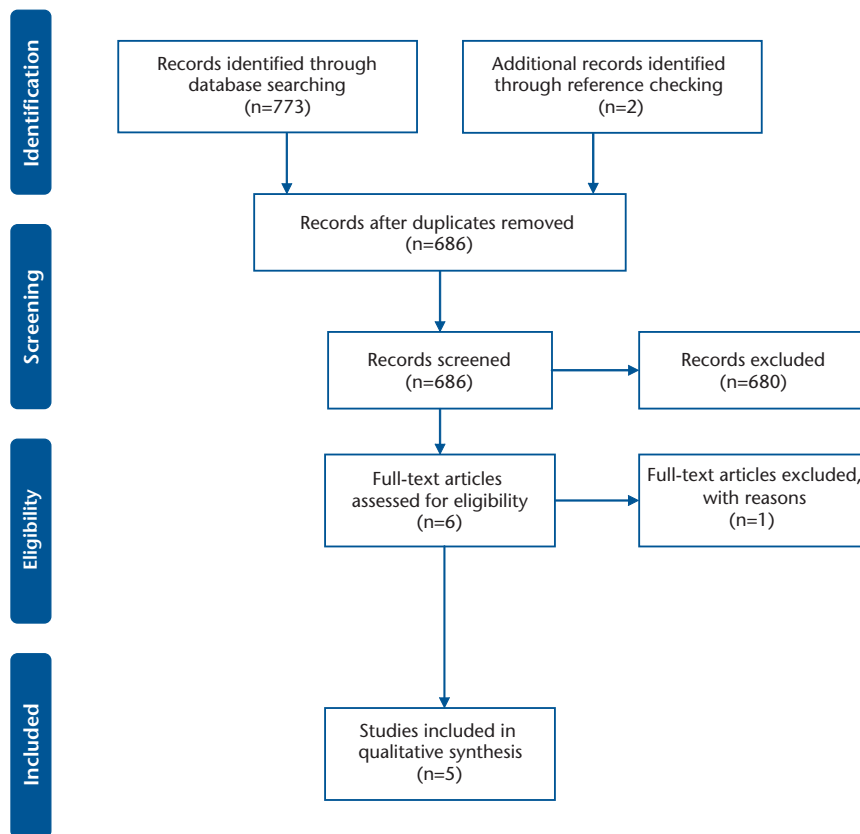
### Quality Assessment

The methodological quality of the included studies was evaluated with QUADAS-2.<sup>13</sup> This tool is designed to assess the quality of primary diagnostic accuracy studies.<sup>13</sup> 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).<sup>13</sup>

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 ( $\kappa$ ). Kappa was categorized as poor ( $\leq .00$ ), slight ( $.00-.20$ ), fair ( $.21-.40$ ), moderate ( $.41-.60$ ), substantial ( $.61-.80$ ) or almost perfect agreement ( $.81-1.00$ ).<sup>14</sup> Prior to statistical pooling, clinical and statistical sources of heterogeneity were assessed. In case of heterogeneity ( $I^2 > 40\%$ ), a descriptive analysis was



**Figure 1.**  
Flow diagram of the included studies.

performed. In case the authors did not report diagnostic accuracy, the raw data were used to calculate diagnostic accuracy data using a  $2 \times 2$  table. We added 0.5 to all fields in case of an empty field in the  $2 \times 2$  table.<sup>15</sup>

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.<sup>16</sup>

## Results

### Study Selection

The search identified 773 potential citations. Figure 1 presents the flow-chart of the study selection process. After removal of double citations and excluding articles not fulfilling the inclusion criteria based on title

and abstract, 4 studies<sup>17–20</sup> were retrieved for full-text assessment, and 3 of them were included.<sup>18–20</sup> One study was excluded because it did not provide sensitivity and specificity values or raw data to calculate these values.<sup>17</sup> Reference checking provided 2 additional studies.<sup>21,22</sup> Finally, 5 studies were found eligible for inclusion in this review.<sup>18–22</sup> For 1 study, we contacted the primary author to provide raw data or sensitivity and specificity data.<sup>19</sup> 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.<sup>18–22</sup> 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.<sup>19,20,22</sup> Four studies included patients with rheumatoid arthritis.<sup>18,19,21,22</sup> In 3 studies, the diagnosis of rheumatoid arthritis was based on the criteria of the American Rheumatism Association,<sup>18,21,22</sup> and in 1 study, the criteria were unclear.<sup>19</sup> One study included patients with whiplash-associated disorders (WAD) diagnosed by a physician.<sup>20</sup> Two studies also included a healthy population.<sup>20,21</sup> Two studies included patients visiting a hospital,<sup>18,22</sup> and 1 study included patients who visited the physical therapy department of the hospital.<sup>21</sup> In another study, the setting was unclear.<sup>19</sup> Prevalence of instability in the included studies varied between 0.07 and 0.44.

### Index Tests

A total of 7 tests were evaluated: SPT,<sup>18,19,21,22</sup> clunking test,<sup>21</sup> palate sign,<sup>21</sup> alar ligament test,<sup>20</sup> transverse ligament test,<sup>20</sup> tectorial membrane test,<sup>20</sup> and posterior atlanto-occipital membrane test<sup>20</sup> (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.<sup>19,23</sup> Of the other tests, no information about the reliability was available.

### Reference Tests

Four studies compared 1 or more instability tests with roentgen-radiation (x-ray films) as the reference test.<sup>18,19,21,22</sup> 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,<sup>18</sup> and in 3 other studies, the ADI was measured only in flexion.<sup>19,21,22</sup> In 4 studies, an  $ADI \geq 3$  mm was classified as abnormal.<sup>18,19,21,22</sup> One study also calculated sensitivity and specificity values when the ADI was  $\geq 4$  mm.<sup>18</sup> One study compared the alar ligament test, transverse ligament test, tectorial membrane test, and poste-

**Table 1.**Information on the 5 Included Studies<sup>a</sup>

Study	Population (n)	Mean Age (Range) (y)	Index Test	Reference Test
Forrester and Barlas, <sup>19</sup> 1999	RA (31)	57.5 (29–74)	Sharp-Purser test	Radiographs (lateral, flexion)
Kaale et al, <sup>20</sup> 2008	WAD (92) HC (30)	39.1 45.3	Alar ligament test Transverse ligament test Tectorial membrane test Atlanto-occipital membrane test	MRI
Mathews, <sup>21</sup> 1969	RA (76) HC (28)		Sharp-Purser test Clunking test Palate sign	Radiographs (lateral, flexion)
Stevens et al, <sup>22</sup> 1971	RA (100)	54.2 (21–75)	Sharp-Purser test	Radiographs (lateral, flexion)
Uitvlugt and Indenbaum, <sup>18</sup> 1988	RA (123)		Sharp-Purser test	Radiographs (lateral, flexion, extension)

<sup>a</sup> RA=rheumatoid arthritis, WAD=whiplash-associated disorder, HC=healthy controls, MRI=magnetic resonance imaging.

rior membrane test with magnetic resonance imaging (MRI).<sup>20</sup>

### Methodological Quality

The interobserver reliability of the methodological quality assessment with QUADAS-2 was kappa=.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.<sup>18,21,22</sup> Most of the tests were assessed in only 1 study. Only the SPT was assessed multiple

times. Statistical pooling of 3 studies investigating the SPT in which raw data were provided was not possible due to statistical heterogeneity (sensitivity  $I^2=87.8\%$ , specificity  $I^2=91.6\%$ ).

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 2.**Assessment of Methodological Quality With QUADAS-2<sup>a</sup>

Study	Risk of Bias				Applicability Concerns		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Forrester and Barlas, <sup>19</sup> 1999	?	?	?	?	↑	↑	?
Kaale et al, <sup>20</sup> 2008	↑	↑	↑	↑	↑	?	↑
Mathews, <sup>21</sup> 1969	↓	↑	?	↑	↑	↑	?
Stevens, <sup>22</sup> 1971	↑	?	?	↑	↑	↑	?
Uitvlugt and Indenbaum, <sup>18</sup> 1988	↑	?	?	?	↑	↑	?

<sup>a</sup> ↑ =Low risk, ↓ =high risk, ?=unclear risk.



## Diagnostic Accuracy of Upper Cervical Spine Instability Tests

### Sharp-Purser Test

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Mathews <sup>21</sup>	6	13	25	32	0.19 (0.07-0.37)	0.71 (0.56-0.84)		
Stevens et al <sup>22</sup>	16	1	20	63	0.44 (0.28-0.62)	0.98 (0.92-1.00)		
Uitvlugt and Indenbaum <sup>18</sup>	22	4	10	87	0.69 (0.50-0.84)	0.96 (0.89-0.99)		

### Right Alar Ligament Test

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Kaale et al <sup>20</sup>	37	0	17	68	0.69 (0.54-0.80)	1.00 (0.95-1.00)		

### Left Alar Ligament Test

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Kaale et al <sup>20</sup>	39	3	15	65	0.72 (0.58-0.84)	0.96 (0.88-0.99)		

### Transverse Ligament Test

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Kaale et al <sup>20</sup>	28	1	15	78	0.65 (0.49-0.79)	0.99 (0.93-1.00)		

### Tectorial Membrane Test

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Kaale et al <sup>20</sup>	15	1	1	104	0.94 (0.70-1.00)	0.99 (0.95-1.00)		

### Atlantoaxial Membrane Test

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Kaale et al <sup>20</sup>	22	0	1	98	0.96 (0.78-1.00)	1.00 (0.96-1.00)		

### Clunking Test

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Mathews <sup>21</sup>	2	17	4	53	0.33 (0.04-0.78)	0.76 (0.64-0.85)		

### Palate Sign

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Mathews <sup>21</sup>	5	14	2	55	0.71 (0.29-0.96)	0.80 (0.68-0.88)		

### Sharp-Purser Test (Atlas Dens Interval 4 mm)

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Uitvlugt and Indenbaum <sup>18</sup>	22	4	3	94	0.88 (0.69-0.97)	0.96 (0.90-0.99)		

**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.

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

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.<sup>24</sup> 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.<sup>24</sup> However, a potential effective treatment might be withheld from the patient.

Likelihood ratios are alternative statistics to express diagnostic accuracy.<sup>25</sup> Likelihood ratios  $\geq 10$  and  $\leq 0.1$  are assumed to provide strong evidence to rule in or rule out a diagnosis.<sup>16</sup> 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 intra-observer reliability was significant, whereas in 4 of 6 pairs of investigators, the interobserver reliability was significant.<sup>23</sup> The SPT and the lateral displacement test were not considered to be reliable.<sup>19,23</sup>

### 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.<sup>20,26</sup> We also found wide confidence intervals for likelihood ratios, indicating a lack of precision,<sup>27</sup> 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.<sup>18,20,21</sup> 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-2<sup>13</sup> 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-2<sup>13</sup> because of unclear reporting. In the future, reporting of studies may be improved by using the STARD guidelines.<sup>28</sup>

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.<sup>29</sup> Because the test-retest reliability of radiographs is reported to be unsatisfactory,<sup>29</sup> 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.<sup>23</sup> 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

**Table 3.**  
Diagnostic Accuracy of the Included Studies<sup>a</sup>

Study	Test	Prevalence	Sensitivity (95% CI)	Specificity (95% CI)	+PV (95% CI)	-PV (95% CI)	+LR (95% CI)	-LR (95% CI)
Forrester and Barlas, <sup>19</sup> 1999	SPT	?	0.43	0.77	0.34	0.89	1.88 (1.14–3.09)	0.74 (0.53–1.02)
Kaale et al, <sup>20</sup> 2008	ALR	0.44	0.69 (0.56–0.81)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	0.80 (0.71–0.89)	94.09 (5.91–1,498.02)	0.33 (0.22–0.47)
	ALL	0.44	0.72 (0.60–0.84)	0.96 (0.91–1.00)	0.93 (0.85–1.01)	0.81 (0.73–0.90)	16.37 (5.35–50.10)	0.29 (0.19–0.48)
	TL	0.35	0.65 (0.51–0.79)	0.99 (0.96–1.01)	0.97 (0.90–1.03)	0.84 (0.76–0.91)	51.44 (7.25–365.12)	0.35 (0.12–0.53)
	TM	0.13	0.94 (0.82–1.06)	0.99 (0.97–1.01)	0.94 (0.82–1.06)	0.99 (0.97–1.01)	98.44 (13.94–695.15)	0.06 (0.01–0.42)
	AOM	0.19	0.96 (0.87–1.04)	1.00 (1.00–1.00)	1.00 (1.00–1.00)	0.99 (0.97–1.01)	185.6 (11.67–2,952.63)	0.04 (0.01–0.30)
Mathews, <sup>21</sup> 1969	SPT	0.41	0.19 (0.08–0.38)	0.71 (0.55–0.83)	0.32 (0.14–0.57)	0.56 (0.42–0.70)	0.67 (0.29–1.57)	1.13 (0.94–1.38)
	CL	0.08	0.33 (0.06–0.76)	0.76 (0.64–0.85)	0.11 (0.02–0.35)	0.93 (0.82–0.98)	1.37 (0.41–4.58)	0.88 (0.50–1.56)
	PS	0.09	0.71 (0.30–0.95)	0.80 (0.68–0.88)	0.26 (0.10–0.51)	0.96 (0.87–0.99)	3.52 (1.82–6.82)	0.36 (0.11–1.16)
Stevens et al, <sup>22</sup> 1971	SPT	0.36	0.44 (0.28–0.62)	0.98 (0.90–1.00)	0.94 (0.70–1.00)	0.76 (0.65–0.84)	28.44 (3.93–205.72)	0.56 (0.42–0.76)
Uitvlugt and Indenbaum, <sup>18</sup> 1988	SPT	0.26	0.69 (0.50–0.83)	0.96 (0.89–0.99)	0.85 (0.64–0.95)	0.90 (0.81–0.95)	15.64 (5.83–41.94)	0.33 (0.20–0.55)
	SPT*	0.20	0.88 (0.68–0.97)	0.96 (0.89–0.99)	0.85 (0.64–0.95)	0.97 (0.91–0.99)	21.56 (8.17–56.91)	0.13 (0.04–0.36)

<sup>a</sup> +PV=positive predictive value, -PV=negative predictive value, +LR=positive likelihood ratio, -LR=negative likelihood ratio, SPT=Sharp-Purser Test, CL=clunking test, PS=palate sign, ALR=right alar ligament, ALL=left alar ligament, TL=transverse ligament, TM=tectorial membrane, AOM=atlanto-occipital membrane. \*With atlas dens interval >4 mm.

extension.<sup>4</sup> Because the ADI is small, assessors need to be aware of possible errors in measurement.<sup>4</sup> Some authors have suggested that the diameter of the spinal canal is a better diagnostic criterion.<sup>4,30</sup> Other authors suggest that computed tomography and MRI are preferable reference standards for diagnosing upper cervical spine instability.<sup>4</sup> Kaale et al<sup>20</sup> 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<sup>31</sup> and is considered to represent the “gold standard.”<sup>20</sup> 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.<sup>20</sup> These findings indicate possible physiologic ligament variants with loose connective tissue.<sup>32,33</sup> One study<sup>20</sup> examined whether results from a clinical test (alar ligament test, transverse ligament test, tectorial membrane test, and posterior atlanto-occipital membrane test) corresponded with signs of physical injuries, as judged by MRI.<sup>20</sup> In case soft tissue structures are injured, an abnormally increased mobility in this region can be expected.<sup>20</sup> Unfortunately, the authors did not classify increased mobility (grades 2 and 3) as instability.

## Conclusions

Overall, the studies suffered from various types of potential bias, and the sensitivity varied. Therefore, we conclude that screening for upper cervical instability cannot be done accurately at the moment. The atlanto-axial membrane test



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 Keesenberg 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.

DOI: 10.2522/ptj.20130186

## References

- Beck RW, Holt KR, Fox MA, Hurtgen-Grace KL. Radiographic anomalies that may alter chiropractic intervention strategies found in a New Zealand population. *J Manipulative Physiol Ther.* 2004;27:554-559.
- Swinkels RA, Beeton K, Alltree J. Pathogenesis of upper cervical instability. *Man Ther.* 1996;1:127-132.
- Aspinall W. Clinical testing for the craniovertebral hypermobility syndrome. *J Orthop Sports Phys Ther.* 1990;12:47-54.
- Swinkels RA, Oostendorp RA. Upper cervical instability: fact or fiction? *J Manipulative Physiol Ther.* 1996;19:185-194.
- Rivett DA, Thomas L, Bolton P. Pre-manipulative testing: where do we go from here? *New Zealand Journal of Physiotherapy.* 2005;33:78-84.
- Colloca CJ, Pickar JG, Slosberg M. Special focus on spinal manipulation. *J Electromyogr Kinesiol.* 2012;22:629-631.
- Refshauge KM, Parry S, Shirley D, et al. Professional responsibility in relation to cervical spine manipulation. *Aust J Physiother.* 2002;48:171-179.
- Mintken PE, Metrick L, Flynn TW. Upper cervical ligament testing in a patient with os odontoideum presenting with headaches. *J Orthop Sports Phys Ther.* 2008;38:465-475.
- Di Fabio RP. Manipulation of the cervical spine: risks and benefits. *Phys Ther.* 1999;79:50-65.
- Hing WA, Reid DA, Monaghan M. Manipulation of the cervical spine. *Man Ther.* 2003;8:2-9.
- Rushton A, Rivett D, Carlesso L, et al. *International Framework for Examination of the Cervical Region for Potential of Cervical Arterial Dysfunction Prior to Orthopaedic Manual Therapy Intervention.* Auckland, New Zealand: International Federation of Orthopaedic Manipulative Physical Therapists; 2012.
- Osmotherly PG, Rivett DA. Knowledge and use of craniovertebral instability testing by Australian physiotherapists. *Man Ther.* 2011;16:357-363.
- Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med.* 2011;155:529-536.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159-174.
- Altman D, Machin D, Bryant T, Gardner MJ. *Statistics With Confidence: Confidence Intervals and Statistical Guidelines.* London, United Kingdom: BMJ Books; 2000.
- Guyatt G, Rennie D, Meade M, Cook M. *Users' Guides to the Medical Literature.* Chicago, IL: McGraw-Hill Professional; 2008.
- Olson KA, Paris SV, Spohr C, Gorniak G. Radiographic assessment and reliability study of the craniovertebral sidebending test. *J Man Manip Ther.* 1998;6:87-96.
- Uitvlugt G, Indenbaum S. Clinical assessment of atlantoaxial instability using the Sharp-Purser test. *Arthritis Rheum.* 1988;31:918-922.
- Forrester G, Barlas P. Reliability and validity of the Sharp-Purser test in the assessment of atlanto-axial instability in patients with rheumatoid arthritis. *Physiotherapy.* 1999;85:376.
- Kaale BR, Krakenes J, Albrektsen G, Wester K. Clinical assessment techniques for detecting ligament and membrane injuries in the upper cervical spine region: a comparison with MRI results. *Man Ther.* 2008;13:397-403.
- Mathews JA. Atlanto-axial subluxation in rheumatoid arthritis. *Ann Rheum Dis.* 1969;28:260-266.
- Stevens JC, Cartlidge NE, Saunders M, et al. Atlanto-axial subluxation and cervical myelopathy in rheumatoid arthritis. *Q J Med.* 1971;40:391-408.
- Cattrysse E, Swinkels R, Oostendorp R, Duquet W. Upper cervical instability: are clinical tests reliable? *Man Ther.* 1997;2:91-97.
- Hutting NV, Verhagen AP, Vijverman V, et al. Diagnostic accuracy of premanipulative vertebrobasilar insufficiency tests: a systematic review. *Man Ther.* 2013;18:177-182.
- Deeks JJ, Altman DG. Diagnostic tests, 4: likelihood ratios. *BMJ.* 2004;329:168-169.
- Neva MH, Hakkinen A, Makinen H, et al. High prevalence of asymptomatic cervical spine subluxation in patients with rheumatoid arthritis waiting for orthopaedic surgery. *Ann Rheum Dis.* 2006;65:884-888.
- Richter RR, Reinking MF. Evidence in practice: how does evidence on the diagnostic accuracy of the vertebral artery test influence teaching of the test in a professional physical therapist education program? *Phys Ther.* 2005;85:589-599.
- Bossuyt PM, Reitsma JB, Bruns DE, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Clin Chem.* 2003;49:7-18.
- Gajdosik CG, Ostertag S. Cervical instability and Down syndrome: review of the literature and implications for physical therapists. *Pediatr Phys Ther.* 1996;31-36.
- White KS, Ball WS, Prenger EC, et al. Evaluation of the craniocervical junction in Down syndrome: correlation of measurements obtained with radiography and MR imaging. *Radiology.* 1993;186:377-382.
- Erickson SJ. High-resolution imaging of the musculoskeletal system. *Radiology.* 1997;205:593-618.
- Vetti N, Krakenes J, Ask T, et al. Follow-up MR imaging of the alar and transverse ligaments after whiplash injury: a prospective controlled study. *AJNR Am J Neuroradiol.* 2011;32:1836-1841.
- Vetti N, Krakenes J, Damsgaard E, Rorvik J, Gilhus NE, Espeland A. Magnetic resonance imaging of the alar and transverse ligaments in acute whiplash-associated disorders 1 and 2: a cross-sectional controlled study. *Spine (Phila Pa 1976).* 2011;36:E434-E440.

---

## Diagnostic Accuracy of Upper Cervical Spine Instability Tests

---

### 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<sup>a</sup>

Domain	Patient Selection	Index Test	Reference Standard	Flow and Timing
<b>Description</b>	Describe methods of patient selection Describe included patients (prior testing, presentation, intended use of index test, and setting)	Describe the index test and how it was conducted and interpreted	Describe the reference standard and how it was conducted and interpreted	Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the $2 \times 2$ table (refer to flow diagram) Describe the time interval and any interventions between index test(s) and reference standard
<b>Signaling questions (yes/no/unclear)</b>	Was a consecutive or random sample of patients enrolled?	Were the index test results interpreted without knowledge of the results of the reference standard?	Is the reference standard likely to correctly classify the target condition?	Was there an appropriate interval between index test(s) and reference standard?
	Was a case-control design avoided?	If a threshold was used, was it prespecified?	Were the reference standard results interpreted without knowledge of the results of the index test?	Did all patients receive a reference standard?
	Did the study avoid inappropriate exclusions?			Did all patients receive the same reference standard?
				Were all patients included in the analysis?
<b>Risk of bias: high/low/unclear</b>	Could the selection of patients have introduced bias?	Could the conduct or interpretation of the index test have introduced bias?	Could the reference standard, its conduct, or its interpretation have introduced bias?	Could the patient flow have introduced bias?
<b>Concerns regarding applicability: high/low/unclear</b>	Are there concerns that the included patients do not match the review question?	Are there concerns that the index test, its conduct, or interpretation differs from the review question?	Are there concerns that the target condition as defined by the reference standard does not match the review question?	

<sup>a</sup> Reproduced with permission from: <http://www.bris.ac.uk/quadas/>.

# Physical Therapy

Journal of the American Physical Therapy Association



## **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

doi: 10.2522/ptj.20130186

---

### **References**

This article cites 29 articles, 8 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/fora.xhtml>

---