Taylor & Francis
Taylor & Francis Group

Check for updates

Systematic review of the diagnostic accuracy, reliability, and safety of the sharp-purser test

Cody J. Mansfield [a,b,c], Charlie Domnisch[b], Laura Iglar[b,c], Laura Boucher[a,d], James Onate[a,d] and Matthew Briggs[a,b,c,d,e]

aSchool of Health and Rehabilitation Sciences, College of Medicine, The Ohio State University, Columbus, OH, USA; bOSU Sports Medicine, The Ohio State University Wexner Medical Center, Columbus, OH, USA; cOrthopaedic Manual Physical Therapy Fellowship, The Ohio State University, Wexner Medical Center, Columbus, OH, USA; dSports Medicine Research Institute, The Ohio State University Wexner Medical Center, Columbus, OH, USA; eDepartment of Orthopaedics, The Ohio State University Wexner Medical Center, Columbus, OH, USA

ABSTRACT

Introduction: The Sharp-Purser Test (SPT) is used to assess for atlantoaxial instability (AI) in patients with rheumatoid arthritis (RA). The test is commonly used by clinicians; however, many experts argue it lacks reliability and validity along with concerns of safety. The primary purpose of this review is to determine the diagnostic accuracy of the SPT to detect AI.

Methods: A search of five databases was performed from inception to 19 December 2018 using search terms related to the SPT. Studies were eligible for inclusion if the SPT was used on a patient/participant. Methodological quality assessment of diagnostic studies was performed with the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) for studies that reported data to calculate sensitivity (SN), specificity (SP), positive likelihood ratio (+LR), and negative likelihood ratio (-LR).

Results: The search yielded 1009 articles, and 32 studies met the inclusion criteria for analysis. Meta-analysis on diagnostic accuracy studies assessing the SPT was not possible due to statistical heterogeneity. Six diagnostic accuracy studies assessed the SN of the SPT ranging from 0.19 to 1.00. Four of the studies assessed SP of the SPT ranging from 0.71 to 0.98. The +LR was identified in 4 studies was 0.655, 1.73, 22, and 17.25. The -LR was 1.14, 0.799, 0.571, and 0.323. Seven RCTs utilized the SPT to screen for AI, and the SPT was used in 18 case reports.

Conclusion: The SPT may be inappropriate to use due to inconsistent validity, poor inter-rater reliability, and potential to cause harm.

Level of evidence: 1

Introduction

The Sharp-Purser Test (SPT) is used to assess for atlantoaxial instability (AI) and was developed by Dr. Sharp and Dr. Purser from their experience in identifying AI in patients with rheumatoid arthritis (RA) [1,2]. The use of the SPT is a contentious topic [3,4]. The test is used by clinicians during the examination of a patient to identify AI, and prior to the performance of manipulation, joint mobilization or dry needling to the upper cervical spine. However, many argue that the test is unreliable, lacks validity, and is not safe to use[4].

There are two steps to the SPT (Figure 1). First, the patient's head is semi-flexed to 20º-30º in a seated position [5]. The clinician assesses for any numbness and/or tingling in the arms or legs [6]. If the transverse ligament is compromised, the dens of C2 may compress the tracts of the spinal cord with upper cervical flexion, thereby producing myelopathic symptoms in the arms and/or legs [6]. Secondly, the clinician will stabilize the C2 spinous process with a pincer grip of one hand and apply an anterior to posterior force

through the forehead of the patient with the clinician's other hand [5]. Some studies have also described this maneuver as moving the head into extension [2]. The test is considered positive if the myelopathic symptoms that were produced with upper cervical flexion resolve, or a sliding motion of the head occurs posteriorly, sometimes resulting in an audible clunk with the approximation of the dens on the posterior aspect of the anterior ring of the atlas [5]. Recommendations following a positive SPT include radiographs to assess for the atlantodental interval (ADI) with the neck in neutral, flexion, and extension to assess for AI [7]. An ADI of ≥ 3 mm is suggestive of AI [7].

RA affects approximately 1% of the population, and cervical pathology is very common in patients with RA [8]. Cervical myelopathy commonly occurs in patients with RA, with AI being the most common cause of spinal cord compression [9]. Serious complications can occur from AI if left untreated, such as myelopathy and quadriparesis [6]. The seminal research on the SPT has been validated in patients with RA and is

Figure 1. The Sharp-Purser Test.

First, the patient's head is semi-flexed to 20º-30º in a seated position (WHITE ARROW). The clinician assesses for any numbness and/or tingling in the arms or legs. If the transverse ligament is compromised the dens of C2 may compress the tracts of the spinal cord with upper cervical flexion, thereby producing myelopathic symptoms such as numbness and/or tingling in the arms and/or legs. Secondly, the clinician will stabilize the C2 spinous process with a pincer grip of one hand and apply an anterior to posterior force through the forehead of the patient with the clinician's other hand (BLACK ARROW). The test is considered positive if the myelopathic symptoms that were produced with upper cervical flexion resolve, or a sliding motion of the head occurs posteriorly, sometimes resulting in an audible clunk with the approximation of the dens on the posterior aspect of the anterior ring of the atlas.

commonly used to screen for AI in clinical practice [10–12]. Many have found the SPT to be useful in screening for AI [3,13,14]. Even though the use of the SPT has been established since the 1960's, there is no consensus amongst medical professions when to appropriately use the SPT [3,4,13]. A previous systematic review by Hutting et al. (2013) assessed the diagnostic accuracy of upper cervical spine instability tests, and identified the seminal research on the SPT, however meta-analysis was not performed due to heterogeneity of studies, and other study types (randomized controlled trials (RCTs) and case reports) were not included [7]. Therefore, **aim 1** of our review was to determine the diagnostic accuracy of the SPT and perform an update to this previous review. Our review is an important contribution to the literature because we are performing an update to the Hutting et al. (2013) review in regard to the diagnostic properties of the SPT, and our review goes a step further with our ancillary aims to assess the reliability, safety and patterns of use of the SPT [7]. **Aim 2** of our review was to determine the reliability of the SPT. **Aim 3** of our review was to determine if there was any

evidence that the SPT is harmful or unsafe to use in clinical practice. The dangerous mechanism behind a positive SPT involves compressing the spinal cord via the dens of C2 and performing a maneuver to alleviate pressure from the spinal cord, which could be harmful or unsafe in high risk populations. Lastly, **aim 4** of this review was to understand patterns for use of the SPT in clinical practice by analyzing case reports and RCTs. We hypothesize that the SPT may be used inappropriately in clinical practice and research trials and that the SPT is not harmful or unsafe to use in clinical practice.

Methods

Study design

This review was registered with PROSPERO (CRD42019121657), and utilized the *Preferred Reporting Items for Systematic Reviews and Metanalyses* (PRISMA) guidelines [15]. The protocol guided the review process for **AIM 1**, however a broad search strategy was implemented in order to assess **AIMS 2–4**. After **AIM 1** was completed, the process of analysis of other studies to answer **AIMS 2–4** was developed.

Data sources and searches

An electronic search of five databases (PubMed, CINAHL, SPORTDiscus, EMBASE, and Google Scholar) was performed from inception to 19 December 2018. Hutting et al. only searched 4 databases (PubMed, CINAHL, EMBASE, and RECAL Legacy). PubMed, CINAHL, and EMBASE were selected to identify similar key databases that Hutting et al. (2013) used, with the addition of SPORTDiscus and Google Scholar. Google Scholar was included since the database typically yields the most articles compared to other databases when implementing a search strategy. MEDLINE was not searched since PubMed accesses the MEDLINE database, and although searches in both databases may differ slightly, we felt the inclusion of 4 other databases accounted for this. The exact search terms inserted into each database were the following: (((sharp purser test) OR sharp purser) OR sharp-purser) OR sharp-purser test. Our search terms account for variability in the spelling of the SPT. The database search was performed by the primary author (CM), and studies were uploaded to Covidence systematic review software [Veritas Health Innovation, Melbourne, Australia (www.covidence.org)] [16].

Eligibility criteria

The research question was formulated using PICO (P: patient, problem or population; I: intervention; C: comparison; O: outcome) to formulate the core concepts of our question. In addition to performing an

update to the diagnostic accuracy specific to the SPT reported in Hutting et al. (2013) systematic review on upper cervical instability tests we expanded the inclusion criteria to include any study that used the SPT on a patient or group of participants and published the results in a peer reviewed journal [7]. Intervention was not specified and no comparison group was defined. Outcomes of interest were different for each AIM, but included data related to diagnostic properties of the SPT related to assessing the validity, reliability, and evidence of the SPT causing harm. Study types of interest included cohort studies assessing incidence, reliability or validity, case reports and RCTs. **AIM 1** of our study was in line with Hutting et al. (2013), however we had to include other study types in order to adequately address **AIMS 2–4** [7]. Studies were excluded if full text was not available, English language was not used, the article was not published from a peer-reviewed journal, or the SPT was not used. Although systematic reviews and literature reviews were excluded, they were screened for content and references.

Study selection

Any study where the SPT was used on a participant was included. After articles were identified and uploaded to Covidence (Veritas Health Innovation, Melbourne, Australia), titles and abstracts were independently screened for inclusion by two authors (CM, LI) [16]. A third author resolved conflicts during the screening phase (CD). The two-reviewer process (CM, CD), with third reviewer to resolve conflicts (LI), was repeated for the full text review.

Data extraction and diagnostic accuracy measures

Extraction of pertinent information from each study included in the review was performed by the primary author (CM) and confirmed by coauthor (CD). Custom designed tables were formed with pertinent information from all studies included. Pertinent information for diagnostic accuracy and reliability studies included: author, year, number of subjects, age, sex, imaging reference standard, setting and complications with use of the test (Table 1). For research trials that used the SPT we included pertinent information such as author, year, number of subjects, age, sex, number of subjects excluded based on positive SPT, intervention assessed in research trial (Table 2). For case reports included in this review we included information such as author, year, age, sex, pathology, imaging, complication with use of the SPT, and manual therapy intervention used (Tables 3 and 4). Diagnostic studies that used advanced imaging such as radiographs, computed tomography scan or magnetic resonance imaging as

Table 1. Diagnostic Accuracy and Reliability Studies of that Sharp-Purser Test.

Author and Year	Study Type	Subjects	Age	Sex	Setting	Imaging Reference Standard	Complications with use of Test
Forrester and Barlas 1999[19]	Reliability and validity study cross sectional design	31 patients with RA	Mean age 57.5 years; range 29–24 yrs; mean duration of RA was 15 yrs	21 women, 10 men	Physical therapy practice	Radiographs	No
Cattrysse 1997[3]	Reliability study	4 examiners and 11 children with Down's Syndrome	Mean age of 9 yrs; age range from 5–14 yrs	Sex of subjects unknown; sex of examiners is 50% male/female	Physical therapy practice	Radiographs	No
Montemerani 1994[28]	Incidence study with cross-sectional design	183 subjects with RA	Mean age ± SD: 33.35 ± 14.72	36 males; 147 females	Physician practice	Radiographs/CT scan	No
Katayama 1990[25]	Validity study with cross-sectional design	104 subjects with RA	Mean age of 53.3 yrs; age range 28 to 76 yrs	25 males; 79 females	Physician practice	Radiographs/CT scan	No
Stevens 1971[12]	Incidence study with cross-sectional design	100 subjects with RA	Mean age of 54.2 yrs; age range from 21–75 yrs	27 males; 73 females	Physician practice	Radiographs	No
Mathews 1969[13]	Incidence study with case-control study design	76 subjects with RA; 28 healthy control subjects	Under 50 yrs: 18 50–59 yrs: 24 60–69 yrs: 24 Over 70 yrs: 10	26 males; 50 females	Physician practice	Radiographs	No
Uitvlugt 1988[20]	Validity study with cross-sectional design	123 subjects with RA	NR	NR	Physician practice	Radiographs	No

Yrs: years; RA: rheumatoid arthritis; SD: standard deviation; CT scan: computed tomography scan; NR: not reported

Table 2. Research Trials that used the Sharp-Purser Test to Screen/Exclude Participants.

Author and Year	Subjects (mean ± SD)	Subjects Excluded based on positive Sharp-Purser Test	Intervention Assessed
Bortolazzo 2015[10]	10 women with TMD (25.8 ± 6.8 yrs)	0	Cervical manipulation
Llamos-Ramos 2014[6]	94 subjects with chronic neck pain; 66% female (31 ± 3 yrs)	0	Dry needling
Mejuto-Vazques 2014[27]	17 subjects with acute neck pain; 53% female (25 ± 4 yrs)	0	Dry needling
Saavedra-Hernández 2013[1]	82 subjects with neck pain; 50% female (45 ± 9 yrs)	0	Cervical, cervicothoracic and thoracic manipulation
Martínez-Segura 2012[4]	90 subjects with chronic neck pain; 51% female (37 ± 8 yrs)	0	Cervical and thoracic manipulation
Puentedura 2012[5]	82 subjects with neck pain; 59% female (38.3 ± 14.7)	0	Cervical manipulation
Puentedura 2011[42]	24 subjects with neck pain; 67% female (33.7 ± 6.4 yrs)	0	Cervical and thoracic manipulation

SD: standard deviation; TMD: temporomandibular disorder; yrs: years

Table 3. Case Reports with a Positive Sharp-Purser Test.

Author and Year	Age	Sex	Pathology	Imaging	Complications with use of Test	Manual Therapy Used
Atlantoaxial Pathology						
Lyons 2018[40]	31	Male	Atlantoaxial instability with ADI of 6 mm on flexion radiograph	Radiograph	No	No
Mourad 2016[29]	37	Male	Basilar Impression/invagination	MRI	No	No
Mintken 2008[33]	23	Female	C2-C3 Klippel-Feil congenital fusion and os odontoideum with cervical spine narrowing	Radiograph/MRI	No	No
Rebbeck 2006[30]	14	Male	Congenital defect of atlas (spina bifida atlanto)	Radiograph/MRI/CT scan	No	No
Kosaka 2002[32]	70	Female	Periodontoid pannus migration into spinal canal	Radiograph/myelogram/MRI	No	No
Cervical Myelopathy						
Mansfield 2018[13]	55	Female	Myelopathy due to disc osteophytes in central canal from C4-C7	Radiograph/MRI/CT scan	No	No

MRI: magnetic resonance imaging; ADI: atlantodental interval; CT scan: computed tomography scan

Table 4. Case reports with a negative sharp-purser test.

Author and Year	Age	Sex	Pathology	Imaging	Complications with use of Test	Manual Therapy Used
Head/Neck pain/Headache						
Jayaseelan 2016[9]	46	Female	Facial pain with temporomandibular disorder	None	No	Yes
Sillevis 2015[11]	20	Female	Cervicogenic Headache	Radiograph	No	Yes
Rodeghero 2013[47]	18	Female	Migraines/Cervicogenic Headache	Radiographs	No	Yes
Lowry 2011[26]	37	Female	Neck/shoulder pain secondary to whiplash	Radiographs	No	Yes
Mathers 2011[31]	32	Female	Headache, neck pain, and alar ligament laxity	Radiographs/MRI	No	Yes
Wong 2011[8]	50	Female	Neck/thoracic pain with known Eagle syndrome	MRI/CT scan/Doppler	No	Yes
Young 2009[24]	64	Female	C5-C7 degenerative disc disease	Radiographs	No	Unknown
van Duijn 2007[23]	40	Female	Cervicogenic headache	Radiograph	No	Yes
Glynn 2006[7]	46	Female	Cervical/thoracic pain	MRI	No	Yes
Derrick 1992[36]	25	Female	Headache/alar ligament laxity	Radiograph/CT scan	No	Yes
Cervical Radiculopathy						
Naimi 2015[34]	64	Female	Radiculopathy with right C6-C7/C7-C8 foramina narrowing	Radiographs/MRI	No	Yes
Internal Carotid Artery Dissection						
Willet 2011[45]	45	Male	Internal carotid artery dissection with neck pain and headache after water skiing accident	MRI	No	No

MRI: magnetic resonance imaging; CT scan: computed tomography scan
Table 5 to be made available online.

a references standard, and reported information to calculate sensitivity (SN) and specificity (SP) were considered for meta-analysis. If the SPT was used in a similar population (i.e. RA, AI, etc), then the true positives, false positives, false negatives and true negatives were pooled to calculate SN and SP. Descriptive statistics were reported as SN and SP using Review Manager 5 software [(RevMan 5) The Nordic Cochrane Center, The

Cochrane Collaboration] [17]. Positive likelihood ratio (+LR) and negative likelihood ratio (-LR) were calculated from the reported SN and SP. Interpretation of +LR and -LR values were based on threshold LR ranges available from the *Journal of the American Medical Association* [18]. Statistical heterogeneity was measured using the I [10] statistic for diagnostic accuracy studies that reported data to calculate both SN and SP. It was

Table 5. Methodological quality assessment of diagnostic accuracy studies with QUADAS-2.

	Risk of Bias				Applicability Concerns		
Author and Year	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Forrester 1999[19]	?	?	?	?	☹	☹	?
Montemerani 1994[28]	☺	☹	☺	?	☺	?	?
Katayama 1990[25]	☹	☹	☺	?	☺	?	?
Stevens 1971[12]	☺	?	?	☺	☺	☺	?
Mathews 1969[13]	☹	☹	?	☺	☺	☺	?
Uitvlugt 1988[20]	☺	?	?	?	☺	☺	?

☺= low risk; ☹= high risk; ? = unclear risk

decided a priori that if statistical heterogeneity was low (<40%) as measured by the I [10] statistic then meta-analysis would be performed [19].

Risk of bias

Methodological quality assessment of diagnostic studies was performed with the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) for studies that reported data to calculate SN and SP. The Hutting et al. (2013) review also used the QUADAS-2 to appraise the quality of studies included. Scores on the QUADAS-2 tool were completed by two authors (CM, CD) and discrepancies resolved by a third author (MB) [20]. Publication bias was assessed if greater than 10 studies were yielded for meta-analysis. Risk of bias was limited to diagnostic accuracy studies only. Reliability studies, RCTs and case reports were not included in methodological quality assessment.

Analysis of the reliability and safety of the sharp-purser test

A narrative synthesis was conducted for any study that assessed the reliability of the SPT. Pertinent information such as population, rater information, reference standard, methods and results would be reported. All included studies will be assessed for any evidence of the SPT causing harm or resulting in a participant of a study dropping out or being excluded. When it was unclear if the SPT resulted in exclusion of a participant or harm to a patient, the corresponding author was contacted to clarification.

Results

The search yielded 1009 articles, and 32 studies met inclusion for analysis (Figure 2). Two studies assessed the reliability, and six studies assessed the validity of the SPT (Table 1) [10–13,21,22]. SN, SP, +LR, and -LR was determined for 4 studies (Figure 3) [11–13,23]. Two other studies were identified, however only SN could be determined (Figure 3) [21,22]. Seven RCTs used the SPT to exclude subjects from participating in studies prior to randomization into a treatment group involving dry needling, mobilization or manipulation directed toward

the upper cervical spine (Table 2) [24–30]. Eighteen case reports were included (Tables 3 and 4). Six case reports yielded a positive SPT (Table 3) [6,14,31–34]. Twelve case reports yielded a negative SPT (Table 4) [35–46].

Methodological quality results

☺ ☹Methodological quality of 6 studies with diagnostic accuracy data on the SPT were assessed using the QUADAS-2 (Table 5 available online) [11–13,21,22]. There was 100% agreement between the two reviewers on all 6 studies assessed with the QUADAS-2 tool, and the 4 studies reported by Hutting et al. yielded the same QUADAS-2 tool results as this review [7]. Using the QUADAS-2 tool, there were several domains that indicated an unclear or high risk of bias in the included studies (Figure 4). Two studies did not demonstrate evidence of enrolling patients consecutively or utilizing a random sample [13,21]. All studies, except 2 [13,23], did not demonstrate evidence that researchers were blinded and may have known the results of imaging prior to performing the SPT. All 6 studies used radiographs as the index test to identify ADI, and only two studies used computed tomography scan [21,22]. Radiographs may not be the most accurate or reliable imaging modality to identify AI [47]. The minimum threshold of the SPT to detect AI is an ADI ≥ 3 mm in adults, however the SN and SP of this test may change depending on how AI is defined based on the ADI [10]. Publication bias was not assessed due to only 6 studies qualifying for meta-analysis.

Diagnostic accuracy of the sharp-purser test

Statistical heterogeneity was assessed for the 4 studies that reported data to calculate both SN and SP [11–13]. There was high statistical heterogeneity amongst all 4 studies (I [10]> 40%). Based on the high heterogeneity, pooling of the studies and meta-analysis was not advisable [7]. Six diagnostic accuracy studies assessed the SN of the SPT ranging from 0.19 to 1.00. Four of the studies assessed SP of the SPT ranging from 0.71 to 0.98 (Figure 3). One study assessed the differences in SN and SP with an ADI of >3mm versus ADI >4 mm [12]. For ADI >3mm the SN of the SPT was reported as 69% with a SP of 96%, and
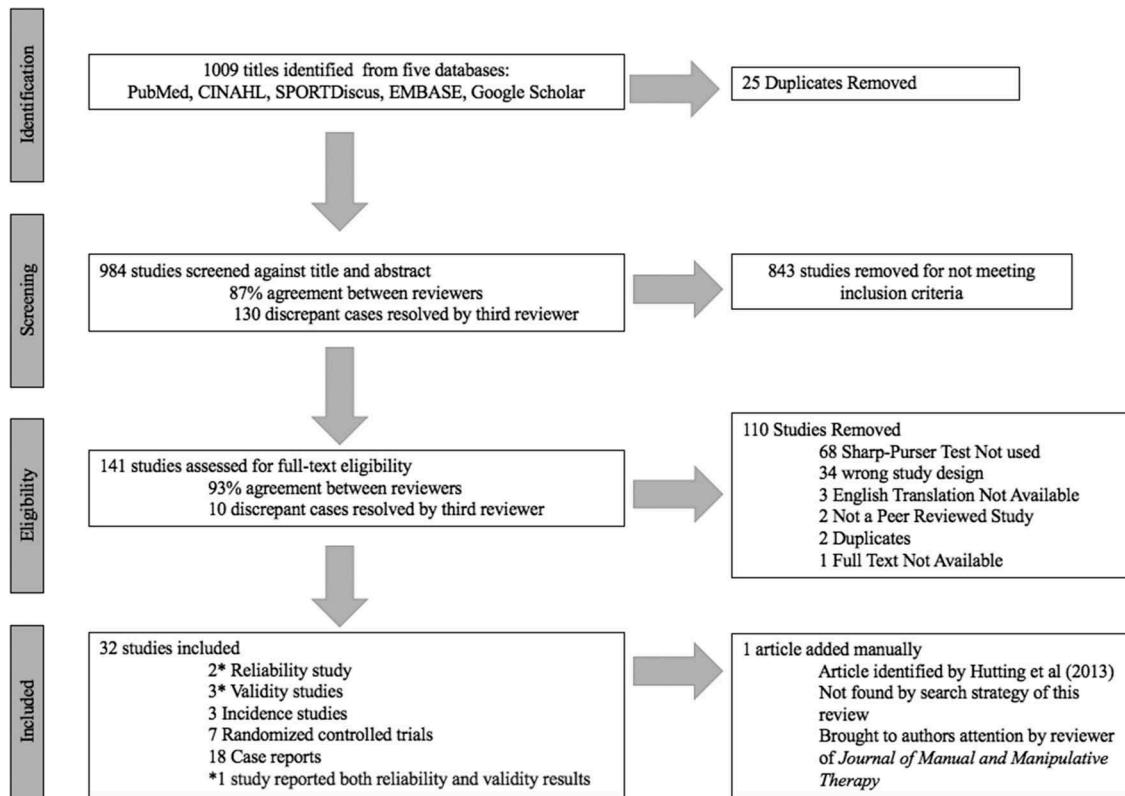
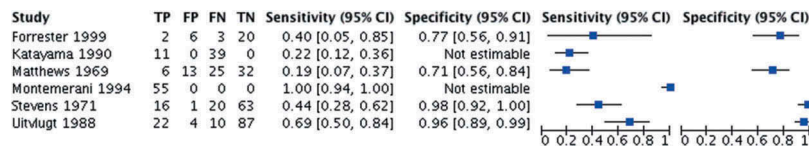**Figure 2.** Search and selection strategy.

Figure 3. Diagnostic properties of the Sharp-Purser Test.

for an ADI >4 mm the SN improved to 88% [12]. Matthews (1969) and Forrester (1999) reported +LR values of 0.655 and 1.73, respectively, and Stevens (1971) and Uitvlugt (1988) reported +LR of 22 and 17.25, respectively [11–13,23]. In regard to the -LR, Mathews (1969), Forrester (1999), Stevens (1971) and Uitvlugty (1988) reported 1.14, 0.77, 0.571 and 0.323, respectively. [11–13,23]

Reliability of the sharp-purser test

There have only been 2 studies published assessing the reliability of the SPT [10]. One study used 4 examiners in 11 children with Down's syndrome, where only 2 of 11 children were radiographically unstable [10]. Intra-observer reliability for the SPT (modified kappa) was 0.67, 0.45, 0.29, and 0.67, respectively [10]. Inter-observer reliability ranged from 0.09 to 0.67 [10]. Forrester (1999) assessed the reliability of 6 physical therapists with manual therapy experience using the SPT on 31 participants with an average diagnosis of RA of 15 years [23]. To determine inter-

therapist reliability, overall kappa value was 0.20 with a standard error mean of 0.05, and 95% confidence interval of 0.09 and 0.30.

Using the sharp-purser test to screen for subject participation in research trials

Seven manual therapy RCTs attempted to exclude subjects from their study with a positive SPT. [24,25,27–30] The goal was to exclude subjects that may have AI and to eliminate any subjects that would be unsafe to receive manual therapy to the upper cervical spine. Performing upper cervical manual therapy interventions on a subject with AI could lead to catastrophic consequences (Table 2) [24,25,27–30]. All seven trials did not yield a single positive SPT to exclude a subject.

Case reports

Six case reports yielded a positive SPT (Table 3) [6,14,31–34]. Five involved pathological conditions

Proportion of studies with low, high, or unclear risk of bias on the QUADAS-2

Proportion of studies with low, high or unclear concerns regarding applicability on the QUADAS-2
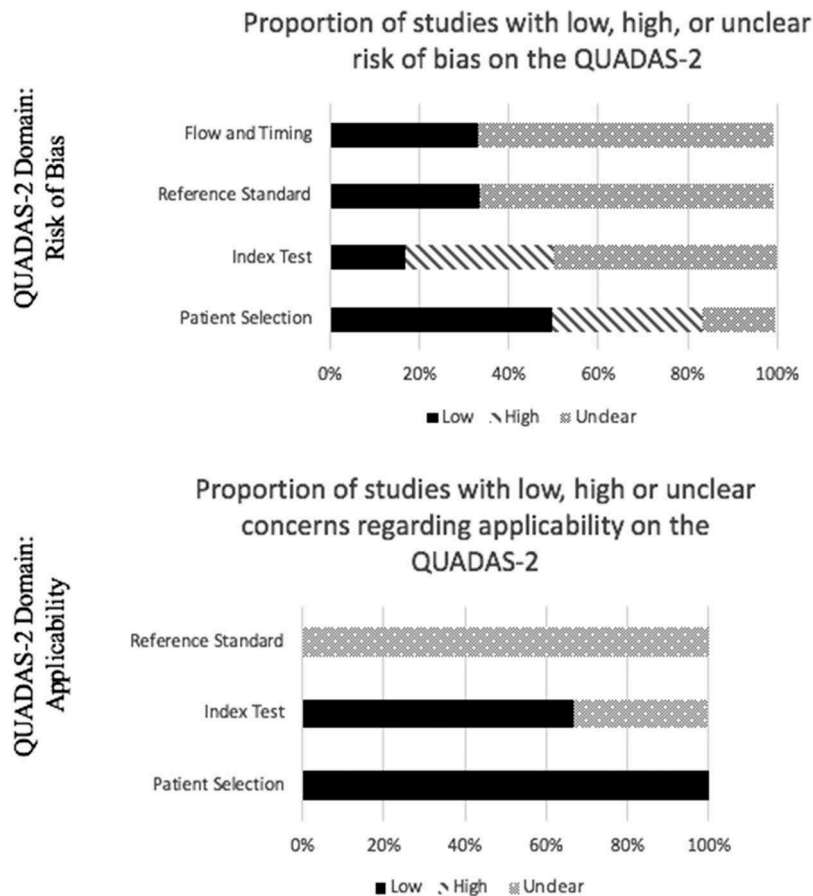

Figure 4. Proportion of studies with low, high or unclear concerns regarding risk of bias and applicability, %.

affecting the atlantoaxial joint, such as AI, basilar invagination, periodontoid pannus, os odontoideum or spina bifida atlanto [14,31–34]. One study identified a false positive SPT with confirmed cervical myelopathy from C4 to C7 [6]. According to the majority of case reports, a negative SPT preceded cervical manual therapy techniques to treat neck pain or headaches (Table 4) [35–46].

Safety

The SPT has come under scrutiny because it may be an unsafe test to perform. Patients with AI are at risk of myelopathy if not identified and treated. In all 31 studies included in this review, there was no evidence that the SPT was unsafe to perform or caused harm (Table 1–4). Even though no evidence exists in the scientific literature to suggest the SPT is harmful or unsafe, no study has specifically assessed the safety of the SPT as part of a research question. Therefore, this limits our review to make a definitive conclusion about the relative safety of the SPT.

Discussion

The SPT is used to assess for AI and is commonly performed prior to manipulation, joint mobilization and dry needling to the upper cervical spine. [1,2] There is division amongst professionals on whether or not the SPT should be used in clinical practice with experts citing a lack of validity, reliability, and potential perilous consequences of performing the test in high risk individuals [4]. The goal of this review of the SPT is to perform an update to the previous review by Hutting et al. (2013) performed on upper cervical instability tests and broaden the inclusion criteria to elucidate the potential benefits and pitfalls of the SPT [7]. Therefore, **AIM 1** was to determine the diagnostic accuracy of the SPT, **AIM 2** was to determine the reliability, **AIM 3** was to determine if the SPT was harmful or unsafe to use, and **AIM 4** was to understand patterns of use of the SPT by analyzing case reports and RCTs. Although our original systematic review protocol published on PROSPERO focused on the synthesis of data for **AIM 1**, we did not create a priori protocol for synthesizing data from other study types to answer **AIMS 2–4**. Method for synthesizing data for **AIMS 2–4** was completed after analysis of **AIM 1** was finished.

Our update was specific to the SPT only, whereas Hutting et al. (2013) assessed the diagnostic accuracy of several upper cervical instability specials tests [7]. We found two studies that Hutting et al. (2013) did not include where SN was able to be estimated but SP

was not [21,22]. Hutting et al. (2013) found 1 study, Forrester (1999), that our search strategy did not identify [23]. We manually added this article after our search protocol failed to identify it, and using algebraic principles, we estimated the number of true positives, false positives, true negatives and false negatives from the study [23].

Six diagnostic accuracy studies assessed the SN of the SPT ranging from 0.19 to 1.00. Four of the studies assessed SP of the SPT ranging from 0.71 to 0.98 **(AIM 1)**. The +LR lacked consistency amongst the studies, with Matthews (1969) and Forrester (1999) reporting 0.655 and 1.73, respectively, and Stevens (1971) and Uitvlugt (1988) reporting 22 and 17.25, respectively [11–13,23]. The higher + LR scores from Stevens (1971) and Uitvlugt (1988) suggests a significant change in posttest probability with a positive SPT, whereas the study by Matthews (1969) and Forrester (1999) suggest minimal to no increase in post-test probability with a positive SPT. [11–13,23] In regard to the -LR, Mathews (1969), Forrester (1999), Stevens (1971) and Uitvlugty (1988) reported 1.14, 0.77, 0.57 and 0.32, respectively [11–13,23]. The studies suggests small to no effect on posttest probability with a negative SPT [11–13,23]. Matthews (1969) included patients with RA and controls and Forrester (1999) only included individuals with RA and had a higher risk of bias than Stevens (1971) and Uitvlugt (1988) according to the QUADAS-2 domain of patient selection [11–13,23].

AIM 2 of our review was to determine the reliability of the SPT. Only two studies assessed the reliability of the SPT [10,23]. One study assessed the reliability of 4 raters using the SPT in a population of children with Down's syndrome [10] and the other used six raters in a population of participants with RA [23]. Both studies demonstrated poor reliability which limits the use of the SPT in clinical practice and the use of the SPT as a screening test to exclude participants from research studies [10,23].

AIM 3 of our study was to determine if the SPT was harmful or unsafe to use and **AIM 4** was to determine patterns of use of the SPT in clinical practice and research. There was no evidence in the scientific literature to suggest the SPT is harmful. Currently there is a paucity of evidence supporting or refuting performing the SPT in high risk populations, which is why we implemented a broad search strategy to identify any peer reviewed study that used the SPT to identify any record of compromise due to the performance of the test. Future research is needed to assess if the SPT is harmful or unsafe when performed on individuals with upper cervical instability. However it is unlikely that a study like this will ever be performed due to the inherent risk of performing the test in a high risk population, and likely there are other subjective and objective examination findings that would warrant referral for imaging in the presence of cervical instability.

Many manual therapy focused RCTs utilized the SPT to screen for AI, and in many case reports, a negative SPT preceded manual therapy treatment to the cervical or thoracic spine. Six case reports yielded a positive SPT in individuals at increased risk of compromising the spinal cord in the upper cervical region, and there was no evidence the SPT led to a complication and in most cases assisted the clinical reasoning of the clinician (Table 3).

There is evidence to suggest that the SPT may be beneficial to use in patients with RA **(AIM 1)**, however the reliability of the SPT is poor **(AIM 2)**. Although there was no evidence that performing the SPT is harmful or unsafe in the scientific literature, extreme caution should be utilized when using the SPT in high risk populations since no researcher has thoroughly assessed the safety of the SPT in various populations **(AIM 3)**. Manual therapy research trials studying the effectiveness of manipulation and dry needling in the cervical and thoracic spine used the SPT to exclude any participant with a positive SPT **(AIM 4)**. There was no evidence any participant was excluded based on a positive SPT. In case reports, a positive SPT aided in the clinical reasoning of the clinician to facilitate the patient's care, and all case reports identified pathology in the upper cervical spine **(AIM 4)**. A negative SPT identified in case reports often preceded manual therapy intervention to the cervical or thoracic spine **(AIM 4)**. More research is needed to elucidate the diagnostic properties of the SPT and its application to other populations and to establish the reliability of the study.

Limitations

Although this review is the most exhaustive regarding the SPT in the literature, there are limitations worth noting. We only searched five databases with a very simple, but broad, search strategy. Our search strategy failed to find one diagnostic accuracy study that the review from Hutting et al. (2013) found, however we found two other diagnostic studies that were not included by Hutting et al. (2013) [7]. Upon further inspection, the Forrester (1999) article, the article we failed to include in our review, is searchable in Google Scholar, however was missed by the two reviewers in the screening process [23]. The article was published in *Physiotherapy*, and was likely missed by our screeners because it was only seven paragraphs long and was not identified by screening Hutting et al. because they did not report diagnostic properties of it alongside other studies. Also, even though the SPT was performed in subjects with varying degrees of severity of AI and no consequences of the SPT were reported, we may be

unable to fully answer the question of the dangers of performing the SPT in high risk populations with RA and Down's syndrome. Since the SPT has not been validated in any other populations, clinicians should use the SPT with extreme caution and in combination with other subjective and objective signs and symptoms for the general population with neck pain and symptoms suggestive of AI. To date, only case reports have been published on the use of the SPT in populations without AI or RA warranting further exploration by researchers into the used of the SPT in populations other than those with RA. Another limitation of this systematic review is that it only included articles written in English.

Conclusion

The SPT is used during the physical examination of a patient to identify AI. Many argue the SPT is unreliable, dangerous, lacks validity, and does little to change clinical management of a patient that presents with AI [4]. Our results suggest there is insufficient evidence to perform the SPT outside of patients presenting with RA as the only studies with diagnostic properties were performed in this patient population. No study to date has established the SPT as a reliable test. There appears to be no evidence the SPT is dangerous, however a definitive conclusion about the relative safety of the SPT could not be made due to insufficient evidence.

Disclosure statement

Funding

ORCID

Cody J. Mansfield (iD) http://orcid.org/0000-0003-3266-463X

References

[1] Sharp J, Purser DWLJ. Rheumatoid arthritis of the cervical spine in the adult. Ann Rheum Dis. 1958;17 (3):303–313.

[2] Sharp J, Purser DW. Spontaneous atlanto-axial dislocation in ankylosing spondylitis and rheumatoid arthritis. Ann Rheum Dis. 1961;20(1):47–77.

[3] Cook C, Brismée J-M, Fleming R, et al. Identifiers suggestive of clinical cervical spine instability: a Delphi study of physical therapists. Phys Ther. 2005;85(9PG–895–906):895–906.

[4] Meadows J. The Sharp-Purser test: a useful clinical tool or an exercise in futility and risk? J Man Manip Ther. 1998;6(2):97–100.

[5] Reiman MP. Orthopedic clinical examination. Champaign (IL): Human kinetics; 2015.

[6] Mansfield CJ. Cervical myelopathy causing numbness and paresthesias in lower extremities: A case report identifying the cause of a false positive Sharp–purser test. Physiother Theory Pract. 2018 Feb;1–8. DOI:10.1080/09593985.2018.1443361.

[7] Hutting N, Scholten-Peeters GGM, Vijverman V, et al. Diagnostic accuracy of upper cervical spine instability tests: a systematic review. Phys Ther. 2013;93(12PG–1686–1695):1686–1695.

[8] van der Woude D, van der Helm-van Mil AHM. Update on the epidemiology, risk factors, and disease outcomes of rheumatoid arthritis. Best Pract Res Clin Rheumatol. 2018;32(2):174–187.

[9] Keersmaekers A, Truyen L, Ramon F, et al. Cervical myelopathy due to rheumatoid arthritis. Case report and review of the literature. Acta Neurol Belg. 1998;98 (3):284–288.

[10] Cattrysse E, Swinkels RA, Oostendorp RADW. Upper cervical instability: are clinical tests reliable? Man Ther. 1997;2(2):91–97.

[11] Stevens JC, Cartlidge NEF, Saunders M, et al. Atlanto-axial subluxation and cervical myelopathy in rheumatoid arthritis. QJM An Int J Med. 1971;40(3):391–408.

[12] Uitvlugt G, Indenbaum S. Clinical assessment of atlantoaxial instability using the sharp-purser test. Arthritis Rheum. 1988;31(7):918–922.

[13] Mathews JA. Atlanto-axial subluxation in rheumatoid arthritis. Ann Rheum Dis. 1969;28(3PG–260):260.

[14] Mintekn P, Metrick L, Flynn T. Upper cervical ligament testing in a patient with os odontoideum presenting with headaches. J Orthop Sports Phys Ther. 2008;38 (8):465–475.

[15] Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg. 2010;8(5):336–341.

[16] Covidence systematic review software. Veritas health innovation, Melbourne, Australia. [cited 2018 December]. Available from: https://www.covidence. org/reviews/active

[17] Copenhagen: The Nordic Cochrane Centre TCC. RevMan 5. Version 5.3. Cochrane community; 2014. [cited 2018 December]. Available from: https://community.cochrane.org/help/tools-and-software/revman-5

[18] Jaeschke R, Guyatt GH, Sackett DL, et al. Users' guides to the medical literature. JAMA. 1994;271(9):703.

[19] Higgins JP Cochrane handbook for systematic reviews of interventions. Version 5.1.0. The Cochrane Collaboration; 2011 Mar. [cited 2018 December]. Available from: www.cochrane-handbook.org https://ci.nii.ac.jp/naid/20000796633/.

[20] Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011;155(8):529.

[21] Katayama M. Clinical and radiological evaluation of the spine due to rheumatoid arthritis. Bull Yamaguchi Med Sch. 1990;37(1–2PG–31–43):31–43.

[22] Montemerani M, Venturi C, Bracco S, et al. Involvement of atlanto-axial joint in rheumatoid arthritis: rare or frequent? Clin Rheumatol. 1994;13 (3PG–459–463):459–463.

[23] 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(7):376.

[24] Bortolazzo GL, Pires PF, Dibai-Filho AV, et al. Effects of upper cervical manipulation on the electromyographic activity of the masticatory muscles and the opening range of motion of the mouth in women with temporomandibular disorder: randomized and blind clinical trial. Fisioter E Pesqui. 2015;22(4PG–426–434):426–434.

[25] Llamas-Ramos R, Pecos-Martín D, Gallego-Izquierdo T, et al. Comparison of the short-term outcomes between trigger point dry needling and trigger point manual therapy for the management of chronic mechanical neck pain: a randomized clinical trial. J Orthop Sports Phys Ther. 2014;44(11):852–861.

[26] Martínez-Segura R, De-la-Llave-Rincón AI, Ortega-Santiago R, et al. Immediate changes in widespread pressure pain sensitivity, neck pain, and cervical range of motion after cervical or thoracic thrust manipulation in patients with bilateral chronic mechanical neck pain: a randomized clinical trial. J Orthop Sports Phys Ther. 2012;42(9PG–806–814):806–814.

[27] Mejuto-Vázquez MJ, Salom-Moreno J, Ortega-Santiago R, et al. Short-term changes in neck pain, widespread pressure pain sensitivity, and cervical range of motion after the application of trigger point dry needling in patients with acute mechanical neck pain: a randomized clinical trial. J Orthop Sports Phys Ther. 2014;44(4PG–252–260):252–260.

[28] Puentedura EJ. Development of a clinical prediction rule to identify patients with neck pain likely to benefit from cervical spine manipulation and a range of motion exercise. Ann Arbor (MI): Nova Southeastern University, ProQuest Dissertations Publishing; 2011.

[29] Puentedura EJ, Cleland JA, Landers MR, et al. Development of a clinical prediction rule to identify patients with neck pain likely to benefit from thrust joint manipulation to the cervical spine. J Orthop Sports Phys Ther. 2012;42(7PG–577–592):577–592.

[30] Saavedra-Hernández M, Arroyo-Morales M, Cantarero-Villanueva I, et al. Short-term effects of spinal thrust joint manipulation in patients with chronic neck pain: a randomized clinical trial. Clin Rehabil. 2013;27(6PG–504–512):504–512.

[31] Kosaka E, Ishihara H, Osada R, et al. Periodontoid pannus migration into the spinal canal with reduction of rheumatoid atlantoaxial subluxation: a case report. J Orthop Sci. 2002;7(6PG–703–706):703–706.

[32] Lyons C, Ross M, Elliott R, et al. Atlantoaxial instability in a patient with neck pain and ankylosing spondylitis. Mil Med. 2018;183(9/10):e654–e657.

[33] Mourad F, Giovannico G, Maselli F, et al. Basilar impression presenting as intermittent mechanical neck pain: a rare case report. BMC Musculoskelet Disord. 2016;17:1–5.

[34] Rebbeck T, Liebert A. Clinical management of cranio-vertebral instability after whiplash, when guidelines should be adapted: A case report. Man Ther. 2014;19(6PG–618–621):618–621.

[35] Derrick LJ, Chesworth BM. Post-motor vehicle accident alar ligament laxity. J Orthop Sports Phys Ther. 1992;16(1 PG–6–11):6–11.

[36] van Duijn J, van Duijn AJ, Nitsch W. Orthopaedic manual physical therapy including thrust manipulation and exercise in the management of a patient with cervicogenic headache: a case report. J Man Manip Ther. 2007;15(1PG–10–24):10–24.

[37] Glynn PE, Cleland JA. Evidence-based approach to the physical therapy diagnosis and management of neck and upper extremity pain using cervical and thoracic spine thrust manipulation: a case report. J Man Manip Ther. 2006;14(3PG–30E–45E):30E–45E.

[38] Jayaseelan DJ, Tow NS. Cervicothoracic junction thrust manipulation in the multimodal management of a patient with temporomandibular disorder. J Man Manip Ther. 2016;24(2PG–90–97):90–97.

[39] Lowry CD, O'Hearn MA, Courtney CA. Resolution of whiplash-associated allodynia following cervicothoracic thrust and non-thrust manipulation. Physiother Theory Pract. 2011;27(6PG–451–459):451–459.

[40] Mathers KS, Schneider MTM. Occult hypermobility of the craniocervical junction: A case report and review. J Orthop Sports Phys Ther. 2011;41(6):444–457.

[41] Naimi C The Effects Of Neuromobilization Combined With Posture Training In The Management Of A Patient With Cervical Radiculopathy: A Case Report. 2015

[42] Rodeghero J, Smith A, Russell J. Role of manual physical therapy and specific exercise intervention in the treatment of a patient with cervicogenic headaches: a case report. J Man Manip Ther. 2006;14(3PG–159–167):159–167.

[43] Sillevis R, Wyss K The management of a positional default of atlas in a patient with cervicogenic headache: A case report. (PG-).

[44] Willett GM, Wachholtz NA. A patient with internal carotid artery dissection. Phys Ther. 2011;91(8PG–1266–1274):1266–1274.

[45] Wong ML, Rossi MD, Groff W, et al. Physical therapy management of a patient with Eagle syndrome. Physiother Theory Pract. 2011;27(4PG–319–327):319–327.

[46] Young BA, Ross MD. Neck pain and headaches in a patient after a fall. J Orthop Sports Phys Ther. 2009;39(5PG–418):418.

[47] Selby KA, Newton RW, Gupta S, et al. Clinical predictors and radiological reliability in atlantoaxial subluxation in Down's syndrome. Arch Dis Child. 1991;66 (7):876–878.