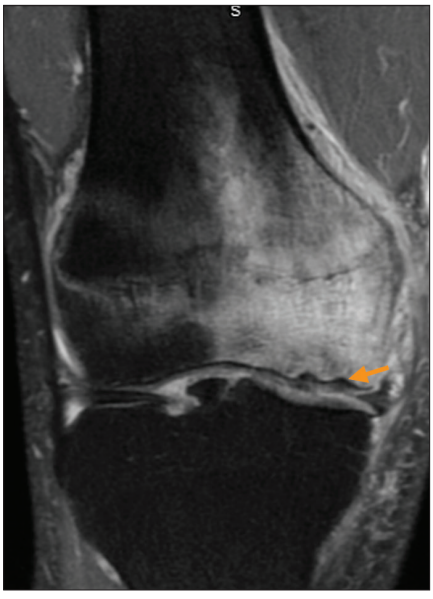


FIGURE 1. Coronal, fat-suppressed, proton density–weighted magnetic resonance image of the right knee, demonstrating subchondral bone collapse (arrow), bone marrow edema, and osteopenia of the medial femoral condyle extending into the metaphysis.

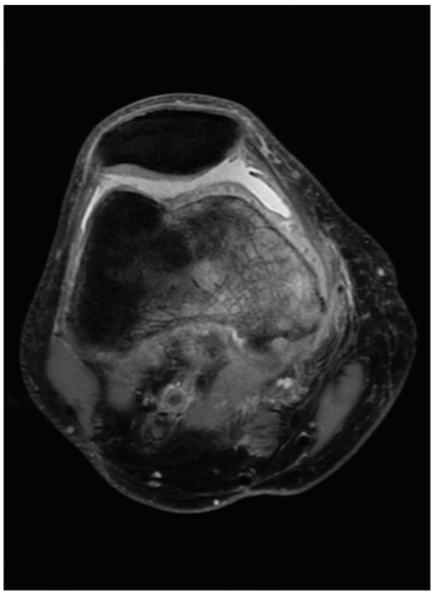


FIGURE 2. Axial, fat-suppressed, proton density–weighted magnetic resonance image of the right knee, demonstrating significant bone marrow edema and osteopenia of the medial femoral condyle.

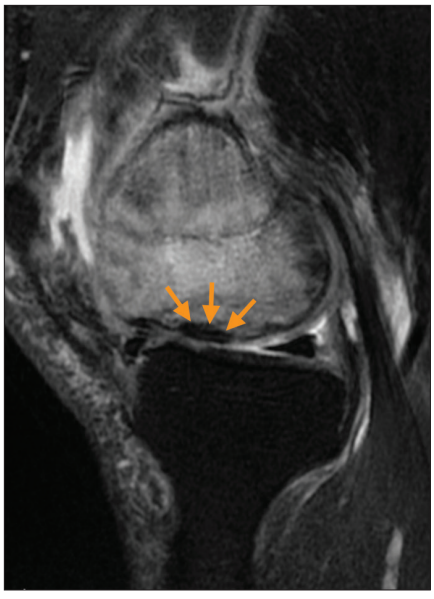


FIGURE 3. Sagittal, fat-suppressed, T2-weighted magnetic resonance image of the right knee, demonstrating cortical irregularity of the medial femoral condyle and curvilinear hypointensity at the articular surface (arrows).

Spontaneous Osteonecrosis of the Knee

JUSTIN BITTNER, PT, DPT, OCS, *Outpatient Therapy Services, Western Maryland Health System, Cumberland, MD.*
AARON HARTSTEIN, PT, DPT, OCS, FAAOMPT, *Shenandoah University Doctoral Program in Physical Therapy, Winchester, VA.*

A 69-year-old man was referred to physical therapy by his primary care physician for an insidious onset of right medial knee pain that had been present for over 4 weeks. Primary complaints included knee pain of 6/10 on the numeric pain-rating scale, swelling, and limited knee range of motion (ROM). He reported full but painful function with activities of daily living, recreational hiking, and his gym routine consisting of aerobic and resistance training (5 days per week). His past medical history was unremarkable.

Physical examination revealed joint effusion, limited flexion and extension ROM, quadriceps weakness, and limited weight-bearing tolerance consistent with a clinical diagnosis of osteoarthritis. He was instructed in non–weight-bearing lower extremity strengthening and ROM exercises and in self-mobilization techniques.

One week later, symptoms worsened, with a numeric pain-rating scale score of 8/10, decreased weight-bearing tolerance, and warmth to palpation, without constitutional signs or symptoms. Due to the disproportionate response following seemingly benign initial management, he was referred to an orthopaedist for imaging, with suspicion of an articular lesion.

Magnetic resonance imaging revealed medial femoral condyle osteopenia, bone marrow edema, and subchondral bone collapse (**FIGURES 1** and **2**). He was diagnosed with spontaneous osteonecrosis of the knee, with associated impaction fracturing (**FIGURE 3**).

Four weeks of non–weight bearing and immobilization in extension, as recommended by the orthopaedist, produced a significant reduction in pain with functional activities. Currently, he has resumed gym routines 3 days per week and performs activities of daily living pain free.

Clinicians should be aware of rare differential diagnoses of common knee pain presentation that worsen after initial conservative treatment. In this case, worsening symptoms and appropriate clinical decision making led to magnetic resonance imaging, which showed spontaneous osteonecrosis of the knee and altered the clinician's management of the patient. ◉ *J Orthop Sports Phys Ther 2018;48(10):824. doi:10.2519/jospt.2018.7923*

PAVLOS BOBOS, PT, MSc[1] • **JOY C. MACDERMID,** PT, PhD[2-4] • **DAVID M. WALTON,** PT, PhD[5]
ANITA GROSS, PT, MSc[4] • **P. LINA SANTAGUIDA,** PT, PhD[6]

Patient-Reported Outcome Measures Used for Neck Disorders: An Overview of Systematic Reviews

BACKGROUND: The evaluation of patient-reported outcome measures for the neck from multiple systematic reviews will provide a broader view of, and may identify potential conflicting or consistent results for, their psychometric properties.

OBJECTIVES: The purpose of this study was to conduct an overview of systematic reviews and synthesize evidence to establish the current state of knowledge on psychometric properties of patient-reported outcome measures for patients with neck disorders.

METHODS: In this overview of systematic reviews, an electronic search of 6 databases (MEDLINE, Embase, CINAHL, ILC, the Cochrane Central Register of Controlled Trials, and LILACS) was conducted to identify reviews that addressed at least one measurement property of outcome measures for people with neck pain. Only systematic reviews with patient-reported outcome measures were included in the analysis. Risk of bias was assessed with A MeaSurement Tool to Assess systematic Reviews (AMSTAR). Data on measurement properties were extracted from each systematic review.

RESULTS: From 13 systematic reviews, 8 patient-reported outcome measures were evaluated in 2 or more reviews. Risk-of-bias scores ranged from moderate (5-7) to high (4 and lower). Findings on internal consistency, test-retest reliability, construct validity, responsiveness to change, and content and structural validity were synthesized for the Neck Disability Index (NDI) in 11 systematic reviews; the Northwick Park Neck Pain Questionnaire and Neck Pain and Disability scale (NPDS) in 6 systematic reviews; the Copenhagen Neck Functional Disability Scale in 5 systematic reviews; the Neck Bournemouth Questionnaire in 4 systematic reviews; the Core Neck Pain Questionnaire and Patient-Specific Functional Scale in 3 systematic reviews, and the Whiplash Disability Questionnaire in 2 systematic reviews.

CONCLUSION: High-quality evidence was found of good to excellent internal consistency and moderate to excellent test-retest reliability for the NDI. Moderate-quality evidence was found of good to excellent internal consistency and good test-retest reliability for the Northwick Park Neck Pain Questionnaire. High-quality evidence was found of excellent test-retest reliability and good to strong construct validity with pain scales for the Copenhagen Neck Functional Disability Scale. Moderate-quality evidence was found of unclear to excellent internal consistency and moderate to strong concurrent associations with the NDI and global assessment of change for the Neck Pain and Disability scale. Moderate-quality evidence was found of excellent internal consistency for the Whiplash Disability Questionnaire and of high test-retest reliability for the Patient-Specific Functional Scale. *J Orthop Sports Phys Ther 2018;48(10):775-788. Epub 22 Jun 2018. doi:10.2519/jospt.2018.8131*

KEY WORDS: *neck pain, overview, psychometric, questionnaires*

Neck pain has been associated with high disability and increased health care cost, and it is considered a major musculoskeletal burden.[4] About 70% of adults experience an event of neck pain during their life span, and many of those are chronic or recurring.[9,20] People with neck pain regard it as a substantial burden that limits their daily activity or work, reporting recurrent cycles of pain and disability that extend for short or long periods.[23]

Diagnostic imaging is rarely able to identify the cause of neck pain after serious pathology has been ruled out, and only weak associations between imaging and patient symptoms have been found.[13,15] As a result, patient-reported outcome measures have recently been proposed as a critical component of evaluating and monitoring spinal musculoskeletal pain.[3,40] However, there is currently little consistency in selection or implementation of patient-reported outcome measures for neck pain.[38] An international survey revealed a variety of patient-report and impairment measures used inconsistently by

[1]Western Bone and Joint Institute, Department of Health and Rehabilitation Sciences, School of Physical Therapy, Western University, London, Canada. [2]Physical Therapy and Surgery, Western University, London, Canada. [3]Clinical Research Lab, Hand and Upper Limb Centre, St Joseph's Health Care London, London, Canada. [4]School of Rehabilitation Sciences, McMaster University, Hamilton, Canada. [5]Department of Health and Rehabilitation Sciences, School of Physical Therapy, Western University, London, Canada. [6]Department of Clinical Epidemiology and Biostatistics and Department of Physiotherapy, School of Rehabilitation Sciences, McMaster University, Hamilton, Canada. Mr Bobos is supported by a Western Graduate Research Scholarship and an Ontario Graduate Scholarship. Dr MacDermid is supported by a Canadian Institutes of Health Research Chair in Measurement and Rehabilitation of Musculoskeletal Work Disability and the Dr James Roth Research Chair in Musculoskeletal Measurement and Knowledge Translation. This study was funded by the Canadian Institutes of Health Research with financial reference numbers FRN: 102084 and FRN: 117420. Address correspondence to Pavlos Bobos, DB-220, Clinical Research Lab, Hand and Upper Limb Centre, St Joseph's Health Care London, 268 Grosvenor Street, London, ON N6A 4L6 Canada. E-mail: pbobos@uwo.ca

health care providers. Of those, evidence to support their measurement properties was often lacking.[39]

Neck pain is associated with a range of impairments and disabilities, from pain and stiffness through balance, cognitive, and emotional dysfunction.[8,26,36] Individual studies typically provide a limited view of the full breadth of the experience of neck pain. As a first step toward creation of a core set of outcome measures (patient-reported and observational), the current evidentiary base needs to be synthesized. The evaluation of psychometric properties of patient-reported outcome measures from multiple systematic reviews may provide a broader view, and identify conflicting or consistent results, of the current knowledge in the field. The development of critical appraisal tools designed specifically for psychometric studies has supported the recent emergence of systematic reviews in this area of research. Evidence-based recommendations on patient-reported outcome measures should assist researchers and clinicians when selecting appropriate tools. The purpose of this study was to conduct an overview of systematic reviews and synthesize evidence to establish the current state of knowledge on psychometric properties of patient-reported outcome measures for patients with neck disorders.

METHODS

Study Design

An overview of systematic reviews is a procedure to synthesize findings from systematic reviews.[10] An overview of systematic reviews can be useful when a broader array of literature needs to be summarized and there is a substantial pool of systematic reviews addressing relevant research. The detailed methods of our overview of systematic reviews are published in a separate paper[30] and summarized below.

Search Methods

A search for reviews published between January 2000 and September 2017 was conducted in the following databases without language restriction: MEDLINE, CINAHL, Embase, ILC, LILACS, and the Cochrane Central Register of Controlled Trials. The search strategy was designed to locate a systematic review that addressed at least one measurement property of an outcome measure in patients with neck pain or musculoskeletal neck conditions. The search strategy, including key words and Boolean operators, is shown in the **APPENDIX** (available at www.jospt.org).

Study Selection

Retrieved articles were entered into DistillerSR software (Evidence Partners, Ottawa, Canada) and reviewed independently by 2 authors. Titles and abstracts were reviewed, and articles were included for full-text review if they assessed at least 1 outcome measure for neck pain and at least 1 of the following psychometric properties: validity, reliability, responsiveness, Rasch analysis, factor analysis, cross-cultural validation, interpretability, and floor/ceiling effect. Reviews that addressed clinician-based outcomes and performance-based tests only were excluded from the review.

Risk-of-Bias Assessment

Two review authors applied A MeaSurement Tool to Assess systematic Reviews (AMSTAR),[33] a tool for the assessment of the risk of bias of systematic reviews. The AMSTAR is composed of 11 items and has adequate support for face and content validity[33] for measuring the quality of systematic reviews. Reviews that achieved a score of 8 or higher on the 11-point AMSTAR scale were considered to have low risk of bias, moderate risk of bias was assigned for scores ranging between 5 and 7, and high risk of bias for scores of 4 or less.[33] Cohen's kappa was used to measure agreement between the 2 raters. A score of 0.60 to 0.80 is considered to be good agreement. Discrepancies were resolved by consensus.

Data Extraction

The standardized data-extraction form was adapted from previous systematic reviews of outcome measures.[7,14,21,28,29,37] Two review authors were trained and calibrated on the use of the data-extraction form. Data extraction was completed by a single person (P.B.) and checked by a second (J.M.). Descriptive elements extracted from systematic reviews included the type of studies included in the reviews, sample size, the patient population, the purpose of the study, and the psychometric properties investigated. Also, data extracted from systematic reviews included recommendations made by different authors about improving the neck pain patient-reported outcome measures.

Data Synthesis

Due to heterogeneity across studies, a quantitative synthesis of the studies was not possible; therefore, a qualitative synthesis was conducted. *High-quality evidence* was defined as instances where similar findings were reported in at least 2 systematic reviews with low to moderate risk of bias; *moderate-quality evidence* was defined as that coming from 1 or more systematic reviews with moderate risk of bias and similar findings, with or without conflicting findings from high-risk-of-bias reviews; and *low-quality evidence* was defined as that occurring in 1 or more reviews with high risk of bias. *Conflicting evidence* was defined as when reviews of similar quality reported different conclusions. The strength of measurement properties reported in the evidence was based on COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) criteria ("poor," "fair," "good," or "excellent") for each separate measurement property (eg, reliability, concurrent validity). These were extracted and synthesized as reported in the included systematic reviews, relying on the judgment and reporting of the review authors, without extracting or rescoring data from the primary studies included in the review. Data were analyzed and presented from those patient-reported outcome measures that were included in 2 or more systematic reviews.

RESULTS

The results of the search and screening are included in the **FIGURE**. The electronic search resulted in a total of 490 articles. After the duplications and prior versions of updated systematic reviews were removed, 95 articles and accompanying abstracts were screened for inclusion in full-text review. Forty articles were retained and assessed at a full-text level. Eleven reviews were excluded at this stage because psychometric properties of at least one patient-reported outcome measure were not evaluated. At the end of the selection procedure, 13 reviews were eligible for extraction and, from these, 26 different patient-reported outcome measures were evaluated in the systematic reviews (**TABLE 1**), sorted from greatest to lowest frequency of evaluation. All the included reviews targeted patients with neck pain and incorporated a wider testing population, such as cervical radiculopathy, cervical myelopathy, whiplash, nonspecific neck pain, chronic neck pain, and neck dysfunction. The characteristics of the included reviews are described and summarized in **TABLE 2**. One systematic review[6] did not include additional details about the tested population and was assumed to be applicable to patients with both specific and nonspecific neck pain.

Risk-of-Bias Assessment of Systematic Reviews

All systematic reviews provided an a priori hypothesis, and 12 of the 13 systematic reviews conducted a comprehensive search by including multiple databases. More than half of the reviews reported duplicate study selection and data extraction. Eleven reviews provided a list with the characteristics of the included studies. The ninth component of AMSTAR, which describes the methods that were used to combine the findings of the studies, was rated as "not applicable" for all the included reviews. None of the systematic reviews assessed publication bias, and none of them provided sources of support in both the systematic review

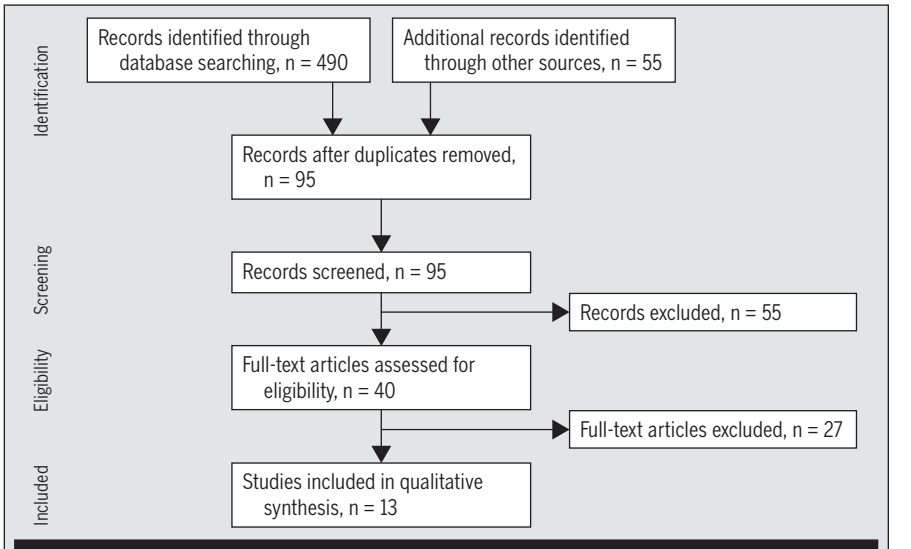


FIGURE. PRISMA flow diagram of citations screened and eligible studies for this overview.

TABLE 1 Neck PROMs Evaluated in the Included Systematic Reviews

Measure	Reviews, n
Neck Disability Index	11
Northwick Park Neck Pain Questionnaire	6
Neck Pain and Disability scale	6
Copenhagen Neck Functional Disability Scale	5
Neck Bournemouth Questionnaire	5
Patient-Specific Functional Scale	4
Core Neck Pain Questionnaire	2
Whiplash Disability Questionnaire	2
Core Outcome Measure Index	2
Core Whiplash Outcome Measure	1
Functional Rating Index	1
Short Core Neck Pain Questionnaire	1
NHANES ADL Scale	1
Current Perceived Health-42 Questionnaire	1
Global Assessment of Neck Pain	1
Myelopathy Disability Index	1
Japanese Orthopaedic Association Scale	1
MOS 36-Item Short-Form Health Survey	1
NASS Lumbar Spine Outcome Assessment	1
Pain Disability Index	1
McGill Pain Questionnaire	1
Visual analog scale	1
Numeric pain-rating scale	1
Neck and Upper Limb Index	1
DASH	1
QuickDASH	1

Abbreviations: ADL, activities of daily living; DASH, Disabilities of the Arm, Shoulder and Hand questionnaire; MOS, Medical Outcomes Study; NASS, North American Spine Society; NHANES, National Health and Nutrition Examination Survey; PROM, patient-reported outcome measure; QuickDASH, shortened version of the Disabilities of the Arm, Shoulder and Hand questionnaire.

and the included studies. The agreement (Cohen's kappa) between the 2 raters on quality was very good (κ = 0.89). Among the systematic reviews (n = 13), 9 had moderate risk of bias and 4 had high risk of bias. **TABLE 3** shows the details of the AMSTAR ratings.

Properties of Specific Patient-Reported Outcome Measures

Neck Disability Index Psychometric properties of the original version of the Neck Disability Index (NDI) were summarized in 4 reviews with moderate bias[11,14,24,32] and 3 reviews with high risk of bias[6,17,27] (**TABLE 4**). Two reviews with moderate risk of bias consistently reported internal consistency (reported as Cronbach's alpha) as moderate to high. Schellingerhout and colleagues[32] (moderate quality) reported the alpha range to be .87 to .92, while MacDermid and colleagues[14] (moderate quality) reported it to be .70 to .93. Only 1 high-risk-of-bias review reported alpha values, of .79 to .93.[6]

Test-retest reliability was reported in 5 systematic reviews, 3 of moderate risk[11,14,32] and 2 of high risk of bias.[6,27] Findings were consistent across the reviews that test-retest reliability (most commonly indicated by the intraclass correlation coefficient [ICC]) was moderate (ICC = 0.50; Schellingerhout et al[32]) to excellent (ICC = 0.50-0.98; MacDermid et al[14]). Two high-risk-of-bias reviews[6,27] reported test-retest reliability as moderate to high. More specifically, Rodine and Vernon[27] reported an ICC range of 0.30 to 0.90, Holly and colleagues[11] reported an ICC of 0.68, and Ferreira and colleagues[6] reported an ICC range of 0.62 to 0.97. Minimum detectable change (MDC) was reported by Holly et al[11] (moderate risk) and MacDermid et al[14] (moderate risk) as 10 to 10.2 points for patients with cervical radiculopathy.

Three reviews[14,24,32] of moderate risk of bias consistently reported moderate to strong (r>0.53) concurrent associa-

TABLE 2 Characteristics of the Eligible Systematic Reviews

Study/Risk of Bias	Studies, n	Study Population*	Patient-Reported Outcomes
Alreni et al[1] Moderate	11	Workers with neck and upper-limb dysfunction, shoulder and/or arm pain	NULI, DASH, QuickDASH
Pellicciari et al[22] Moderate	66	Neck pain (acute, subacute, and chronic), mechanical neck pain, neck complaints	Italian translated versions of the NDI, NPDS, NBQ, COMI, and NeckPix
Yao et al[42] Moderate	24	Neck pain with or without radicular findings with degenerative joint or disc disease, neck pain with or without WAD	Translated versions of the NDI (Arabic, Catalan, Chinese, Finnish, French, Greek, Persian, Japanese, Korean, Polish, Spanish, Thai, and Turkish)
Murphy and Lopez[19] High	10	Chronic neck pain, nonspecific neck pain, posttraumatic neck pain	Spanish versions of the NPQ, COMI, and NDI
Schellingerhout et al[32] Moderate	25	Nonspecific and mechanical neck pain, cervical radiculopathy, WAD	NDI, NPDS, NBQ, NPQ, WDQ, CNFDS, CNPQ, CWOM
Rodine and Vernon[27] High	6	Cervical radiculopathy	NDI
Horn et al[12] Moderate	13	Cervical radiculopathy, neck dysfunction	PSFS
Schellingerhout et al[31] Moderate	27	Neck pain	Translated versions of the NDI, NPDS, NPQ, NBQ, CNFDS, CNPQ
Misailidou et al[17] High	86	Neck pain, chronic neck pain, nonspecific neck pain	NBQ, CNFDS, NPDS, NDI, NPQ, PSFS, WDQ
Ferreira et al[6] High	18	Neck pain	NDI, NPQ, CNFDS, NPDS, NBQ
Holly et al[11] Moderate	11	Cervical radiculopathy	PSFS, NDI
MacDermid et al[14] Moderate	36	Nonspecific neck pain, chronic neck pain	NDI
Pietrobon et al[24] Moderate	15	Mechanical neck pain, WAD, chronic neck pain, neck dysfunction	NDI, PSFS, CNFDS, NPQ, NPDS

Abbreviations: CNFDS, Copenhagen Neck Functional Disability Scale; CNPQ, Core Neck Pain Questionnaire; COMI, Core Outcome Measure Index; CWOM, Core Whiplash Outcome Measure; DASH, Disabilities of the Arm, Shoulder and Hand questionnaire; NBQ, Neck Bournemouth Questionnaire; NDI, Neck Disability Index; NPDS, Neck Pain and Disability scale; NPQ, Northwick Park Neck Pain Questionnaire; NULI, Neck and Upper Limb Index; PSFS, Patient-Specific Functional Scale; QuickDASH, shortened version of the Disabilities of the Arm, Shoulder and Hand questionnaire; WAD, whiplash-associated disorder; WDQ, Whiplash Disability Questionnaire.
**Included study populations within the eligible studies of the review.*

tions between the NDI and pain intensity using a visual analog scale (VAS) or the McGill Pain Questionnaire. One review[27] with high risk of bias also found strong associations with the shortened version of the Disabilities of the Arm, Shoulder and Hand (QuickDASH) questionnaire (r = 0.83), and 1 high-risk-of-bias review[6] found moderate (at least r = 0.70) associations with the McGill Pain Questionnaire. Only 1 moderate-risk-of-bias review, by Holly et al,[11] reported that the NDI failed to demonstrate construct validity.

Two reviews with moderate risk of bias[14,32] summarized responsiveness, as indicated by area under the receiver operating characteristic curve (AUC), for discriminating between those with and without meaningful change. Those 2 systematic reviews reported an AUC of 0.79 from the same primary study. However, the most recent systematic review[32] reported 2 primary studies, one with moderate responsiveness (AUC = 0.79) and the other with low responsiveness (AUC = 0.57).

Properties of translated versions of the NDI were evaluated and summarized in 3 moderate-risk-of-bias reviews[22,32,42] and 1 high-risk-of-bias review.[19] Three systematic reviews[22,31,42] (moderate risk of bias) summarized the psychometric properties of 14 different translations (Arabic, Catalan, Chinese, Finnish, French, Greek, Persian, Italian, Japanese, Korean, Polish, Spanish, Thai, and Turkish) and reported moderate to strong internal consistency (α = .70-.90) and good test-retest reliability (ICC>0.70) for all translations. The high-risk-of-bias systematic review[19] explored the properties of the Spanish version and found that translation to have excellent test-retest reliability (ICC = 0.97). Concurrent associations of the translated versions with pain intensity were consistent with those of the English version only (r = 0.64).[19]

Our findings indicate high-quality evidence of good to excellent internal consistency, high-quality evidence of moderate to excellent test-retest reliability, moderate- to high-quality evidence of moderate to good cross-sectional convergent validity with related constructs (pain and upper extremity function), moderate-quality evidence of poor to moderate responsiveness, and high-quality evidence that the internal consistency and test-retest reliability of up to 13 translations share similar properties with those of the original English version.

Northwick Park Neck Pain Questionnaire The psychometric properties of the original version of the Northwick Park Neck Pain Questionnaire (NPQ) were reported in 4 systematic reviews, 2 of moderate risk of bias[24,32] and 2 of high risk of bias[6,17] (**TABLE 5**). Only 1 high-risk-of-bias review[6] reported an alpha value (.79). Two reviews[6,24] reported on test-retest reliability, with an ICC of 0.62 and a kappa coefficient of 0.62 (range, 0.53-0.73). Only 1 review[32] reported a positive correlation (r = 0.56) between the NPQ and the problem elicitation technique. In the same review, a moderately positive finding was reported for responsiveness

TABLE 3 Risk-of-Bias Assessment of the Systematic Reviews That Examined the Psychometric Properties of the Neck Pain Patient-Reported Outcomes*

	AMSTAR Items†											
Study	1	2	3	4	5	6	7	8	9	10	11	Quality
Alreni et al[1]	Y	Y	Y	N	Y	Y	Y	Y	NA	N	N	Moderate
Pellicciari et al[22]	Y	Y	Y	N	Y	Y	Y	Y	NA	N	N	Moderate
Yao et al[42]	Y	Y	Y	Y	Y	Y	N	Y	NA	N	N	Moderate
Murphy and Lopez[19]	Y	N	Y	N	N	Y	N	NA	NA	N	N	High
Schellingerhout et al[32]	Y	Y	Y	N	N	Y	Y	Y	NA	N	N	Moderate
Horn et al[12]	Y	Y	Y	Y	N	Y	Y	Y	NA	N	N	Moderate
Rodine and Vernon[27]	Y	N	N	N	N	Y	N	NA	NA	N	N	High
Schellingerhout et al[31]	Y	Y	Y	N	N	Y	Y	Y	NA	N	N	Moderate
Misailidou et al[17]	Y	N	Y	N	N	N	N	NA	NA	N	N	High
Ferreira et al[6]	Y	CA	Y	N	N	N	N	NA	NA	N	N	High
Holly et al[11]	Y	N	Y	N	N	Y	Y	Y	NA	N	N	Moderate
MacDermid et al[14]	Y	Y	Y	N	N	Y	Y	Y	NA	N	N	Moderate
Pietrobon et al[24]	Y	N	Y	Y	N	Y	N	NA	NA	N	N	High

Abbreviations: AMSTAR, A MeaSurement Tool to Assess systematic Reviews; CA, can't assess; N, no; NA, not applicable; Y, yes.
**Scores of 8 or higher were considered as low risk of bias, scores of 5 through 7 as moderate risk of bias, and scores of 4 or less as high risk of bias.*
†1, Was an a priori design provided? 2, Was there duplicate study selection and data extraction? 3, Was a comprehensive literature search performed? 4, Was the status of publication (ie, gray literature) used as an inclusion criterion? 5, Was a list of studies (included and excluded) provided? 6, Were the characteristics of the included studies provided? 7, Was the scientific quality of the included studies assessed and documented? 8, Was the scientific quality of the included studies used appropriately in formulating conclusions? 9, Were the methods used to combine the findings of studies appropriate? 10, Was the likelihood of publication bias assessed? 11, Was the conflict of interest included?

(r = 0.60), without specifying further details. No floor or ceiling effects were detected. Minimum important change was reported as unclear, and differences in scores between subgroups have not been evaluated. Responsiveness was reported as 0.93 (large) from a single high-risk-of-bias review.[6]

Two additional systematic reviews[19,31] of high[19] and moderate[31] risk of bias reported the psychometric properties of translated versions of the NPQ. The review with high risk of bias[19] reported the test-retest reliability of the Spanish version (ICC = 0.63). The review with moderate risk of bias[31] reported that the Chinese and Turkish versions showed excellent and fair reliability, respectively. Chinese, Turkish, French, and Spanish versions were reported to have poor internal consistency and large measurement error.

A correlation between the VAS and the Spanish NPQ was reported by a review with high risk of bias[19] (r = 0.51), which also found a retest association of r = 0.74 between days 1 and 15. Chinese, French, and Spanish versions were rated as having poor content and structural validity and poor responsiveness, and hypothesis testing ranged from poor to fair as reported in a review with moderate risk of bias.[31]

Our findings indicate moderate-quality evidence of good to excellent internal consistency; moderate-quality evidence of good test-retest reliability; limited

TABLE 4 Summary of the Psychometric Properties of the NDI in Included Systematic Reviews

Study/Risk of Bias	Reliability	Validity	Responsiveness	Other Reports
Pellicciari et al[22] Moderate	Internal consistency was found to be high, ranging from .72 to .99, with ICC values ranging from 0.81 to 0.99; 2 other studies reported very low reliability values	Moderate to strong correlations with the VAS for pain, NRS, SF-36, and other disability PROMs (eg, NPDS and NBQ)	MCID ranging from 3.5 to 9.5 points on a 50-point scale in studies of excellent quality, but the MDC showed very large variability, ranging from 1.66 to 23.3 points, in studies of fair quality	For reliability, all studies were of poor to fair quality, and no firm conclusions can be drawn
Yao et al[42] Moderate	Cronbach α = .70-.90; ICC>0.70	Content validity	No reports	Japanese version found to show floor or ceiling effect (direction not specified)
Murphy and Lopez[19] High	Cronbach α = .93-.94; ICC = 0.97	Construct validity of the NDI and VAS; r = 0.64-0.74	No reports	NA
Schellingerhout et al[32] Moderate	Cronbach α = .87-.92; MIC was undetermined; ICC = 0.50	Good correlation (r = 0.53-0.70) with pain and/or physical functioning	AUC = 0.79 and 0.57	No floor and ceiling effects were detected
Rodine and Vernon[27] High	ICC = 0.68, 0.30, 0.90	QuickDASH and NDI: r = 0.83	No reports	NA
Schellingerhout et al[31] Moderate	ICC>0.70	Inconsistent results	Inconsistent results	NA
Misailidou et al[17] High	Good to high reliability	Strong correlation	Medium responsiveness	NA
Ferreira et al[6] High	Cronbach α = .79-.93; ICC = 0.62-0.97	r = 0.6 (pain VAS), r = 0.7 (MPQ)	No reports	The NDI had all its sections classified into the ICF, with 4 items categorized as body functions and structures and 6 items classified as activity and participation
Holly et al[11] Moderate	ICC = 0.68; MDC, 10.2	NDI failed to demonstrate construct validity	...	No reports
MacDermid et al[14] Moderate	Cronbach α = .70-.93; ICC = 0.50-0.98	Strongly correlated (>0.70) to several similar indices and moderately related to physical and mental aspects of general health	AUC = 0.79	MDC is 5/50 for uncomplicated neck pain and up to 10/50 for cervical radiculopathy. The CID is inconsistent, ranging from 5/50 to 19/50
Pietrobon et al[24] Moderate	No report	r = 0.6 between the NDI and VAS, r = 0.7 between the NDI and MPQ	No reports	NA

Abbreviations: AUC, area under the curve; CID, clinically important difference; ICC, intraclass correlation coefficient; ICF, International Classification of Functioning, Disability and Health; MCID, minimum clinically important difference; MDC, minimum detectable change; MIC, minimum important change; MPQ, McGill Pain Questionnaire; NA, not applicable; NBQ, Neck Bournemouth Questionnaire; NDI, Neck Disability Index; NPDS, Neck Pain and Disability scale; NRS, numeric rating scale; PROM, patient-reported outcome measure; QuickDASH, shortened version of the Disabilities of the Arm, Shoulder and Hand questionnaire; SF-36, Medical Outcomes Study 36-Item Short-Form Health Survey; VAS, visual analog scale.

findings of moderate-quality evidence of good construct validity; limited findings of low-quality evidence of responsiveness to change; low-quality evidence of internal consistency of the Spanish version and low-quality evidence of fair to excellent reliability for Chinese, Turkish, and French versions; and limited findings of moderate-quality evidence of poor content and structural validity for Spanish, Chinese, and French versions of the NPQ.

Copenhagen Neck Functional Disability Scale Five systematic reviews reported the psychometric properties of the original version of the Copenhagen Neck Functional Disability Scale (CNFDS), 2 of moderate[24,32] and 2 of high[6,17] risk of bias (**TABLE 6**). One high-risk-of-bias review[6] reported internal consistency (Cronbach α = .90-.92) and test-retest reliability (ICC = 0.83). One moderate-risk-of-bias review[24] reported test-retest reliability with short-term (same day) and between-day (evaluated by mail 2 days later) analyses, achieving Pearson correlation coefficients of 0.99 and 0.98, respectively.

Consistent findings from 2 reviews with moderate risk of bias[24,32] and 1 review with high risk of bias[6] reported a good correlation ($r \geq 0.64$) with pain scales, such as the VAS and numeric pain-rating scale (NPRS). A single high-risk-of-bias review[6] reported a correlation between the CNFDS and the doctor global assessment of change, with a value of r = 0.56.

A single high-risk-of-bias review[6] reported a small effect size for detecting change over 6 weeks (0.49), a small effect over 4 months (0.48), and a moderate effect at 12 months (0.54).

Only 1 systematic review[31] with moderate risk of bias provided the psychometric properties of English, French, and Turkish versions of the CNFDS. The English and French versions were reported to have poor internal consistency, reliability, and measurement error, while the Turkish version was found to have fair reliability. Poor validity and fair responsiveness were reported for the Turkish and English versions. The quality of the translation process was underreported or not described.

Our findings indicate low-quality evidence of excellent internal consistency; high-quality evidence of excellent test-retest reliability; high-quality evidence of good to strong construct validity with pain scales; limited findings of low-quality evidence of moderate responsiveness to change; and limited findings of moderate-quality evidence of poor internal consistency, reliability, measurement error, validity, and responsiveness for the translated versions (English, Turkish, and French).

Neck Pain and Disability Scale The psychometric properties of the original version of the Neck Pain and Disability scale (NPDS) were examined in 4 systematic reviews (**TABLE 7**).[6,17,24,32] Two systematic reviews[6,17] with high risk of bias reported conflicting findings regarding internal consistency. Schellingerhout et

TABLE 5 Summary of the Psychometric Properties of the NPQ in Included Systematic Reviews

Study	Reliability	Validity	Responsiveness	Other Reports
Murphy and Lopez[19]	Test-retest reliability was good (ICC = 0.63)	Correlation between pain (VAS) and the NPQ score was acceptable in the test (r = 0.51) and retest (r = 0.74)	Not reported	Not reported
Schellingerhout et al[32]	No methodologically sound studies evaluating the internal consistency, measurement error, or reliability	No methodologically sound studies evaluating the content validity or structural validity. Positive correlation (r = 0.56) between the NPQ and PET	There is moderate positive evidence for responsiveness (r = 0.60)	No floor or ceiling effects were detected. Differences in scores between subgroups were not evaluated. The MIC is unclear
Schellingerhout et al[31]	Chinese, Turkish, French, and Spanish versions: poor internal consistency and measurement error, excellent reliability (Chinese version), fair reliability (Turkish version)	Chinese, French, and Spanish versions had poor content and structural validity	Poor responsiveness	Hypothesis testing was poor to fair
Misailidou et al[17]	Similar to NDI	Similar to NDI	Similar to NDI	Similar to NDI
Ferreira et al[6]	Internal consistency, Cronbach α = .79; test-retest reliability, ICC = 0.62	Not reported	Responsiveness to change, 0.93; large effect size	Not reported
Pietrobon et al[24]	Test-retest reliability with a 3-day interval between tests had a kappa coefficient of 0.62 (range, 0.53-0.73)	Not reported	Total scores correlated linearly with the question for pain from 3 days to 1 month, but correlation was not linear between 1 and 3 months	Not reported

Abbreviations: ICC, intraclass correlation coefficient; MIC, minimum important change; NDI, Neck Disability Index; NPQ, Northwick Park Neck Pain Questionnaire; PET, problem elicitation technique; VAS, visual analog scale.

al[32] reported no methodologically sound studies for internal consistency, while Pietrobon et al[24] reported internal consistency of α = .93. A high-risk-of-bias systematic review[6] reported internal consistency of α = .93 and test-retest reliability of ICC = 0.97.

Two high-risk-of-bias[6,17] reviews reported correlations between the NPDS and NDI (r = 0.86) and between the NPDS and global assessment of change (r = 0.59). One moderate-risk-of-bias systematic review[32] reported the results of hypothesis-based (concurrent) validity testing (r = 0.52-0.78) and responsiveness (r = 0.59).

Two moderate-risk-of-bias systematic reviews[22,31] reported the psychometric properties of 9 translated versions of the NPDS (Dutch, Finnish, French, German, Italian, Hindi, Persian, Korean, Turkish). Only the Finnish, German, and Italian translations of the NPDS were recommended to be used in clinical practice.

In summary, we found moderate-quality evidence of unclear to excellent internal consistency, limited findings of low-quality evidence of excellent test-retest reliability, limited findings of moderate-quality evidence of moderate to strong ($r \geq 0.59$) concurrent associations between the NPDS and NDI and global assessment of change, and limited findings of moderate-quality evidence of translated versions of the NPDS (Finnish, German, and Italian) recommended by a moderate-quality systematic review to be used in clinical practice.

Whiplash Disability Questionnaire The psychometric properties of the Whiplash Disability Questionnaire were examined in 2 systematic reviews, 1 of moderate risk and 1 of high risk of bias (**TABLE 8**).[17,32]

One systematic review with moderate risk of bias[32] reported an internal consistency (Cronbach's alpha) of .95 to .96, and a systematic review with high risk of bias[17] found an ICC of 0.96 for test-retest reliability over time in patients with whiplash. Only 1 systematic review with moderate risk of bias[32] reported a responsiveness value (r = 0.67), without specifying what was being correlated. No floor or ceiling effects were detected, and information on other aspects of interpretability is lacking.

In summary, we found limited findings of moderate-quality evidence of excellent internal consistency, limited findings of low-quality evidence of excellent test-retest reliability, and limited findings of low-quality evidence of positive responsiveness.

Patient-Specific Functional Scale The psychometric properties of the Patient-Specific Functional Scale (PSFS) were reported in 3 systematic reviews, 1 with high risk of bias[17] and 1 of moderate risk of bias[12] (**TABLE 8**).

We found consistent findings from 2 reviews with moderate[11,12] of high test-retest reliability (ICC = 0.82) in patients with cervical radiculopathy.

There were consistent findings, from 2 reviews with moderate[11,12] risk of bias, of concurrent validity associations between PSFS activities and the NDI in patients with neck dysfunction (r = 0.82; lower 95% confidence limit, 0.68) and at discharge (r = 0.83; lower 95% confidence limit, 0.69). The PSFS scores (r = 0.64;

TABLE 6 Summary of the Psychometric Properties of the CNFDS in Included Systematic Reviews

Study	Reliability	Validity	Responsiveness	Other Reports
Schellingerhout et al[32]	No methodologically sound studies evaluating internal consistency, measurement error, or reliability	The CNFDS correlates with the NRS for pain (r = 0.64)	No methodologically sound studies evaluating responsiveness	There is no information available on floor effects, differences in scores between subgroups, or the MIC
Schellingerhout et al[31]	Poor internal consistency (English, French, and Turkish versions), reliability (fair for Turkish version), and measurement error	Poor validity	Fair responsiveness for Turkish and English versions	The translation process is not described, so the quality of this process is unknown
Misailidou et al[17]	Similar to NDI	Similar to NDI	Similar to NDI	Similar to NDI
Ferreira et al[6]	Internal consistency, Cronbach α = .90-.92 (excellent); test-retest reliability, ICC = 0.83 (excellent)	Pain NRS: excellent validity (patient global assessment; 0.89) and adequate validity (doctor global assessment; 0.56)	Responsiveness sensitivity to change: 0.49 at 6 weeks (small), not reported, 0.48 at 4 months (small), 0.54 at 12 months (moderate)	Not reported
Pietrobon et al[24]	Test-retest reliability was evaluated with short-term (same-day) and between-day (evaluated by mail 2 days later) reliability analysis, achieving Pearson correlation coefficients of 0.99 and 0.98, respectively	Pearson correlation coefficients were high for pain scores (r = 0.83) and for the patients' global assessment (r = 0.89). It was moderate for doctors' global assessment (r = 0.56)	Moderate correlations were demonstrated with pain scores at 6 weeks (r = 0.49), 4 months (r = 0.48), and 12 months (r = 0.54)	No quantification-estimated MCID was reported

Abbreviations: CNFDS, Copenhagen Neck Functional Disability Scale; ICC, intraclass correlation coefficient; MCID, minimum clinically important difference; MIC, minimum important change; NDI, Neck Disability Index; NRS, numeric rating scale.

lower 95% confidence limit, 0.40) had a similar strength relationship to that of the NDI scores at discharge (r = 0.66; lower 95% confidence limit, 0.40).[12]

Score change between the PSFS and the NPRS was r = 0.80 (P<.001). A high correlation between change scores of the PSFS and the global rating of change (r = 0.82, P<.001) in patients with cervical radiculopathy was also reported.[12]

Limited findings were reported for the MDC: 0.99 PSFS points for an average of 3 activities and 1.18 PSFS points for an individual activity. Limited findings were reported for the minimum clinically important difference in patients with cervical radiculopathy, ranging from 2.0 to 3.0 points.[12]

In summary, we found moderate-quality evidence of high test-retest reliability; low-quality evidence of strong concurrent associations between the PSFS and NDI, as well as between the PSFS and the NPRS and global rating of change; and low-quality evidence of MDC, as well as of the minimum clinically important difference.

Core Neck Pain Questionnaire The psychometric properties of the Core Neck Pain Questionnaire (CNPQ) were evaluated in 2 systematic reviews of moderate risk of bias (**TABLE 8**).[31,32] Limited findings were reported for reliability (ICC>0.70). Also, limited findings were reported for a positive correlation of the CNPQ with the NDI (r>0.60). No floor or ceiling effects were detected, and there was no information on other aspects of interpretability.

Only 1 systematic review[31] with moderate risk of bias reported a translated version of the CNPQ (in Spanish). Limited findings were reported of an excellent translation process, although internal consistency, responsiveness, and hypothesis testing were rated as poor.

In summary, we found limited findings of low-quality evidence of good test-retest reliability and limited findings of low-quality evidence (Spanish version of the CNPQ) of poor internal consistency, responsiveness, and hypothesis testing.

Neck Bournemouth Questionnaire Two systematic reviews[6,32] reported the psychometric properties of the original version of the Neck Bournemouth Questionnaire (**TABLE 9**).

One systematic review[32] with moderate risk of bias reported no methodologically sound studies for internal consistency, measurement error, reliability, and content or structural validity. A systematic review[6] with high risk of bias reported (Cronbach α = .86) high internal consistency and test-retest reliability (ICC = 0.91). Convergent validity was found to have a value of r = 0.69 (VAS).

Moreover, 2 systematic reviews[22,31] reported the psychometric properties of the Neck Bournemouth Questionnaire in

TABLE 7 Summary of the Psychometric Properties of the NPDS in Included Systematic Reviews

Study	Reliability	Validity	Responsiveness	Other Reports
Pellicciari et al[22]	Internal consistency was high, with Cronbach's alpha for the total score ranging from .86 to .97 The ICC values were above 0.75, but only in a few studies of lower quality did they exceed the minimum required value of 0.90	Strong correlations with the NPQ and NDI, moderate to strong correlations with the VAS, weak to moderate correlations with the SF-36, good face validity	The MCID was close to 10 points for the Italian version in a study of excellent quality (AUC = 0.91; sensitivity, 0.93; specificity, 0.83)	No floor or ceiling effects were found
Schellingerhout et al[32]	No methodologically sound studies evaluating the internal consistency, measurement error, or reliability	No content validity, 4-factor structure, hypothesis testing (r = 0.52-0.78)	Responsiveness, r = 0.59	No floor or ceiling effects were detected
Schellingerhout et al[31]	Translated versions	Translated versions	Translated versions	Translated versions
Misailidou et al[17]	No reports	Scores strongly correlated with the NDI (r = 0.86); good content validity	No reports	No reports
Ferreira et al[6]	Internal consistency, Cronbach α = .93 (excellent); test-retest reliability, ICC = 0.97 (excellent)	Adequate validity for patient global assessment of improvement (r = 0.59)	No reports	The NPDS had 11 items classified in the ICF as body functions and structures, 8 in activity and participation, and 1 in environmental factors
Pietrobon et al[24]	Internal consistency reliability, .93; test-retest reliability not reported	Validity scores compared between patients with neck pain or dysfunction and controls to establish face validity	No reports	No reports

Abbreviations: AUC, area under the curve; ICC, intraclass correlation coefficient; ICF, International Classification of Functioning, Disability and Health; MCID, minimum clinically important difference; NDI, Neck Disability Index; NPDS, Neck Pain and Disability scale; NPQ, Northwick Park Neck Pain Questionnaire; SF-36, Medical Outcomes Study 36-Item Short-Form Health Survey; VAS, visual analog scale.

Dutch, Italian, and French. The translation process of the Dutch version[31] was rated as "excellent" and the French version as "poor." Hypothesis testing was rated as "poor" for the Dutch version and "fair" for the French version. The reliability and measurement error were rated as "indeterminate" for the Dutch version. For the French version, the reliability was found to be "poor" and responsiveness was rated as "moderate." The Italian version of the Neck Bournemouth Questionnaire was found to have Cronbach alpha values ranging from .79 to .92 and moderate to strong correlations with the NDI, NPDS, and CNFDS.

In summary, we found limited findings of low-quality evidence of excellent internal consistency, low-quality evidence of unknown to excellent test-retest reliability, and limited findings of low-quality evidence of strong correlations with the pain VAS and with responsiveness to change.

DISCUSSION

A systematic synthesis of systematic reviews resulted in an overview of the current state of knowledge for 8 different neck-specific patient-reported outcome measures. Only the NDI and CNFDS were supported by high-quality evidence of internal consistency, and only the CNFDS was supported by high-quality evidence of test-retest reliability and construct validity. Consistent problems in reporting of psychometrics for many of the studies included the following: (1) the specific type of psychometric property was not reported, (2) the population was not specifically defined other than stating only neck pain, and (3) values of psychometric properties were not reported. While many properties drawn largely from classical test theory–based analyses have been reported for several of the tools, there is a lack of reporting of

TABLE 8 Summary of Psychometric Properties of the WDQ, PSFS, and CNPQ Reported in Systematic Reviews

Measure/Study	Reliability	Validity	Responsiveness	Other Reports
WDQ				
Schellingerhout et al[32]	Internal consistency, Cronbach α = .95-.96	Not reported	There is limited positive evidence for responsiveness (r = 0.67)	No floor or ceiling effects were detected; information on other aspects of interpretability is lacking
Misailidou et al[17]	Whiplash-specific ICC = 0.96	Not reported	Not reported	Not reported
PSFS				
Horn et al[12]	Test-retest reliability, ICC = 0.82	PSFS activity and the NDI in neck dysfunction (r = 0.82; lower 95% confidence limit, 0.68) and at discharge (r = 0.83; lower 95% confidence limit, 0.69). PSFS scores (r = 0.64; lower 95% confidence limit, 0.40) at discharge (r = 0.66; lower 95% confidence limit, 0.40) PSFS and NPRS (r = 0.80), PSFS scores and the global rating of change (r = 0.82 in cervical radiculopathy)	Responsiveness for neck dysfunction: the MDC was calculated to be 0.99 PSFS points for an average of 3 activities, and 1.18 PSFS points for an individual activity	The MCID for cervical radiculopathy was found to be between 2.0 and 3.0 points
Misailidou et al[17]	Not reported	Not reported	Not reported	Very sensitive to functional changes in individual patients
Holly et al[11]	The PSFS demonstrated excellent test-retest reliability (ICC = 0.82) for patients with cervical radiculopathy	The PSFS showed construct validity, as there was a significant difference in scores between stable and improved patients based on the global rating of change (P<.001)	Not reported	Not reported
CNPQ				
Schellingerhout et al[32]	The reliability of the total score of the CNPQ has not been studied, but 4 of the 6 items have an ICC greater than 0.70	There was a positive correlation of the CNPQ with the NDI (r>0.60)	Not reported	No floor or ceiling effects were detected; there is no information on other aspects of interpretability
Schellingerhout et al[31]	Poor internal consistency (Spanish version)	...	Poor responsiveness	Poor hypothesis testing and excellent translation process

Abbreviations: CNPQ, Core Neck Pain Questionnaire; ICC, intraclass correlation coefficient; MCID, minimum clinically important difference; MDC, minimum detectable change; NDI, Neck Disability Index; NPRS, numeric pain-rating scale; PSFS, Patient-Specific Functional Scale; WDQ, Whiplash Disability Questionnaire.

clinical responsiveness that will limit easy translation to practice.

Our synthesis found a moderate to high risk of bias in the reporting of many systematic reviews of patient-reported outcome measures for the neck, as described by the AMSTAR criteria. Common sources of bias included the following: (1) did not assess publication bias, (2) did not describe appropriate methods to combine findings, and (3) did not report sources of financial support in either the systematic review or the included primary studies. Several authors[6,32,42] have highlighted the importance of following predefined criteria and guidelines for the evaluation of psychometric properties of neck patient-reported outcomes. There are at least 3 checklists available for evaluation of the properties of patient-reported outcome measures, including the guidelines for the process of cross-cultural adaptation of self-report measures,[2,16] the COSMIN checklist,[18] and quality criteria for psychometric properties of health status questionnaires.[35] Authors of future reviews are strongly encouraged to make use of these tools to facilitate transparency and knowledge syntheses.

The selection of patient-reported outcome measures should be informed by what is known about the measurement properties of the available tools, and must go beyond statistics alone. Face validity, being the interpretation that items in a scale appear "on the face" to be tapping into correct constructs, is generally considered necessary but not sufficient to support the adequate performance of a tool; however, it can be quite important in real-world application.[18] Clinicians should consider questions such as, Do the items make sense? Can they likely be interpreted properly by my patients? Will the responses give me useful information for making clinical decisions? Clinicians should also consider the construct being captured (eg, function, symptom severity, well-being) and the information that can be gleaned from the response options (eg, intensity of experience, frequency of experience, opinion or agreement with a set of statements). One way to standardize patient-reported outcome measures by construct would be to map the items within a tool to the International Classification of Functioning, Disability and Health[41] to determine the level and type of disability construct being captured (eg, impairment, activity, or participation).

While the intention of this overview was not to re-examine the primary studies within the systematic reviews, it was apparent from many of the systematic reviews included in our synthesis that design and evaluation for many neck-specific patient-reported outcome measures appear to suffer from consistent threats to validity. These include small or unjustified sample sizes for the analyses performed, lack of a priori hypotheses defined for concurrent or longitudinal validity assessment, and poorly defined constructs that impair the ability to appraise content validity. Common threats to external validity appear to be poorly defined patient recruitment and unspecified inclusion/exclusion criteria (ie, often referred to simply as "neck pain"), which do not recognize the heterogeneity of this musculoskeletal condition.[14] We also note that face validity of the patient-reported outcome measures was not considered in our overview, as the criteria within the COSMIN checklist[25] list face validity as an aspect of content validity, the measurement property that is to be evaluated; the current checklist does not require assessment of face validity independent from content validity. Our overview

TABLE 9 Summary of the Psychometric Properties of the NBQ in Included Systematic Reviews

Study	Reliability	Validity	Responsiveness	Other Reports
Pellicciari et al[22]	Cronbach's alpha for the total score ranged from .79 to .92	Moderate to strong correlations with the NDI, NPDS, and CNFDS, and weak to moderate correlations with a pain VAS	The MCID ranged from 4.4 to 5.5 points, but there were higher raw change scores of 13 points or more	The MDC of the questionnaire has never been calculated
Schellingerhout et al[32]	No methodologically sound studies evaluating the internal consistency, measurement error, or reliability	No methodologically sound studies evaluating content validity or structural validity	Limited positive evidence for responsiveness (r = 0.42-0.82)	Limited positive evidence for hypothesis testing (r = 0.63)
Schellingerhout et al[31]	The Dutch NBQ: ICC = 0.92; measurement error of the NBQ is indeterminate	The French NBQ with an instrument measuring pain and physical functioning: r = 0.61-0.67	Limited negative evidence for the responsiveness of the French NBQ (r = 0.42)	The Dutch NBQ had excellent methodological quality of the translation process, but the French NBQ had poor quality
Ferreira et al[6]	Internal consistency, Cronbach α = .86 (excellent); test-retest reliability, ICC = 0.91 (excellent)	Convergent validity correlation of 0.69 (excellent) with a pain VAS	Not reported	The NBQ had 3 items classified in ICF body functions and structures and 3 in activity and participation

Abbreviations: CNFDS, Copenhagen Neck Functional Disability Scale; ICC, intraclass correlation coefficient; ICF, International Classification of Functioning, Disability and Health; MCID, minimum clinically important difference; MDC, minimum detectable change; NBQ, Neck Bournemouth Questionnaire; NDI, Neck Disability Index; NPDS; Neck Pain and Disability scale; VAS, visual analog scale.

showed that the original studies did not consistently report on the specific subtype of participants with neck pain used in the study. As such, we cannot state with any confidence whether the patient-reported outcome measures included in this overview function differently with respect to the type of neck pain (eg, traumatic, insidious, radicular, coordination deficit); the exception is the Whiplash Disability Questionnaire, which is specific to this neck disorder subtype. The original reviews and studies were also limited in specifying general participant characteristics (eg, between sexes, age groups, ethnic or cultural backgrounds), which makes it difficult to allow clinicians to interpret scale scores across clinically relevant subgroups. These are critical gaps in knowledge given the global ubiquity of neck pain.

Internal consistency, most commonly reported using the Cronbach alpha statistic, appears to have been one of the most common properties assessed. This statistic, ranging from .0 to 1.0, is calculated using the pairwise correlations of each item on a scale with each of the other items, and can be interpreted as the degree to which items all represent a similar latent construct. It has been suggested that a desirable range for alpha is between .80 and .95.[18] This is one of the few statistics where a very high value (above .95) may not be desirable, as it indicates likely redundancy in items, suggesting that not all are required to adequately locate the patient along the latent construct of interest.[34] Beyond providing some degree of confidence that the items on the scale all appear to be informed by the latent construct, it is arguably one of the least valuable properties for selecting a patient-reported outcome measure to assist clinical decision making. Despite the almost universal reporting of Cronbach's alpha, we note during our overview that it has rarely been interpreted to a degree beyond reporting the value, and we suggest it is of questionable value in defending the adequacy of a patient-reported outcome measure. Surprisingly, other measurement properties (especially responsiveness) were conspicuously absent in the eligible systematic reviews; this points to the need for more studies to assess these critical measurement properties of these patient-reported outcome measures.

Responsiveness to change is a different construct from reliability; the former explores the meaning of changes in score when clinical change has occurred, while the latter explores stability of scores when no change is expected to have occurred. While the different types of reliability are important properties to estimate the natural "noise" in a score, responsiveness is more valuable to clinicians by providing a threshold value that is most likely to represent a minimum level of clinically meaningful change (the minimal important change, also known as the clinically important change).[5] Different methods to estimate responsiveness have been described, including anchor-based and distribution-based approaches. While it is beyond the scope of this overview to contrast the different approaches, our synthesis did reveal that very few patient-reported outcome measures have either had an evaluation of their responsiveness or have not been synthesized in the eligible systematic review. Clearly, this important knowledge gap will not assist clinicians in making decisions about which patient-reported outcome measure to use in their daily practice and serves as a key barrier to widespread implementation of any of them.

This overview study has some limitations that must be taken into account when interpreting our findings. Overview methodology requires the reporting of measurement properties and the conclusions that were determined within the original systematic review.[30] We observed an overlap among the primary studies included in different systematic reviews evaluating the same patient-reported outcome measure, and therefore overlap of the measurement properties of patient-reported outcome measures. Although the included systematic reviews had different objectives and eligibility criteria (ie, different populations), the overlap resulted in syntheses and conclusions that were generally consistent across reviews. Furthermore, we limited our analysis to patient-reported outcome measures that were evaluated in 2 or more systematic reviews, and, therefore, we did not report the assessment of the measurement properties of other patient-reported outcome measures (n = 15) evaluated in the eligible systematic reviews. We acknowledge that patient-reported outcome measures that are frequently evaluated in systematic reviews (ie, more than 2 systematic reviews) may not have better psychometric properties than other patient-reported outcome measures that were not included in our analysis; however, they do likely reflect patient-reported outcome measures frequently selected for use in clinical practice. Finally, we note that the AMSTAR risk-of-bias tool may not adequately capture all aspects of importance in the critical appraisal of systematic review methods and does not assess the adequacy with which the measurement properties of the patient-reported outcome measures included in the reviews were assessed.

There are several important research gaps that our overview has identified, and this leads to several key areas for future research. These include reporting issues such as (1) clear specification of the population used to establish the measurement, with adequate description of the participant characteristics as well as neck disorder type; (2) specification of a priori hypotheses defined for concurrent or longitudinal validity assessment; and (3) clearly defined constructs to assess content validity. Moreover, future research in establishing patient-reported outcome measure measurement properties could include (1) more head-to-head comparison of different patient-reported outcome measures in the same patient population and clinical context to evaluate the relative performance of these instruments; (2) assessment of responsiveness and both anchor- and distribution-based approaches (from a clinical perspective, this

is critical); and (3) adequate sample-size calculations. Finally, the COSMIN checklist has been modified to be used to assess risk of bias,[25] and its use in future systematic reviews will allow for a better understanding of the methodological quality of the studies.

CONCLUSION

High-quality evidence found the NDI to have good to excellent internal consistency and moderate to excellent test-retest reliability. Moderate-quality evidence found the NPQ to have good to excellent internal consistency and good test-retest reliability. High-quality evidence found the CNFDS to have excellent test-retest reliability and good to strong construct validity with pain scales. Moderate-quality evidence found the NPDS to have unclear to excellent internal consistency and moderate to strong concurrent associations with the NDI and the global assessment of change. Moderate-quality evidence found the Whiplash Disability Questionnaire to have excellent internal consistency, and high test-retest reliability was found for the PSFS. ◉

KEY POINTS

FINDINGS: Only the Neck Disability Index (NDI) and Copenhagen Neck Functional Disability Scale (CNFDS) were supported by high-quality evidence of internal consistency, and only the CNFDS was supported by high-quality evidence of test-retest reliability and construct validity.

IMPLICATIONS: Very few patient-reported outcome measures have either had an evaluation of their responsiveness or have not been synthesized in the eligible systematic reviews. Clearly, this important knowledge gap will not assist clinicians in making decisions about which patient-reported outcome measures to use in their daily practice and serves as a key barrier to widespread implementation of them.

CAUTION: The patient-reported outcome measures that are frequently evaluated in systematic reviews (more than 2) may not have better psychometric properties than other patient-reported outcome measures that were not included in our analysis.

ACKNOWLEDGMENTS: *We would like to thank the International Collaboration On Neck, which is a multidisciplinary collaborative group that includes scientist-authors (listed below) and support staff (Margaret Lomotan) who conduct knowledge synthesis and translation aimed at reducing the burden of neck pain. The International Collaboration On Neck authors are, in alphabetical order, Gert Bronfort, Norm Buckley, Lisa Carlesso, Linda Carroll, Pierre Côté, Jeanette Ezzo, Paulo Ferreira, Tim Flynn, Charlie Goldsmith, Anita Gross, Ted Haines, Jan Hartvigsen, Wayne Hing, Gwendolen Jull, Faith Kaplan, Ron Kaplan, Helge Kasch, Justin Kenardy, Per Kjær, Janet Lowcock, Joy MacDermid, Jordan Miller, Margareta Nordin, Paul Peloso, Jan Pool, Duncan Reid, Sidney Rubinstein, P. Lina Santaguida, Anne Söderlund, Natalie Spearing, Michele Sterling, Grace Szeto, Robert Teasell, Arianne Verhagen, David M. Walton, and Marc White. The expert panel includes Gert Bronfort, DC, PhD, Northwestern Health Sciences University (Minnesota); Duncan Reid, PT, PhD, Auckland University of Technology (New Zealand); Ron Kaplan, PhD; Anne Söderlund, PT, PhD, Mälardalen University College (Sweden); Grace P.Y. Szeto, PT, PhD, Hong Kong Polytechnic University (Hong Kong); and Marc White, MD, Canadian Institute for the Relief of Pain and Disability (Canada).*

REFERENCES

1. Alreni ASE, Harrop D, Lowe A, Potia T, Kilner K, McLean SM. Measures of upper limb function for people with neck pain. A systematic review of measurement and practical properties. *Musculoskelet Sci Pract*. 2017;29:155-163. https://doi.org/10.1016/j.msksp.2017.02.004
2. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)*. 2000;25:3186-3191. https://doi.org/10.1097/00007632-200012150-00014
3. Black N. Patient reported outcome measures could help transform healthcare. *BMJ*. 2013;346:f167. https://doi.org/10.1136/bmj.f167
4. Carroll LJ, Hogg-Johnson S, van der Velde G, et al. Course and prognostic factors for neck pain in the general population: results of the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. *Spine (Phila Pa 1976)*. 2008;33:S75-S82. https://doi.org/10.1097/BRS.0b013e31816445be
5. Engel L, Beaton DE, Touma Z. Minimal clinically important difference: a review of outcome measure score interpretation. *Rheum Dis Clin North Am*. 2018;44:177-188.
6. Ferreira ML, Borges BM, Rezende IL, et al. Are neck pain scales and questionnaires compatible with the International Classification of Functioning, Disability and Health? A systematic review. *Disabil Rehabil*. 2010;32:1539-1546. https://doi.org/10.3109/09638281003611045
7. Forhan M, Vrkljan B, MacDermid J. A systematic review of the quality of psychometric evidence supporting the use of an obesity-specific quality of life measure for use with persons who have class III obesity. *Obes Rev*. 2010;11:222-228. https://doi.org/10.1111/j.1467-789X.2009.00612.x
8. Griggs RK, Cook J, Gargan M, Bannister G, Amirfeyz R. Mid-term follow-up of whiplash with Bournemouth Questionnaire: the significance of the initial depression to pain ratio. *J Back Musculoskelet Rehabil*. 2015;28:443-446. https://doi.org/10.3233/BMR-140538
9. Guzman J, Hurwitz EL, Carroll LJ, et al. A new conceptual model of neck pain: linking onset, course, and care: the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. *Spine (Phila Pa 1976)*. 2008;33:S14-S23. https://doi.org/10.1097/BRS.0b013e3181643efb
10. Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions*. Oxford, UK: The Cochrane Collaboration; 2011.
11. Holly LT, Matz PG, Anderson PA, et al. Functional outcomes assessment for cervical degenerative disease. *J Neurosurg Spine*. 2009;11:238-244. https://doi.org/10.3171/2009.2.SPINE08715
12. Horn KK, Jennings S, Richardson G, van Vliet D, Hefford C, Abbott JH. The Patient-Specific Functional Scale: psychometrics, clinimetrics, and application as a clinical outcome measure. *J Orthop Sports Phys Ther*. 2012;42:30-42. https://doi.org/10.2519/jospt.2012.3727
13. Li Q, Shen H, Li M. Magnetic resonance imaging signal changes of alar and transverse ligaments not correlated with whiplash-associated disorders: a meta-analysis of case-control studies. *Eur Spine J*. 2013;22:14-20. https://doi.org/10.1007/s00586-012-2490-x
14. MacDermid JC, Walton DM, Avery S, et al. Measurement properties of the Neck Disability Index: a systematic review. *J Orthop Sports Phys Ther*. 2009;39:400-417. https://doi.org/10.2519/jospt.2009.2930
15. Matsumoto M, Okada E, Ichihara D, et al. Prospective ten-year follow-up study comparing

patients with whiplash-associated disorders and asymptomatic subjects using magnetic resonance imaging. *Spine (Phila Pa 1976)*. 2010;35:1684-1690. https://doi.org/10.1097/BRS.0b013e3181c9a8c7

16. Menezes Costa LC, Maher CG, McAuley JH, Costa LO. Systematic review of cross-cultural adaptations of McGill Pain Questionnaire reveals a paucity of clinimetric testing. *J Clin Epidemiol*. 2009;62:934-943. https://doi.org/10.1016/j.jclinepi.2009.03.019
17. Misailidou V, Malliou P, Beneka A, Karagiannidis A, Godolias G. Assessment of patients with neck pain: a review of definitions, selection criteria, and measurement tools. *J Chiropr Med*. 2010;9:49-59. https://doi.org/10.1016/j.jcm.2010.03.002
18. Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol*. 2010;10:22. https://doi.org/10.1186/1471-2288-10-22
19. Murphy DR, Lopez M. Neck and back pain specific outcome assessment questionnaires in the Spanish language: a systematic literature review. *Spine J*. 2013;13:1667-1674. https://doi.org/10.1016/j.spinee.2013.08.046
20. Murray CJ, Barber RM, Foreman KJ, et al. Global, regional, and national disability-adjusted life years (DALYs) for 306 diseases and injuries and healthy life expectancy (HALE) for 188 countries, 1990-2013: quantifying the epidemiological transition. *Lancet*. 2015;386:2145-2191. https://doi.org/10.1016/S0140-6736(15)61340-X
21. Packham T, MacDermid JC, Henry J, Bain J. A systematic review of psychometric evaluations of outcome assessments for complex regional pain syndrome. *Disabil Rehabil*. 2012;34:1059-1069. https://doi.org/10.3109/09638288.2011.626835
22. Pellicciari L, Bonetti F, Di Foggia D, Monesi M, Vercelli S. Patient-reported outcome measures for non-specific neck pain validated in the Italian-language: a systematic review. *Arch Physiother*. 2016;6:9. https://doi.org/10.1186/s40945-016-0024-2
23. Picavet HS, Schouten JS. Musculoskeletal pain in the Netherlands: prevalences, consequences and risk groups, the DMC_3-study. *Pain*. 2003;102:167-178. https://doi.org/10.1016/s0304-3959(02)00372-x
24. Pietrobon R, Coeytaux RR, Carey TS, Richardson WJ, DeVellis RF. Standard scales for measurement of functional outcome for cervical pain or dysfunction: a systematic review. *Spine (Phila Pa 1976)*. 2002;27:515-522. https://doi.org/10.1097/00007632-200203010-00012
25. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27:1147-1157. https://doi.org/10.1007/s11136-018-1798-3
26. Robinson JP, Burwinkle T, Turk DC. Perceived and actual memory, concentration, and attention problems after whiplash-associated disorders (grades I and II): prevalence and predictors. *Arch Phys Med Rehabil*. 2007;88:774-779. https://doi.org/10.1016/j.apmr.2007.03.004
27. Rodine RJ, Vernon H. Cervical radiculopathy: a systematic review on treatment by spinal manipulation and measurement with the Neck Disability Index. *J Can Chiropr Assoc*. 2012;56:18-28.
28. Roy JS, MacDermid JC, Tang K, Beaton DE. Construct and predictive validity of the Chronic Pain Grade in workers with chronic work-related upper-extremity disorders. *Clin J Pain*. 2013;29:891-897. https://doi.org/10.1097/AJP.0b013e318278d455
29. Roy JS, MacDermid JC, Woodhouse LJ. A systematic review of the psychometric properties of the Constant-Murley score. *J Shoulder Elbow Surg*. 2010;19:157-164. https://doi.org/10.1016/j.jse.2009.04.008
30. Santaguida PL, Keshavarz H, Carlesso LC, et al. A description of the methodology used in an overview of reviews to evaluate evidence on the treatment, harms, diagnosis/classification, prognosis and outcomes used in the management of neck pain. *Open Orthop J*. 2013;7:461-472. https://doi.org/10.2174/1874325001307010461
31. Schellingerhout JM, Heymans MW, Verhagen AP, de Vet HC, Koes BW, Terwee CB. Measurement properties of translated versions of neck-specific questionnaires: a systematic review. *BMC Med Res Methodol*. 2011;11:87. https://doi.org/10.1186/1471-2288-11-87
32. Schellingerhout JM, Verhagen AP, Heymans MW, Koes BW, de Vet HC, Terwee CB. Measurement properties of disease-specific questionnaires in patients with neck pain: a systematic review. *Qual Life Res*. 2012;21:659-670. https://doi.org/10.1007/s11136-011-9965-9
33. Shea BJ, Bouter LM, Peterson J, et al. External validation of a measurement tool to assess systematic reviews (AMSTAR). *PLoS One*. 2007;2:e1350. https://doi.org/10.1371/journal.pone.0001350
34. Streiner DL. Starting at the beginning: an introduction to coefficient alpha and internal consistency. *J Pers Assess*. 2003;80:99-103. https://doi.org/10.1207/S15327752JPA8001_18
35. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60:34-42. https://doi.org/10.1016/j.jclinepi.2006.03.012
36. Treleaven J. Sensorimotor disturbances in neck disorders affecting postural stability, head and eye movement control—part 2: case studies. *Man Ther*. 2008;13:266-275. https://doi.org/10.1016/j.math.2007.11.002
37. Vernon H. The psychometric properties of the Neck Disability Index [letter]. *Arch Phys Med Rehabil*. 2008;89:1414-1415. https://doi.org/10.1016/j.apmr.2008.05.003
38. Walton D. A review of the definitions of 'recovery' used in prognostic studies on whiplash using an ICF framework. *Disabil Rehabil*. 2009;31:943-957. https://doi.org/10.1080/09638280802404128
39. Walton DM, MacDermid JC, Santaguida PL, Gross A, Carlesso L, International Collaboration On Neck. Results of an international survey of practice patterns for establishing prognosis in neck pain: the ICON project. *Open Orthop J*. 2013;7:387-395. https://doi.org/10.2174/1874325001307010387
40. Whipple JL, Lambert MJ. Outcome measures for practice. *Annu Rev Clin Psychol*. 2011;7:87-111. https://doi.org/10.1146/annurev-clinpsy-040510-143938
41. World Health Organization. *International Classification of Functioning, Disability and Health: ICF*. Geneva, Switzerland: World Health Organization; 2001.
42. Yao M, Sun YL, Cao ZY, et al. A systematic review of cross-cultural adaptation of the Neck Disability Index. *Spine (Phila Pa 1976)*. 2015;40:480-490. https://doi.org/10.1097/BRS.0000000000000788

VIEW Videos on *JOSPT*'s Website

Videos posted with select articles on the *Journal*'s website (**www.jospt.org**) show how conditions are diagnosed and interventions performed. To view the associated videos for an article, click on **Supplementary Material** and scroll down to stream the videos online or download them to your computer or device.

APPENDIX

SEARCH TERMS USED WITHIN MEDLINE FOR ALL AREAS IN THE OVERVIEW OF REVIEWS ACROSS DIFFERENT CLINICAL AREAS FOR THE MANAGEMENT OF NECK PAIN

ICON-Diagnosis

MEDLINE-Ovid (September 13, 2010)

1. Neck Pain/
2. exp Brachial Plexus Neuropathies/
3. exp neck injuries/ or exp whiplash injuries/
4. cervical pain.mp.
5. neckache.mp.
6. whiplash.mp.
7. cervicodynia.mp.
8. cervicalgia.mp.
9. brachialgia.mp.
10. brachial neuritis.mp.
11. brachial neuralgia.mp.
12. neck pain.mp.
13. neck injur*.mp.
14. brachial plexus neuropath*.mp.
15. brachial plexus neuritis.mp.
16. thoracic outlet syndrome/ or cervical rib syndrome/
17. Torticollis/
18. exp brachial plexus neuropathies/ or exp brachial plexus neuritis/
19. cervico brachial neuralgia.ti,ab.
20. cervicobrachial neuralgia.ti,ab.
21. (monoradicul* or monoradicl*).tw.
22. or/1-21
23. exp headache/ and cervic*.tw.
24. exp genital diseases, female/
25. genital disease*.mp.
26. or/24-25
27. 23 not 26
28. 22 or 27
29. neck/
30. neck muscles/
31. exp cervical plexus/
32. exp cervical vertebrae/
33. atlanto-axial joint/
34. atlanto-occipital joint/
35. Cervical Atlas/
36. spinal nerve roots/
37. exp brachial plexus/
38. (odontoid* or cervical or occip* or atlant*).tw.
39. axis/ or odontoid process/
40. Thoracic Vertebrae/
41. cervical vertebrae.mp.
42. cervical plexus.mp.
43. cervical spine.mp.
44. (neck adj3 muscles).mp.
45. (brachial adj3 plexus).mp.
46. (thoracic adj3 vertebrae).mp.

APPENDIX

47. neck.mp.
48. (thoracic adj3 spine).mp.
49. (thoracic adj3 outlet).mp.
50. trapezius.mp.
51. cervical.mp.
52. cervico*.mp.
53. 51 or 52
54. exp genital diseases, female/
55. genital disease*.mp.
56. exp *Uterus/
57. 54 or 55 or 56
58. 53 not 57
59. 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 58
60. exp pain/
61. exp injuries/
62. pain.mp.
63. ache.mp.
64. sore.mp.
65. stiff.mp.
66. discomfort.mp.
67. injur*.mp.
68. neuropath*.mp.
69. or/60-68
70. 59 and 69
71. Radiculopathy/
72. exp temporomandibular joint disorders/ or exp temporomandibular joint dysfunction syndrome/
73. myofascial pain syndromes/
74. exp "Sprains and Strains"/
75. exp Spinal Osteophytosis/
76. exp Neuritis/
77. Polyradiculopathy/
78. exp Arthritis/
79. Fibromyalgia/
80. spondylitis/ or discitis/
81. spondylosis/ or spondylolysis/ or spondylolisthesis/
82. radiculopathy.mp.
83. radiculitis.mp.
84. temporomandibular.mp.
85. myofascial pain syndrome*.mp.
86. thoracic outlet syndrome*.mp.
87. spinal osteophytosis.mp.
88. neuritis.mp.
89. spondylosis.mp.
90. spondylitis.mp.
91. spondylolisthesis.mp.
92. or/71-91
93. 59 and 92
94. exp neck/
95. exp cervical vertebrae/
96. Thoracic Vertebrae/
97. neck.mp.
98. (thoracic adj3 vertebrae).mp.

APPENDIX

99. cervical.mp.
100. cervico*.mp.
101. 99 or 100
102. exp genital diseases, female/
103. genital disease*.mp.
104. exp *Uterus/
105. or/102-104
106. 101 not 105
107. (thoracic adj3 spine).mp.
108. cervical spine.mp.
109. 94 or 95 or 96 or 97 or 98 or 106 or 107 or 108
110. Intervertebral Disk/
111. (disc or discs).mp.
112. (disk or disks).mp.
113. 110 or 111 or 112
114. 109 and 113
115. herniat*.mp.
116. slipped.mp.
117. prolapse*.mp.
118. displace*.mp.
119. degenerat*.mp.
120. (bulge or bulged or bulging).mp.
121. 115 or 116 or 117 or 118 or 119 or 120
122. 114 and 121
123. intervertebral disk degeneration/ or intervertebral disk displacement/
124. intervertebral disk displacement.mp.
125. intervertebral disc displacement.mp.
126. intervertebral disk degeneration.mp.
127. intervertebral disc degeneration.mp.
128. 123 or 124 or 125 or 126 or 127
129. 109 and 128
130. 28 or 70 or 93 or 122 or 129
131. animals/ not (animals/ and humans/)
132. 130 not 131
133. exp *neoplasms/
134. exp *wounds, penetrating/
135. 133 or 134
136. 132 not 135
137. (sensitiv* or diagnos*).mp. or di.fs.
138. 136 and 137
139. meta-analysis.pt,ti,ab,sh.
140. (meta anal$ or metaanal$).ti,ab,sh.
141. ((methodol$ or systematic$ or quantitativ$) adj3 (review$ or overview$ or survey$)).ti.
142. ((methodol$ or systematic$ or quantitativ$) adj3 (review$ or overview$ or survey$)).ab.
143. ((pool$ or combined or combining) adj (data or trials or studies or results)).ti,ab.
144. (medline or embase or cochrane or pubmed or pub med).ti,ab.
145. or/142-144
146. review.pt,sh.
147. 145 and 146
148. or/139-141
149. 147 or 148
150. guidelines as topic/

APPENDIX

151. practice guidelines as topic/
152. guideline.pt.
153. practice guideline.pt.
154. (guideline? or guidance or recommendations).ti.
155. consensus.ti.
156. or/150-155
157. 149 or 156
158. 138 and 157
159. limit 158 to yr="2000 -Current"

ICON-Outcomes

MEDLINE-Ovid

1. exp "outcome and process assessment (health care)"/ or "outcome assessment (health care)"/ or treatment outcome/
2. outcome?.ti.
3. exp "Range of Motion, Articular"/
4. Pain Measurement/
5. exp disability evaluation/
6. "Recovery of Function"/
7. Questionnaires/
8. self-report.tw.
9. ((impairment or disability or function) adj2 (measure? or scale? or evaluation?)).tw.
10. range of motion.tw.
11. (strength adj2 (measure? or scale? or evaluation?)).tw.
12. (outcome? adj2 (measure* or scale? or indicator?)).tw.
13. or/1-12
14. "reproducibility of results"/
15. exp "Sensitivity and Specificity"/
16. reliability.mp.
17. validity.mp.
18. responsiveness.mp.
19. Psychometrics/
20. rasch.mp.
21. factor analysis, statistical/
22. factor analysis.tw.
23. differential functioning.mp.
24. (validity or validation).mp. [mp=title, original title, abstract, name of substance word, subject heading word, unique identifier]
25. (validity or validation).mp.
26. item difficulty.mp.
27. translation.tw.
28. or/14-27
29. 13 and 28
30. Neck Pain/
31. exp Brachial Plexus Neuropathies/
32. exp neck injuries/ or exp whiplash injuries/
33. cervical pain.mp.
34. neckache.mp.
35. whiplash.mp.
36. cervicodynia.mp.
37. cervicalgia.mp.
38. brachialgia.mp.
39. brachial neuritis.mp.
40. brachial neuralgia.mp.

APPENDIX

41. neck pain.mp.
42. neck injur*.mp.
43. brachial plexus neuropath*.mp.
44. brachial plexus neuritis.mp.
45. thoracic outlet syndrome/ or cervical rib syndrome/
46. Torticollis/
47. exp brachial plexus neuropathies/ or exp brachial plexus neuritis/
48. cervico brachial neuralgia.ti,ab.
49. cervicobrachial neuralgia.ti,ab.
50. (monoradicul* or monoradicl*).tw.
51. or/30-50
52. exp headache/ and cervic*.tw.
53. exp genital diseases, female/
54. genital disease*.mp.
55. or/53-54
56. 52 not 55
57. 51 or 56
58. neck/
59. neck muscles/
60. exp cervical plexus/
61. exp cervical vertebrae/
62. atlanto-axial joint/
63. atlanto-occipital joint/
64. Cervical Atlas/
65. spinal nerve roots/
66. exp brachial plexus/
67. (odontoid* or cervical or occip* or atlant*).tw.
68. axis/ or odontoid process/
69. Thoracic Vertebrae/
70. cervical vertebrae.mp.
71. cervical plexus.mp.
72. cervical spine.mp.
73. (neck adj3 muscles).mp.
74. (brachial adj3 plexus).mp.
75. (thoracic adj3 vertebrae).mp.
76. neck.mp.
77. (thoracic adj3 spine).mp.
78. (thoracic adj3 outlet).mp.
79. trapezius.mp.
80. cervical.mp.
81. cervico*.mp.
82. 80 or 81
83. exp genital diseases, female/
84. genital disease*.mp.
85. exp *Uterus/
86. 83 or 84 or 85
87. 82 not 86
88. 58 or 59 or 60 or 61 or 62 or 63 or 64 or 65 or 66 or 67 or 68 or 69 or 70 or 71 or 72 or 73 or 74 or 75 or 76 or 77 or 78 or 79 or 87
89. exp pain/
90. exp injuries/
91. pain.mp.
92. ache.mp.

APPENDIX

93. sore.mp.
94. stiff.mp.
95. discomfort.mp.
96. injur*.mp.
97. neuropath*.mp.
98. or/89-97
99. 88 and 98
100. Radiculopathy/
101. exp temporomandibular joint disorders/ or exp temporomandibular joint dysfunction syndrome/
102. myofascial pain syndromes/
103. exp "Sprains and Strains"/
104. exp Spinal Osteophytosis/
105. exp Neuritis/
106. Polyradiculopathy/
107. exp Arthritis/
108. Fibromyalgia/
109. spondylitis/ or discitis/
110. spondylosis/ or spondylolysis/ or spondylolisthesis/
111. radiculopathy.mp.
112. radiculitis.mp.
113. temporomandibular.mp.
114. myofascial pain syndrome*.mp.
115. thoracic outlet syndrome*.mp.
116. spinal osteophytosis.mp.
117. neuritis.mp.
118. spondylosis.mp.
119. spondylitis.mp.
120. spondylolisthesis.mp.
121. or/100-120
122. 88 and 121
123. exp neck/
124. exp cervical vertebrae/
125. Thoracic Vertebrae/
126. neck.mp.
127. (thoracic adj3 vertebrae).mp.
128. cervical.mp.
129. cervico*.mp.
130. 128 or 129
131. exp genital diseases, female/
132. genital disease*.mp.
133. exp *Uterus/
134. or/131-133
135. 130 not 134
136. (thoracic adj3 spine).mp.
137. cervical spine.mp.
138. 123 or 124 or 125 or 126 or 127 or 135 or 136 or 137
139. Intervertebral Disk/
140. (disc or discs).mp.
141. (disk or disks).mp.
142. 139 or 140 or 141
143. 138 and 142
144. herniat*.mp.

APPENDIX

145. slipped.mp.
146. prolapse*.mp.
147. displace*.mp.
148. degenerat*.mp.
149. (bulge or bulged or bulging).mp.
150. 144 or 145 or 146 or 147 or 148 or 149
151. 143 and 150
152. intervertebral disk degeneration/ or intervertebral disk displacement/
153. intervertebral disk displacement.mp.
154. intervertebral disc displacement.mp.
155. intervertebral disk degeneration.mp.
156. intervertebral disc degeneration.mp.
157. 152 or 153 or 154 or 155 or 156
158. 138 and 157
159. 57 or 99 or 122 or 151 or 158
160. animals/ not (animals/ and humans/)
161. 159 not 160
162. exp *neoplasms/
163. exp *wounds, penetrating/
164. 162 or 163
165. 161 not 164
166. 29 and 165
167. guidelines as topic/
168. practice guidelines as topic/
169. guideline.pt.
170. practice guideline.pt.
171. (guideline? or guidance or recommendations).ti.
172. consensus.ti.
173. or/167-172
174. meta-analysis/
175. exp meta-analysis as topic/
176. (meta analy* or metaanaly* or met analy* or metanaly*).tw.
177. review literature as topic/
178. (collaborative research or collaborative review* or collaborative overview*).tw.
179. (integrative research or integrative review* or intergrative overview*).tw.
180. (quantitative adj3 (research or review* or overview*)).tw.
181. (research integration or research overview*).tw.
182. (systematic* adj3 (review* or overview*)).tw.
183. (methodologic* adj3 (review* or overview*)).tw.
184. exp technology assessment biomedical/
185. (hta or thas or technology assessment*).tw.
186. ((hand adj2 search*) or (manual* adj search*)).tw.
187. ((electronic adj database*) or (bibliographic* adj database*)).tw.
188. ((data adj2 abstract*) or (data adj2 extract*)).tw.
189. (analys* adj3 (pool or pooled or pooling)).tw.
190. mantel haenszel.tw.
191. (cohrane or pubmed or pub med or medline or embase or psycinfo or psyclit or psychinfo or psychlit or cinahl or science citation indes).ab.
192. or/174-191
193. 173 or 192
194. 166 and 193
195. limit 194 to yr="2000 -Current"

APPENDIX

ICON-Prognosis

MEDLINE-Ovid

1. Neck Pain/
2. exp Brachial Plexus Neuropathies/
3. exp neck injuries/ or exp whiplash injuries/
4. cervical pain.mp.
5. neckache.mp.
6. whiplash.mp.
7. cervicodynia.mp.
8. cervicalgia.mp.
9. brachialgia.mp.
10. brachial neuritis.mp.
11. brachial neuralgia.mp.
12. neck pain.mp.
13. neck injur*.mp.
14. brachial plexus neuropath*.mp.
15. brachial plexus neuritis.mp.
16. thoracic outlet syndrome/ or cervical rib syndrome/
17. Torticollis/
18. exp brachial plexus neuropathies/ or exp brachial plexus neuritis/
19. cervico brachial neuralgia.ti,ab.
20. cervicobrachial neuralgia.ti,ab.
21. (monoradicul* or monoradicl*).tw.
22. or/1-21
23. exp headache/ and cervic*.tw.
24. exp genital diseases, female/
25. genital disease*.mp.
26. or/24-25
27. 23 not 26
28. 22 or 27
29. neck/
30. neck muscles/
31. exp cervical plexus/
32. exp cervical vertebrae/
33. atlanto-axial joint/
34. atlanto-occipital joint/
35. Cervical Atlas/
36. spinal nerve roots/
37. exp brachial plexus/
38. (odontoid* or cervical or occip* or atlant*).tw.
39. axis/ or odontoid process/
40. Thoracic Vertebrae/
41. cervical vertebrae.mp.
42. cervical plexus.mp.
43. cervical spine.mp.
44. (neck adj3 muscles).mp.
45. (brachial adj3 plexus).mp.
46. (thoracic adj3 vertebrae).mp.
47. neck.mp.
48. (thoracic adj3 spine).mp.
49. (thoracic adj3 outlet).mp.
50. trapezius.mp.

APPENDIX

51. cervical.mp.
52. cervico*.mp.
53. 51 or 52
54. exp genital diseases, female/
55. genital disease*.mp.
56. exp *Uterus/
57. 54 or 55 or 56
58. 53 not 57
59. 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 58
60. exp pain/
61. exp injuries/
62. pain.mp.
63. ache.mp.
64. sore.mp.
65. stiff.mp.
66. discomfort.mp.
67. injur*.mp.
68. neuropath*.mp.
69. or/60-68
70. 59 and 69
71. Radiculopathy/
72. exp temporomandibular joint disorders/ or exp temporomandibular joint dysfunction syndrome/
73. myofascial pain syndromes/
74. exp "Sprains and Strains"/
75. exp Spinal Osteophytosis/
76. exp Neuritis/
77. Polyradiculopathy/
78. exp Arthritis/
79. Fibromyalgia/
80. spondylitis/ or discitis/
81. spondylosis/ or spondylolysis/ or spondylolisthesis/
82. radiculopathy.mp.
83. radiculitis.mp.
84. temporomandibular.mp.
85. myofascial pain syndrome*.mp.
86. thoracic outlet syndrome*.mp.
87. spinal osteophytosis.mp.
88. neuritis.mp.
89. spondylosis.mp.
90. spondylitis.mp.
91. spondylolisthesis.mp.
92. or/71-91
93. 59 and 92
94. exp neck/
95. exp cervical vertebrae/
96. Thoracic Vertebrae/
97. neck.mp.
98. (thoracic adj3 vertebrae).mp.
99. cervical.mp.
100. cervico*.mp.
101. 99 or 100
102. exp genital diseases, female/

APPENDIX

103. genital disease*.mp.
104. exp *Uterus/
105. or/102-104
106. 101 not 105
107. (thoracic adj3 spine).mp.
108. cervical spine.mp.
109. 94 or 95 or 96 or 97 or 98 or 106 or 107 or 108
110. Intervertebral Disk/
111. (disc or discs).mp.
112. (disk or disks).mp.
113. 110 or 111 or 112
114. 109 and 113
115. herniat*.mp.
116. slipped.mp.
117. prolapse*.mp.
118. displace*.mp.
119. degenerat*.mp.
120. (bulge or bulged or bulging).mp.
121. 115 or 116 or 117 or 118 or 119 or 120
122. 114 and 121
123. intervertebral disk degeneration/ or intervertebral disk displacement/
124. intervertebral disk displacement.mp.
125. intervertebral disc displacement.mp.
126. intervertebral disk degeneration.mp.
127. intervertebral disc degeneration.mp.
128. 123 or 124 or 125 or 126 or 127
129. 109 and 128
130. 28 or 70 or 93 or 122 or 129
131. animals/ not (animals/ and humans/)
132. 130 not 131
133. exp *neoplasms/
134. exp *wounds, penetrating/
135. 133 or 134
136. 132 not 135
137. incidence/ or exp mortality/
138. follow-up studies/
139. prognos*.tw.
140. predict*.tw.
141. course*.tw.
142. or/137-141
143. 136 and 142
144. animals/ not (animals/ and humans/)
145. 143 not 144
146. meta-analysis/
147. exp meta-analysis as topic/
148. (meta analy* or metaanaly* or met analy* or metanaly*).tw.
149. review literature as topic/
150. (collaborative research or collaborative review* or collaborative overview*).tw.
151. (integrative research or integrative review* or intergrative overview*).tw.
152. (quantitative adj3 (research or review* or overview*)).tw.
153. (research integration or research overview*).tw.
154. (systematic* adj3 (review* or overview*)).tw.

APPENDIX

155. (methodologic* adj3 (review* or overview*)).tw.
156. exp technology assessment biomedical/
157. (hta or thas or technology assessment*).tw.
158. ((hand adj2 search*) or (manual* adj search*)).tw.
159. ((electronic adj database*) or (bibliographic* adj database*)).tw.
160. ((data adj2 abstract*) or (data adj2 extract*)).tw.
161. (analys* adj3 (pool or pooled or pooling)).tw.
162. mantel haenszel.tw.
163. (cohrane or pubmed or pub med or medline or embase or psycinfo or psyclit or psychinfo or psychlit or cinahl or science citation indes).ab.
164. or/146-163
165. 145 and 164
166. guidelines as topic/
167. practice guidelines as topic/
168. guideline.pt.
169. practice guideline.pt.
170. (guideline? or guidance or recommendations).ti.
171. consensus.ti.
172. or/166-171
173. 145 and 172
174. 165 or 173
175. limit 174 to yr="2000 -Current"

ICON TREATMENT DETAILED SEARCH STRATEGIES FOR MEDLINE

ICON-Physical Medicine/Treatment

MEDLINE-Ovid

1. Neck Pain/
2. exp Brachial Plexus Neuropathies/
3. exp neck injuries/ or exp whiplash injuries/
4. cervical pain.mp.
5. neckache.mp.
6. whiplash.mp.
7. cervicodynia.mp.
8. cervicalgia.mp.
9. brachialgia.mp.
10. brachial neuritis.mp.
11. brachial neuralgia.mp.
12. neck pain.mp.
13. neck injur*.mp.
14. brachial plexus neuropath*.mp.
15. brachial plexus neuritis.mp.
16. thoracic outlet syndrome/ or cervical rib syndrome/
17. Torticollis/
18. exp brachial plexus neuropathies/ or exp brachial plexus neuritis/
19. cervico brachial neuralgia.ti,ab.
20. cervicobrachial neuralgia.ti,ab.
21. (monoradicul* or monoradicl*).tw.
22. or/1-21
23. exp headache/ and cervic*.tw.
24. exp genital diseases, female/
25. genital disease*.mp.

APPENDIX

26. or/24-25
27. 23 not 26
28. 22 or 27
29. neck/
30. neck muscles/
31. exp cervical plexus/
32. exp cervical vertebrae/
33. atlanto-axial joint/
34. atlanto-occipital joint/
35. Cervical Atlas/
36. spinal nerve roots/
37. exp brachial plexus/
38. (odontoid* or cervical or occip* or atlant*).tw.
39. axis/ or odontoid process/
40. Thoracic Vertebrae/
41. cervical vertebrae.mp.
42. cervical plexus.mp.
43. cervical spine.mp.
44. (neck adj3 muscles).mp.
45. (brachial adj3 plexus).mp.
46. (thoracic adj3 vertebrae).mp.
47. neck.mp.
48. (thoracic adj3 spine).mp.
49. (thoracic adj3 outlet).mp.
50. trapezius.mp.
51. cervical.mp.
52. cervico*.mp.
53. 51 or 52
54. exp genital diseases, female/
55. genital disease*.mp.
56. exp *Uterus/
57. 54 or 55 or 56
58. 53 not 57
59. 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 58
60. exp pain/
61. exp injuries/
62. pain.mp.
63. ache.mp.
64. sore.mp.
65. stiff.mp.
66. discomfort.mp.
67. injur*.mp.
68. neuropath*.mp.
69. or/60-68
70. 59 and 69
71. Radiculopathy/
72. exp temporomandibular joint disorders/ or exp temporomandibular joint dysfunction syndrome/
73. myofascial pain syndromes/
74. exp “Sprains and Strains”/
75. exp Spinal Osteophytosis/
76. exp Neuritis/
77. Polyradiculopathy/

APPENDIX

78. exp Arthritis/
79. Fibromyalgia/
80. spondylitis/ or discitis/
81. spondylosis/ or spondylolysis/ or spondylolisthesis/
82. radiculopathy.mp.
83. radiculitis.mp.
84. temporomandibular.mp.
85. myofascial pain syndrome*.mp.
86. thoracic outlet syndrome*.mp.
87. spinal osteophytosis.mp.
88. neuritis.mp.
89. spondylosis.mp.
90. spondylitis.mp.
91. spondylolisthesis.mp.
92. or/71-91
93. 59 and 92
94. exp neck/
95. exp cervical vertebrae/
96. Thoracic Vertebrae/
97. neck.mp.
98. (thoracic adj3 vertebrae).mp.
99. cervical.mp.
100. cervico*.mp.
101. 99 or 100
102. exp genital diseases, female/
103. genital disease*.mp.
104. exp *Uterus/
105. or/102-104
106. 101 not 105
107. (thoracic adj3 spine).mp.
108. cervical spine.mp.
109. 94 or 95 or 96 or 97 or 98 or 106 or 107 or 108
110. Intervertebral Disk/
111. (disc or discs).mp.
112. (disk or disks).mp.
113. 110 or 111 or 112
114. 109 and 113
115. herniat*.mp.
116. slipped.mp.
117. prolapse*.mp.
118. displace*.mp.
119. degenerat*.mp.
120. (bulge or bulged or bulging).mp.
121. 115 or 116 or 117 or 118 or 119 or 120
122. 114 and 121
123. intervertebral disk degeneration/ or intervertebral disk displacement/
124. intervertebral disk displacement.mp.
125. intervertebral disc displacement.mp.
126. intervertebral disk degeneration.mp.
127. intervertebral disc degeneration.mp.
128. 123 or 124 or 125 or 126 or 127
129. 109 and 128

APPENDIX

130. 28 or 70 or 93 or 122 or 129
131. animals/ not (animals/ and humans/)
132. 130 not 131
133. exp *neoplasms/
134. exp *wounds, penetrating/
135. 133 or 134
136. 132 not 135
137. Neck Pain/rh [Rehabilitation]
138. exp Brachial Plexus Neuropathies/rh
139. exp neck injuries/rh or exp whiplash injuries/rh
140. thoracic outlet syndrome/rh or cervical rib syndrome/rh
141. Torticollis/rh
142. exp brachial plexus neuropathies/rh or exp brachial plexus neuritis/rh
143. 137 or 138 or 139 or 140 or 141 or 142
144. Radiculopathy/rh
145. exp temporomandibular joint disorders/rh or exp temporomandibular joint dysfunction syndrome/rh
146. myofascial pain syndromes/rh
147. exp "Sprains and Strains"/rh
148. exp Spinal Osteophytosis/rh
149. exp Neuritis/rh
150. Polyradiculopathy/rh
151. exp Arthritis/rh
152. Fibromyalgia/rh
153. spondylitis/rh or discitis/rh
154. spondylosis/rh or spondylolysis/rh or spondylolisthesis/rh
155. or/144-154
156. 59 and 155
157. exp Combined Modality Therapy/
158. Exercise/
159. Physical Exertion/
160. exp Exercise Therapy/
161. exp Electric Stimulation Therapy/
162. Transcutaneous Electric Nerve Stimulation/
163. pulsed electro magnetic field.mp.
164. pulsed electromagnetic field.tw.
165. Electromagnetic Fields/
166. Magnetic Field Therapy/
167. Electric Stimulation/
168. exp Orthotic Devices/
169. kinesiotaping.tw.
170. taping.tw.
171. oral splints.tw.
172. Occlusal Splints/
173. pillow?.tw.
174. collar?.tw.
175. Traction/
176. traction.tw.
177. exp Laser Therapy/
178. laser therapy.tw.
179. exp Rehabilitation/
180. Ultrasonic Therapy/
181. exp Phototherapy/

APPENDIX

182. Lasers/
183. exp Physical Therapy Modalities/
184. repetitive magnetic stimulation.tw.
185. exp Cryotherapy/
186. Hydrotherapy/
187. exp Hyperthermia, Induced/
188. vapocoolant spray.mp.
189. Cryoanesthesia/
190. Ice/
191. postur* correction.mp.
192. Feldenkrais.mp.
193. (alexander adj (technique or method)).tw.
194. Relaxation Therapy/
195. Biofeedback, Psychology/
196. faradic stimulation.mp.
197. or/157-196
198. 136 and 197
199. 143 or 156 or 198
200. animals/ not (animals/ and humans/)
201. 199 not 200
202. guidelines as topic/
203. practice guidelines as topic/
204. guideline.pt.
205. practice guideline.pt.
206. (guideline? or guidance or recommendations).ti.
207. consensus.ti.
208. or/202-207
209. 201 and 208
210. 136 and 208
211. 209 or 210
212. limit 211 to yr="2006 -Current"
213. limit 211 to yr="1902 - 2005"
214. meta-analysis/
215. exp meta-analysis as topic/
216. (meta analy* or metaanaly* or met analy* or metanaly*).tw.
217. review literature as topic/
218. (collaborative research or collaborative review* or collaborative overview*).tw.
219. (integrative research or integrative review* or intergrative overview*).tw.
220. (quantitative adj3 (research or review* or overview*)).tw.
221. (research integration or research overview*).tw.
222. (systematic* adj3 (review* or overview*)).tw.
223. (methodologic* adj3 (review* or overview*)).tw.
224. exp technology assessment biomedical/
225. (hta or thas or technology assessment*).tw.
226. ((hand adj2 search*) or (manual* adj search*)).tw.
227. ((electronic adj database*) or (bibliographic* adj database*)).tw.
228. ((data adj2 abstract*) or (data adj2 extract*)).tw.
229. (analys* adj3 (pool or pooled or pooling)).tw.
230. mantel haenszel.tw.
231. (cohrane or pubmed or pub med or medline or embase or psycinfo or psyclit or psychinfo or psychlit or cinahl or science citation indes).ab.
232. or/214-231
233. 201 and 232

APPENDIX

234. limit 233 to yr="2006 -Current"
235. limit 233 to yr="1902 - 2005"

OR

ICON-Manual Therapy/Treatment

MEDLINE-Ovid

1. Neck Pain/
2. exp Brachial Plexus Neuropathies/
3. exp neck injuries/ or exp whiplash injuries/
4. cervical pain.mp.
5. neckache.mp.
6. whiplash.mp.
7. cervicodynia.mp.
8. cervicalgia.mp.
9. brachialgia.mp.
10. brachial neuritis.mp.
11. brachial neuralgia.mp.
12. neck pain.mp.
13. neck injur*.mp.
14. brachial plexus neuropath*.mp.
15. brachial plexus neuritis.mp.
16. thoracic outlet syndrome/ or cervical rib syndrome/
17. Torticollis/
18. exp brachial plexus neuropathies/ or exp brachial plexus neuritis/
19. cervico brachial neuralgia.ti,ab.
20. cervicobrachial neuralgia.ti,ab.
21. (monoradicul* or monoradicl*).tw.
22. or/1-21
23. exp headache/ and cervic*.tw.
24. exp genital diseases, female/
25. genital disease*.mp.
26. or/24-25
27. 23 not 26
28. 22 or 27
29. neck/
30. neck muscles/
31. exp cervical plexus/
32. exp cervical vertebrae/
33. atlanto-axial joint/
34. atlanto-occipital joint/
35. Cervical Atlas/
36. spinal nerve roots/
37. exp brachial plexus/
38. (odontoid* or cervical or occip* or atlant*).tw.
39. axis/ or odontoid process/
40. Thoracic Vertebrae/
41. cervical vertebrae.mp.
42. cervical plexus.mp.
43. cervical spine.mp.
44. (neck adj3 muscles).mp.
45. (brachial adj3 plexus).mp.
46. (thoracic adj3 vertebrae).mp.

APPENDIX

47. neck.mp.
48. (thoracic adj3 spine).mp.
49. (thoracic adj3 outlet).mp.
50. trapezius.mp.
51. cervical.mp.
52. cervico*.mp.
53. 51 or 52
54. exp genital diseases, female/
55. genital disease*.mp.
56. exp *Uterus/
57. 54 or 55 or 56
58. 53 not 57
59. 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 58
60. exp pain/
61. exp injuries/
62. pain.mp.
63. ache.mp.
64. sore.mp.
65. stiff.mp.
66. discomfort.mp.
67. injur*.mp.
68. neuropath*.mp.
69. or/60-68
70. 59 and 69
71. Radiculopathy/
72. exp temporomandibular joint disorders/ or exp temporomandibular joint dysfunction syndrome/
73. myofascial pain syndromes/
74. exp "Sprains and Strains"/
75. exp Spinal Osteophytosis/
76. exp Neuritis/
77. Polyradiculopathy/
78. exp Arthritis/
79. Fibromyalgia/
80. spondylitis/ or discitis/
81. spondylosis/ or spondylolysis/ or spondylolisthesis/
82. radiculopathy.mp.
83. radiculitis.mp.
84. temporomandibular.mp.
85. myofascial pain syndrome*.mp.
86. thoracic outlet syndrome*.mp.
87. spinal osteophytosis.mp.
88. neuritis.mp.
89. spondylosis.mp.
90. spondylitis.mp.
91. spondylolisthesis.mp.
92. or/71-91
93. 59 and 92
94. exp neck/
95. exp cervical vertebrae/
96. Thoracic Vertebrae/
97. neck.mp.
98. (thoracic adj3 vertebrae).mp.

APPENDIX

99. cervical.mp.
100. cervico*.mp.
101. 99 or 100
102. exp genital diseases, female/
103. genital disease*.mp.
104. exp *Uterus/
105. or/102-104
106. 101 not 105
107. (thoracic adj3 spine).mp.
108. cervical spine.mp.
109. 94 or 95 or 96 or 97 or 98 or 106 or 107 or 108
110. Intervertebral Disk/
111. (disc or discs).mp.
112. (disk or disks).mp.
113. 110 or 111 or 112
114. 109 and 113
115. herniat*.mp.
116. slipped.mp.
117. prolapse*.mp.
118. displace*.mp.
119. degenerat*.mp.
120. (bulge or bulged or bulging).mp.
121. 115 or 116 or 117 or 118 or 119 or 120
122. 114 and 121
123. intervertebral disk degeneration/ or intervertebral disk displacement/
124. intervertebral disk displacement.mp.
125. intervertebral disc displacement.mp.
126. intervertebral disk degeneration.mp.
127. intervertebral disc degeneration.mp.
128. 123 or 124 or 125 or 126 or 127
129. 109 and 128
130. 28 or 70 or 93 or 122 or 129
131. animals/ not (animals/ and humans/)
132. 130 not 131
133. exp *neoplasms/
134. exp *wounds, penetrating/
135. 133 or 134
136. 132 not 135
137. Neck Pain/rh, th [Rehabilitation, Therapy]
138. exp Brachial Plexus Neuropathies/rh, th
139. exp neck injuries/rh, th or exp whiplash injuries/rh, th
140. thoracic outlet syndrome/rh, th or cervical rib syndrome/rh, th
141. Torticollis/rh, th
142. exp brachial plexus neuropathies/rh, th or exp brachial plexus neuritis/rh, th
143. or/137-142
144. Radiculopathy/rh, th
145. exp temporomandibular joint disorders/rh, th or exp temporomandibular joint dysfunction syndrome/rh, th
146. myofascial pain syndromes/rh, th
147. exp "Sprains and Strains"/rh, th
148. exp Spinal Osteophytosis/rh, th
149. exp Neuritis/rh, th
150. Polyradiculopathy/rh, th

APPENDIX

151. exp Arthritis/rh, th
152. Fibromyalgia/rh, th
153. spondylitis/rh, th or discitis/rh, th
154. spondylosis/rh, th or spondylolysis/rh, th or spondylolisthesis/rh, th
155. or/144-154
156. 59 and 155
157. acupuncture/ or chiropractic/
158. exp Musculoskeletal Manipulations/
159. massage.tw.
160. mobili?ation.tw.
161. Acupuncture Therapy/
162. (acupuncture or acu-puncture or needling or acupressure or mox?bustion).tw.
163. ((neck or spine or spinal or cervical or chiropractic* or musculoskeletal* or musculo-skeletal*) adj3 (adjust* or manipulat* or mobiliz* or mobilis*)).tw.
164. (manual adj therap*).tw.
165. (manipulati* adj (therap* or medicine)).tw.
166. (massag* or reflexolog* or rolfing or zone therap*).tw.
167. Nimmo.mp.
168. exp Vibration/tu [Therapeutic Use]
169. (vibration adj5 (therap* or treatment*)).tw.
170. (Chih Ya or Shiatsu or Shiatzu or Zhi Ya).tw.
171. (flexion adj2 distraction*).tw.
172. (myofascial adj3 (release or therap*)).tw.
173. muscle energy technique*.tw.
174. trigger point.tw.
175. proprioceptive Neuromuscular Facilitation*.tw.
176. cyriax friction.tw.
177. (lomilomi or lomi-lomi or trager).tw.
178. aston patterning.tw.
179. (strain adj counterstrain).tw.
180. (craniosacral therap* or cranio-sacral therap*).tw.
181. (amma or ammo or effleuurage or petrissage or hacking or tapotment).tw.
182. Complementary Therapies/
183. ((complement* or alternat* or osteopthic*) adj (therap* or medicine)).tw.
184. (Tui Na or Tuina).tw.
185. or/157-184
186. 136 and 185
187. 143 or 156 or 186
188. animals/ not (animals/ and humans/)
189. 187 not 188
190. guidelines as topic/
191. practice guidelines as topic/
192. guideline.pt.
193. practice guideline.pt.
194. (guideline? or guidance or recommendations).ti.
195. consensus.ti.
196. or/190-195
197. 189 and 196
198. limit 197 to yr="2006 -Current"
199. limit 197 to yr="1902 -2005"
200. meta-analysis/
201. exp meta-analysis as topic/

APPENDIX

202. (meta analy* or metaanaly* or met analy* or metanaly*).tw.
203. review literature as topic/
204. (collaborative research or collaborative review* or collaborative overview*).tw.
205. (integrative research or integrative review* or intergrative overview*).tw.
206. (quantitative adj3 (research or review* or overview*)).tw.
207. (research integration or research overview*).tw.
208. (systematic* adj3 (review* or overview*)).tw.
209. (methodologic* adj3 (review* or overview*)).tw.
210. exp technology assessment biomedical/
211. (hta or thas or technology assessment*).tw.
212. ((hand adj2 search*) or (manual* adj search*)).tw.
213. ((electronic adj database*) or (bibliographic* adj database*)).tw.
214. ((data adj2 abstract*) or (data adj2 extract*)).tw.
215. (analys* adj3 (pool or pooled or pooling)).tw.
216. mantel haenszel.tw.
217. (cohrane or pubmed or pub med or medline or embase or psycinfo or psyclit or psychinfo or psychlit or cinahl or science citation indes).ab.
218. or/200-217
219. 189 and 218
220. limit 219 to yr="2006 -Current"
221. limit 219 to yr="1902 -2005"

OR

ICON-Drug Therapy/Treatment

MEDLINE-Ovid

1. Neck Pain/
2. exp Brachial Plexus Neuropathies/
3. exp neck injuries/ or exp whiplash injuries/
4. cervical pain.mp.
5. neckache.mp.
6. whiplash.mp.
7. cervicodynia.mp.
8. cervicalgia.mp.
9. brachialgia.mp.
10. brachial neuritis.mp.
11. brachial neuralgia.mp.
12. neck pain.mp.
13. neck injur*.mp.
14. brachial plexus neuropath*.mp.
15. brachial plexus neuritis.mp.
16. thoracic outlet syndrome/ or cervical rib syndrome/
17. Torticollis/
18. exp brachial plexus neuropathies/ or exp brachial plexus neuritis/
19. cervico brachial neuralgia.ti,ab.
20. cervicobrachial neuralgia.ti,ab.
21. (monoradicul* or monoradicl*).tw.
22. or/1-21
23. exp headache/ and cervic*.tw.
24. exp genital diseases, female/
25. genital disease*.mp.
26. or/24-25
27. 23 not 26
28. 22 or 27

APPENDIX

29. neck/
30. neck muscles/
31. exp cervical plexus/
32. exp cervical vertebrae/
33. atlanto-axial joint/
34. atlanto-occipital joint/
35. Cervical Atlas/
36. spinal nerve roots/
37. exp brachial plexus/
38. (odontoid* or cervical or occip* or atlant*).tw.
39. axis/ or odontoid process/
40. Thoracic Vertebrae/
41. cervical vertebrae.mp.
42. cervical plexus.mp.
43. cervical spine.mp.
44. (neck adj3 muscles).mp.
45. (brachial adj3 plexus).mp.
46. (thoracic adj3 vertebrae).mp.
47. neck.mp.
48. (thoracic adj3 spine).mp.
49. (thoracic adj3 outlet).mp.
50. trapezius.mp.
51. cervical.mp.
52. cervico*.mp.
53. 51 or 52
54. exp genital diseases, female/
55. genital disease*.mp.
56. exp *Uterus/
57. 54 or 55 or 56
58. 53 not 57
59. 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 58
60. exp pain/
61. exp injuries/
62. pain.mp.
63. ache.mp.
64. sore.mp.
65. stiff.mp.
66. discomfort.mp.
67. injur*.mp.
68. neuropath*.mp.
69. or/60-68
70. 59 and 69
71. Radiculopathy/
72. exp temporomandibular joint disorders/ or exp temporomandibular joint dysfunction syndrome/
73. myofascial pain syndromes/
74. exp “Sprains and Strains”/
75. exp Spinal Osteophytosis/
76. exp Neuritis/
77. Polyradiculopathy/
78. exp Arthritis/
79. Fibromyalgia/
80. spondylitis/ or discitis/

APPENDIX

81. spondylosis/ or spondylolysis/ or spondylolisthesis/
82. radiculopathy.mp.
83. radiculitis.mp.
84. temporomandibular.mp.
85. myofascial pain syndrome*.mp.
86. thoracic outlet syndrome*.mp.
87. spinal osteophytosis.mp.
88. neuritis.mp.
89. spondylosis.mp.
90. spondylitis.mp.
91. spondylolisthesis.mp.
92. or/71-91
93. 59 and 92
94. exp neck/
95. exp cervical vertebrae/
96. Thoracic Vertebrae/
97. neck.mp.
98. (thoracic adj3 vertebrae).mp.
99. cervical.mp.
100. cervico*.mp.
101. 99 or 100
102. exp genital diseases, female/
103. genital disease*.mp.
104. exp *Uterus/
105. or/102-104
106. 101 not 105
107. (thoracic adj3 spine).mp.
108. cervical spine.mp.
109. 94 or 95 or 96 or 97 or 98 or 106 or 107 or 108
110. Intervertebral Disk/
111. (disc or discs).mp.
112. (disk or disks).mp.
113. 110 or 111 or 112
114. 109 and 113
115. herniat*.mp.
116. slipped.mp.
117. prolapse*.mp.
118. displace*.mp.
119. degenerat*.mp.
120. (bulge or bulged or bulging).mp.
121. 115 or 116 or 117 or 118 or 119 or 120
122. 114 and 121
123. intervertebral disk degeneration/ or intervertebral disk displacement/
124. intervertebral disk displacement.mp.
125. intervertebral disc displacement.mp.
126. intervertebral disk degeneration.mp.
127. intervertebral disc degeneration.mp.
128. 123 or 124 or 125 or 126 or 127
129. 109 and 128
130. 28 or 70 or 93 or 122 or 129
131. animals/ not (animals/ and humans/)
132. 130 not 131

APPENDIX

133. exp *neoplasms/
134. exp *wounds, penetrating/
135. 133 or 134
136. 132 not 135
137. Neck Pain/dt
138. exp Brachial Plexus Neuropathies/dt
139. exp neck injuries/dt or exp whiplash injuries/dt
140. thoracic outlet syndrome/dt or cervical rib syndrome/dt
141. Torticollis/dt
142. exp brachial plexus neuritis/dt
143. or/137-142
144. Radiculopathy/dt
145. exp temporomandibular joint disorders/dt or exp temporomandibular joint dysfunction syndrome/dt
146. myofascial pain syndromes/dt
147. exp "Sprains and Strains"/dt
148. exp Spinal Osteophytosis/dt
149. exp Neuritis/dt
150. Polyradiculopathy/dt
151. exp Arthritis/dt
152. fibromyalgia/dt
153. spondylitis/dt or discitis/dt
154. spondylosis/dt or spondylolysis/dt or spondylolisthesis/dt
155. or/144-154
156. 59 and 155
157. exp Drug Therapy/
158. exp analgesics/
159. exp anti-inflammatory agents/
160. exp muscle relaxants, central/
161. exp psychotropic drugs/
162. exp neuromuscular agents/
163. exp antidepressive agents/
164. exp tranquilizing agents/
165. exp Botulinum Toxins/
166. botulin*.tw.
167. botox.tw.
168. Prilocaine/
169. exp Nerve Block/
170. Injections, Intra-Articular/
171. injections, intramuscular/ or injections, epidural/ or injections, subcutaneous/ or injections, intradermal/
172. Lidocaine/
173. Morphine/
174. Methylprednisolone/
175. exp Glucocorticoids/
176. or/157-175
177. 136 and 176
178. 143 or 156 or 177
179. guidelines as topic/
180. practice guidelines as topic/
181. guideline.pt.
182. practice guideline.pt.
183. (guideline? or guidance or recommendations).ti.
184. consensus.ti.

APPENDIX

185. or/179-184
186. 178 and 185
187. limit 186 to yr="2006 -Current"
188. limit 186 to yr="1902 -2005"
189. meta-analysis/
190. exp meta-analysis as topic/
191. (meta analy* or metaanaly* or met analy* or metanaly*).tw.
192. review literature as topic/
193. (collaborative research or collaborative review* or collaborative overview*).tw.
194. (integrative research or integrative review* or intergrative overview*).tw.
195. (quantitative adj3 (research or review* or overview*)).tw.
196. (research integration or research overview*).tw.
197. (systematic* adj3 (review* or overview*)).tw.
198. (methodologic* adj3 (review* or overview*)).tw.
199. exp technology assessment biomedical/
200. (hta or thas or technology assessment*).tw.
201. ((hand adj2 search*) or (manual* adj search*)).tw.
202. ((electronic adj database*) or (bibliographic* adj database*)).tw.
203. ((data adj2 abstract*) or (data adj2 extract*)).tw.
204. (analys* adj3 (pool or pooled or pooling)).tw.
205. mantel haenszel.tw.
206. (cohrane or pubmed or pub med or medline or embase or psycinfo or psyclit or psychinfo or psychlit or cinahl or science citation indes).ab.
207. or/189-206
208. 178 and 207
209. limit 208 to yr="2006 -Current"
210. limit 208 to yr="1902 -2005"

ICON-Treatment/Patient Education

MEDLINE-Ovid

1. Neck Pain/
2. exp Brachial Plexus Neuropathies/
3. exp neck injuries/ or exp whiplash injuries/
4. cervical pain.mp.
5. neckache.mp.
6. whiplash.mp.
7. cervicodynia.mp.
8. cervicalgia.mp.
9. brachialgia.mp.
10. brachial neuritis.mp.
11. brachial neuralgia.mp.
12. neck pain.mp.
13. neck injur*.mp.
14. brachial plexus neuropath*.mp.
15. brachial plexus neuritis.mp.
16. thoracic outlet syndrome/ or cervical rib syndrome/
17. Torticollis/
18. exp brachial plexus neuropathies/ or exp brachial plexus neuritis/
19. cervico brachial neuralgia.ti,ab.
20. cervicobrachial neuralgia.ti,ab.
21. (monoradicul* or monoradicl*).tw.
22. or/1-21
23. exp headache/ and cervic*.tw.

APPENDIX

24. exp genital diseases, female/
25. genital disease*.mp.
26. or/24-25
27. 23 not 26
28. 22 or 27
29. neck/
30. neck muscles/
31. exp cervical plexus/
32. exp cervical vertebrae/
33. atlanto-axial joint/
34. atlanto-occipital joint/
35. Cervical Atlas/
36. spinal nerve roots/
37. exp brachial plexus/
38. (odontoid* or cervical or occip* or atlant*).tw.
39. axis/ or odontoid process/
40. Thoracic Vertebrae/
41. cervical vertebrae.mp.
42. cervical plexus.mp.
43. cervical spine.mp.
44. (neck adj3 muscles).mp.
45. (brachial adj3 plexus).mp.
46. (thoracic adj3 vertebrae).mp.
47. neck.mp.
48. (thoracic adj3 spine).mp.
49. (thoracic adj3 outlet).mp.
50. trapezius.mp.
51. cervical.mp.
52. cervico*.mp.
53. 51 or 52
54. exp genital diseases, female/
55. genital disease*.mp.
56. exp *Uterus/
57. 54 or 55 or 56
58. 53 not 57
59. 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 58
60. exp pain/
61. exp injuries/
62. pain.mp.
63. ache.mp.
64. sore.mp.
65. stiff.mp.
66. discomfort.mp.
67. injur*.mp.
68. neuropath*.mp.
69. or/60-68
70. 59 and 69
71. Radiculopathy/
72. exp temporomandibular joint disorders/ or exp temporomandibular joint dysfunction syndrome/
73. myofascial pain syndromes/
74. exp "Sprains and Strains"/
75. exp Spinal Osteophytosis/

APPENDIX

76. exp Neuritis/
77. Polyradiculopathy/
78. exp Arthritis/
79. Fibromyalgia/
80. spondylitis/ or discitis/
81. spondylosis/ or spondylolysis/ or spondylolisthesis/
82. radiculopathy.mp.
83. radiculitis.mp.
84. temporomandibular.mp.
85. myofascial pain syndrome*.mp.
86. thoracic outlet syndrome*.mp.
87. spinal osteophytosis.mp.
88. neuritis.mp.
89. spondylosis.mp.
90. spondylitis.mp.
91. spondylolisthesis.mp.
92. or/71-91
93. 59 and 92
94. exp neck/
95. exp cervical vertebrae/
96. Thoracic Vertebrae/
97. neck.mp.
98. (thoracic adj3 vertebrae).mp.
99. cervical.mp.
100. cervico*.mp.
101. 99 or 100
102. exp genital diseases, female/
103. genital disease*.mp.
104. exp *Uterus/
105. or/102-104
106. 101 not 105
107. (thoracic adj3 spine).mp.
108. cervical spine.mp. (6506)
109. 94 or 95 or 96 or 97 or 98 or 106 or 107 or 108
110. Intervertebral Disk/
111. (disc or discs).mp.
112. (disk or disks).mp.
113. 110 or 111 or 112
114. 109 and 113
115. herniat*.mp.
116. slipped.mp.
117. prolapse*.mp.
118. displace*.mp.
119. degenerat*.mp.
120. (bulge or bulged or bulging).mp.
121. 115 or 116 or 117 or 118 or 119 or 120
122. 114 and 121 (2018)
123. intervertebral disk degeneration/ or intervertebral disk displacement/
124. intervertebral disk displacement.mp.
125. intervertebral disc displacement.mp.
126. intervertebral disk degeneration.mp.
127. intervertebral disc degeneration.mp.

APPENDIX

128. 123 or 124 or 125 or 126 or 127
129. 109 and 128
130. 28 or 70 or 93 or 122 or 129
131. animals/ not (animals/ and humans/)
132. 130 not 131
133. exp *neoplasms/
134. exp *wounds, penetrating/
135. 133 or 134
136. 132 not 135
137. Patient Education as Topic/
138. exp Professional-Patient Relations/
139. exp Health Education/
140. exp Consumer Satisfaction/
141. Patient Advocacy/
142. Patient Participation/
143. exp Patient Compliance/
144. (professional patient communication or physician patient communication or doctor patient communication or nurse patient communication or dentist patient communication).tw.
145. (professional patient relation: or physician patient relation: or doctor patient relation: or nurse patient relation: or dentist patient relation:).tw.
146. (professional patient interaction: or physician patient interaction: or dentist patient interaction: or chiropractor patient interaction:).tw.
147. (patient physician communication or patient doctor communication or patient nurse communication or patient dentist communication).tw.
148. (patient professional relation: or patient physician relation: or patient doctor relation: or patient nurse relation: or patient dentist relation:).tw.
149. (patient professional interaction: or patient physician interaction: or patient doctor interaction: or patient nurse interaction: or patient dentist interaction:).tw.
150. (educat: adj (patient: or consumer: or health:)).tw.
151. (information adj (patient: or consumer: or health:)).tw.
152. (advice adj (patient: or consumer: or health:)).tw.
153. consumer health information.tw.
154. (shared decisionmaking or informed choice).tw.
155. (shared decision making or informed choice).tw.
156. pamphlets/ or exp teaching materials/
157. Self Care/
158. Information Dissemination/
159. Information Services/
160. Teaching/
161. or/137-160
162. 136 and 161
163. exp randomized controlled trials as topic/
164. randomized controlled trial.pt.
165. controlled clinical trial.pt.
166. (random* or sham or placebo*).tw.
167. placebos/
168. random allocation/
169. single blind method/
170. double blind method/
171. ((singl* or doubl* or trebl* or tripl*) adj25 (blind* or dumm* or mask*)).ti,ab.
172. (rct or rcts).tw.
173. (control* adj2 (study or studies or trial*)).tw.
174. or/163-173
175. 162 and 174
176. limit 175 to yr="2006 -Current"
177. limit 175 to yr="1902 - 2005"

APPENDIX

178. guidelines as topic/
179. practice guidelines as topic/
180. guideline.pt.
181. practice guideline.pt.
182. (guideline? or guidance or recommendations).ti.
183. consensus.ti.
184. or/178-183
185. 162 and 184
186. limit 185 to yr="2006 -Current"
187. limit 185 to yr="1902 - 2005"
188. meta-analysis/
189. exp meta-analysis as topic/
190. (meta analy* or metaanaly* or met analy* or metanaly*).tw.
191. review literature as topic/
192. (collaborative research or collaborative review* or collaborative overview*).tw.
193. (integrative research or integrative review* or intergrative overview*).tw.
194. (quantitative adj3 (research or review* or overview*)).tw.
195. (research integration or research overview*).tw.
196. (systematic* adj3 (review* or overview*)).tw.
197. (methodologic* adj3 (review* or overview*)).tw.
198. exp technology assessment biomedical/
199. (hta or thas or technology assessment*).tw.
200. ((hand adj2 search*) or (manual* adj search*)).tw.
201. ((electronic adj database*) or (bibliographic* adj database*)).tw.
202. ((data adj2 abstract*) or (data adj2 extract*)).tw.
203. (analys* adj3 (pool or pooled or pooling)).tw.
204. mantel haenszel.tw.
205. (cohrane or pubmed or pub med or medline or embase or psycinfo or psyclit or psychinfo or psychlit or cinahl or science citation indes).ab.
206. or/188-205
207. 162 and 206
208. limit 207 to yr="2006 -Current"
209. limit 207 to yr="1902 - 2005"
210. (ae or to or po or co).fs.
211. (safe or safety or unsafe).tw.
212. (side effect* or side event*).tw.
213. ((adverse or undesirable or harm* or injurious or serious or toxic) adj3 (effect* or event* or reaction* or incident* or outcome*)).tw.
214. (abnormalit* or toxicit* or complication* or consequence* or noxious or tolerabilit*).tw.
215. or/210-214
216. 162 and 215
217. limit 216 to yr="2006 -Current"
218. limit 216 to yr="1902 - 2005"
219. limit 185 to yr="2000 -Current"
220. from 219 keep 1-21

EDWIN J. DE RAAIJ, PT, MSc[1,2] • **RAYMOND W. OSTELO,** PhD[2-4] • **FRANCOIS MAISSAN,** PT, MSc[1,2]
JURGEN MOLLEMA[1] • **HARRIËT WITTINK,** PhD[1]

The Association of Illness Perception and Prognosis for Pain and Physical Function in Patients With Noncancer Musculoskeletal Pain: A Systematic Literature Review

BACKGROUND: In the literature, illness perceptions have been reported to be important psychological factors associated with pain intensity and physical function in individuals with musculoskeletal pain.

OBJECTIVE: To assess the relationship of illness perceptions with pain intensity and physical function in individuals with noncancer musculoskeletal pain.

METHODS: In this systematic review, relevant literature databases, including PubMed, Embase, PsycINFO, CINAHL, and SPORTDiscus, were searched from inception through December 12, 2017. Two authors (E.D.R. and H.W.) independently performed the search procedures, according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses and the A MeaSurement Tool to Assess systematic Reviews guidelines, and the risk-of-bias assessment, using the QUality In Prognosis Studies tool. A qualitative best-evidence synthesis was performed.

RESULTS: A total of 26 articles were included in the review. There were 11 cross-sectional studies concerning associations of illness perceptions with pain intensity and 11 cross-sectional studies of associations of illness perceptions with physical function. For the prognosis of pain intensity by illness perceptions, the authors found 4 longitudinal studies, and for the prognosis of physical function by illness perceptions, the authors found 12 longitudinal studies. All studies except 1 had high risk of bias. Across 15 cross-sectional studies on 9 different musculoskeletal conditions, the researchers found limited to moderate evidence for a consistent direction of the relationship of illness perceptions with pain intensity and physical function. Higher maladaptive illness perceptions imply stronger pain intensity and more limitation in physical function. Evidence in longitudinal studies is lacking, especially on pain.

CONCLUSION: There is limited to moderate evidence for the cross-sectional relationship between illness perceptions and various musculoskeletal conditions. The prognostic value, however, remains unclear. Future research is recommended to investigate the longitudinal relationship between illness perception domains and outcomes in greater detail. *J Orthop Sports Phys Ther 2018;48(10):789-800. Epub 10 May 2018. doi:10.2519/jospt.2018.8072*

KEY WORDS: *disability, low back pain, pain management*

Musculoskeletal pain is a common global condition. The prevalence of this condition is high, and musculoskeletal pain causes many years lived with disability. For instance, global prevalence for low back pain (LBP) is 9.4%, and LBP ranks first among causes of years lived with disability.[41,58] Musculoskeletal pain also poses an economic burden on society. Direct health care costs, social compensation, retirement pensions, and other indirect costs contribute to this load.[3,60] To reduce this burden, effective management of pain and physical function for individuals with musculoskeletal pain is a challenge to society and clinicians.

Emotions, thoughts, beliefs, behaviors, and perceptions are increasingly accepted as important elements in the management of musculoskeletal pain.[39] *Illness perceptions* are the organized representations patients have about their illness and belong to the core concepts of the Common-Sense Model of Self-Regulation of Health and Illness (CSM). The CSM is based on a parallel-processing model that describes behavior in response to health threats. In this model, a health threat is theorized to generate both cognitive representations (danger control) and emotional states of fear and distress (fear control).[33] Based on initial clinical research, 5 illness perception dimensions have been identified.

[1]Research Group Lifestyle and Health, University of Applied Sciences Utrecht, Utrecht, the Netherlands. [2]Department of Health Sciences, VU University, Amsterdam, the Netherlands. [3]EMGO Institute for Health and Care Research, Amsterdam, the Netherlands. [4]Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, the Netherlands. This study is supported by a grant from the Dutch government (NWO-023.005.029). This systematic review was registered with PROSPERO in 2016 (CRD42016026759). The authors certify that they have no affiliations with or financial involvement in any organization or entity with a direct financial interest in the subject matter or materials discussed in the article. Address correspondence to Edwin de Raaij, University of Applied Sciences Utrecht, Research Group Lifestyle and Health, Heidelberglaan 7 3584 CS Utrecht, the Netherlands. E-mail: edwin.deraaij@hu.nl

1. Identity: the label or name given to the condition by patients and the symptoms that are perceived to go with it
2. Timeline-chronic: how long the patient believes the illness will last
3. Consequences: how strong the impact of the patient's illness is on, for example, pain or physical function
4. Causal beliefs: the patient's beliefs about what causes the illness
5. Control beliefs: the patient's beliefs about how to control or recover from the illness

Ongoing research has explored and added the dimensions of timeline-cyclical (periodic changes in symptoms), coherence (making sense of the illness), emotional representations (impact on emotional level), and concern (anxiousness about the illness) to the CSM.[6,43]

Recent research shows that illness perceptions have associations with several outcomes in acute and chronic illness, including self-management behaviors and quality of life.[35] These perceptions are associated with outcomes in a variety of diseases.[19] Although promising, the literature is not unambiguous. For instance, the illness perception dimensions of timeline-chronic, consequences, and control beliefs have been recognized as prognostic factors for limitation in physical function in patients with LBP.[12,40] But, other studies have shown different perception dimensions to be associated with outcomes of LBP.[2,15]

It has been suggested that changes in illness perceptions may predict subsequent physical function in conditions such as LBP, but relatively few intervention studies have been conducted. In a randomized controlled trial for LBP, Siemonsma et al[53] concluded that there were improvements in patient-relevant physical activities at 18-week follow-up after cognitive treatment of illness perceptions. This study and others have shown that influencing perceptions can improve physical functioning.[42]

Evaluating and addressing illness perceptions may be an important component in the treatment of patients with musculoskeletal pain. However, to the authors' knowledge, no systematic review has evaluated the relationship between illness perceptions and pain intensity and physical functioning in individuals with musculoskeletal pain. Therefore, this review explores the relationship of illness perceptions with pain intensity and physical function in patients with musculoskeletal pain in both cross-sectional and longitudinal studies. This review specifically asked (1) what associations illness perceptions may have with pain intensity and physical function in patients with musculoskeletal pain and (2) whether illness perceptions may be prognostic for pain intensity and physical function in patients with musculoskeletal pain.

METHODS

This systematic review was written in accordance with Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines[38] and the A MeaSurement Tool to Assess systematic Reviews checklist.[52] Details of the protocol for this study were registered with PROSPERO and can be accessed at http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42016026759.

The following terms with their definitions were used in this review: *musculoskeletal pain* is pain felt within the context of the following musculoskeletal conditions, according to the European Musculoskeletal Conditions Surveillance and Information Network[11]: (1) joint conditions (ie, rheumatoid arthritis, osteoarthritis), (2) bone conditions (ie, osteoporosis), (3) spinal disorders (ie, LBP), (4) regional and widespread pain disorders, (5) musculoskeletal injuries, and (6) multisystem inflammatory diseases. *Illness perceptions* are the organized representations patients have about their illness that belong to the core concepts of the CSM.[34] Illness perceptions can be assessed by 3 validated questionnaires: (1) the Illness Perception Questionnaire (IPQ),[59] (2) the Illness Perception Questionnaire revised (IPQ-R),[43] and (3) the Brief IPQ.[29] All 3 questionnaires have good psychometric properties.[36] **TABLE 1** presents the number of questions per illness perception dimension and their outcome scores.

The authors of this systematic review hypothesized that a high score on the dimensions of consequences, timeline, identity, concern, and emotional representations would be indicative of maladaptive illness perceptions. On the dimensions of control beliefs and coherence, a low score

TABLE 1 Number of Questions Per Illness Perception Dimension and Their Outcome Score Ranges

Domain	IPQ	IPQ-R	Brief IPQ
Consequences	7 items (7-35)*	6 items (6-30)*	1 item (0-10)†
Timeline-chronic	3 items (3-15)*	6 items (6-30)*	1 item (0-10)†
Timeline-cyclical		4 items (4-20)*	
Control beliefs-personal	6 items (6-30)*	6 items (6-30)*	1 item (0-10)†
Control beliefs-treatment		5 items (5-25)*	1 item (0-10)†
Identity	12 items	14 items	1 item (0-10)†
Concern			1 item (0-10)†
Coherence		5 items (5-25)*	1 item (0-10)†
Emotional representations		6 items (6-30)*	1 item (0-10)†
Causal beliefs	18 items (no sum)	19 items (4 categories,‡ no sum)	1 item (open)

Abbreviations: IPQ, Illness Perception Questionnaire; IPQ-R, Illness Perception Questionnaire revised.
**Scored on a 5-point Likert scale (1-5).*
†Scored on a numeric rating scale (0-10).
‡Psychological, risk, immune, chance.

would indicate maladaptive illness perceptions.[4] The authors considered a positive association between illness perceptions and higher pain intensity or limited physical function to constitute maladaptive illness perceptions. Therefore, the associations found for the illness perception dimensions of control beliefs (personal/treatment) and coherence were converted before being presented in this study's results. The illness perception dimension of causal beliefs is the only dimension that has a nominal measurement scale. Because of this nominal scale, it was not possible within this review to report an association or prognostic value of the illness perception dimension of causal beliefs with pain intensity or physical function.

Pain is defined as an unpleasant sensory and emotional experience associated with actual or potential tissue damage or described in terms of such damage.[30] Questionnaires assessing pain intensity may have opposing scores. For instance, a high score on the numeric pain-rating scale indicates higher pain intensity, whereas a high score on the bodily pain dimension of the Medical Outcomes Study 36-Item Short-Form Health Survey indicates less pain. To resolve such discrepancies, the authors converted all pain measurement scales so that higher scores would indicate higher pain intensity.

Physical function is the self-reported capability to perform physical activities, rather than an objective assessment of performance. This includes the functioning of one's upper extremities (dexterity), lower extremities (walking or mobility), and central regions (neck, back), as well as instrumental activities of daily living, such as running errands.[23] Questionnaires assessing physical function may also have opposing scores. For instance, a high score on the Roland-Morris Disability Questionnaire indicates more limitation, but a high score on the physical functioning dimension of the Medical Outcomes Study 36-Item Short-Form Health Survey indicates less limitation. To resolve such differences, the reported association was converted so that higher scores would indicate more limitation in physical function.

Because most longitudinal studies had follow-up assessments at 3, 6, and 12 months, the authors summarized the data from the longitudinal studies by time intervals of less than 6 months, 6 to 12 months, and greater than 12 months.

Data Sources and Searches

Potentially relevant studies were identified through searches in the following electronic databases: PubMed, Embase, PsycINFO, CINAHL, and SPORTDiscus. The databases were searched from inception to December 12, 2017. A comprehensive search strategy was developed in consultation with a medical information specialist (J.M.). The search strategy consisted of 2 major elements: musculoskeletal pain and illness perceptions. The authors used 2 search strategies for musculoskeletal pain and combined the results: one strategy used terms regarding pain in combination with musculoskeletal diseases and/or musculoskeletal systems, and the other strategy used terms regarding musculoskeletal pain. For each search, the researchers used all known synonyms and related terms to develop as sensitive a search as possible. The key terms were mapped to medical subject headings, and title and abstract search words and phrases were added.

The authors built the search string for PubMed and then translated it to the other databases. All databases were individually searched. The researchers imported identified references into RefWorks (ProQuest LLC, Ann Arbor, MI) and removed duplicates with the close deduplication algorithm from RefWorks. They manually verified the result of the automatic deduplication. The search strings for all databases are available on request from the corresponding author. In addition to the database searches, the authors also searched the gray literature, including the following electronic sources up to October 5, 2016: the DART-Europe E-theses Portal, Open Access Theses and Dissertations, Networked Digital Library of Theses and Dissertations, ClinicalTrials.gov, and World Health Organization International Clinical Trials Registry Platform. In addition, the references of the included articles and recently published review articles were screened for additional publications.

Study Selection

In the first round, 2 authors (E.D.R. and H.W.) independently reviewed all titles and abstracts and excluded all studies that did not fulfill the inclusion criteria. If an abstract was noninformative but potentially relevant, the full-text article was read. In the second round, the full texts of all articles were read for fulfillment of all inclusion criteria and selected by 2 independent authors. Articles on psychometric properties or with qualitative designs were excluded. Any disagreement was resolved by discussion and consensus with a third author (R.O.).

The studies had to meet the following criteria for final inclusion: (1) study population of individuals with musculoskeletal pain, as defined by the European Musculoskeletal Conditions Surveillance and Information Network,[11] (2) measures of illness perceptions with questionnaires based on the CSM, (3) measures of pain and physical function with self-reported questionnaires, and (4) study designs that included cohort studies, cross-sectional studies, and randomized controlled trials.

To answer the research questions concerning associations of illness perceptions with pain intensity and physical function, the authors considered cross-sectional studies or longitudinal studies most appropriate. To answer the research questions concerning prognoses by illness perceptions of increased pain intensity and increased limitation in physical function, the researchers considered longitudinal studies most appropriate. They excluded qualitative studies.

Data Extraction and Quality Assessment

Two authors developed and independently completed the data-extraction form,

which included author, publication year, study design, number of participants, study setting, characteristics of the study population (eg, musculoskeletal disorder, type of illness perceptions), measurement instruments of pain and limitation in physical function, and outcome and statistical measures (correlations, odds ratios, and regression coefficients).

The risk of bias was assessed using the QUality in Prognosis Studies (QUIPS) tool[27] by 2 authors independently. This tool has 31 items that are scored as "yes" (fulfilled), "/" (partial), "no" (not fulfilled), or "?" (unclear whether criterion is fulfilled). The 31 items cover 6 domains for potential bias: study participation, study attrition, prognostic factor measurement, outcome measurement, study confounding, and statistical analysis. These domains are labeled "high," "moderate," or "low" risk of bias, based on the individual item's score within each domain, as described by Hayden et al.[22] A study has a low risk of bias if all 6 domains are rated as low risk of bias. Any disagreement was resolved by discussion and consensus. It is recommended not to report a total score of the 31 items for overall study quality.[22]

Data Synthesis and Analysis

Descriptive statistics were used to summarize the medical condition, number of participants, age, sex, study design, and questionnaires used for illness perceptions/pain intensity/physical function across all included studies. Extraction of results focused on obtaining unadjusted and adjusted correlations, regression coefficients, relative risks, odds ratios, and 95% confidence intervals. To explore possible publication bias or outcome reporting bias, funnel plots were made by plotting all extracted data against the number of participants in each study.

To assess statistical heterogeneity, the I^2 test was used. As proposed by Higgins et al,[24] a value higher than 50% was considered an indicator of substantial heterogeneity. As outcomes were considered too heterogeneous, the authors refrained from statistical pooling and summarized the evidence qualitatively, according to Hayden et al[21] (**TABLE 2**).

RESULTS

Study Selection

The literature search generated a total of 1418 references: PubMed, 411; PsycINFO, 381; Embase, 314; CINAHL, 253; and SPORTDiscus, 59. A total of 114 references were identified in the gray literature. After screening for duplicates (J.M.), 1045 were included for screening. Two authors (E.D.R. and H.W.) independently screened all 1159 studies for eligibility, using the inclusion/exclusion criteria. A total of 26 studies met these criteria and were included in the review (**FIGURE**).

Study Characteristics

TABLE 3 shows the study characteristics of the 26 included papers. The number of participating patients in studies varied from 11[17] to 2113.[25] Twelve different mus-

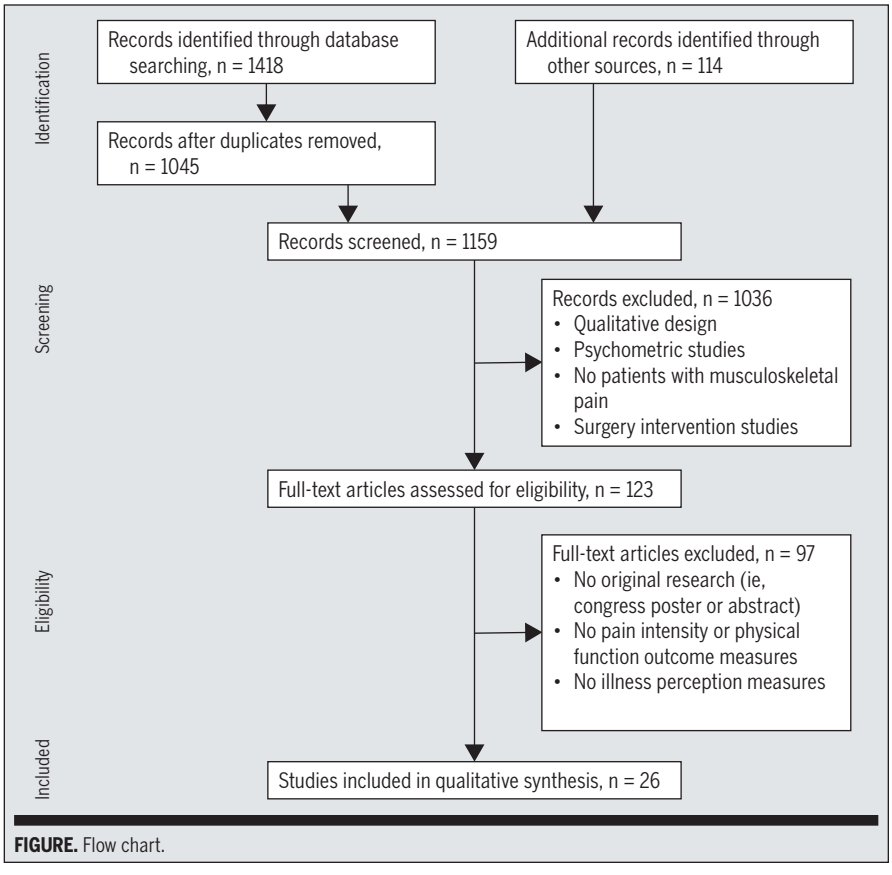


FIGURE. Flow chart.

TABLE 2 Levels of Evidence[21]

Level	Description
Strong	Consistent findings (defined as greater than 75% of studies showing the same direction of effect) in multiple low-risk-of-bias studies
Moderate	Consistent findings in multiple high-risk-of-bias studies and/or 1 study with low risk of bias
Limited	1 study available
Conflicting	Inconsistent findings across studies
No evidence	No association between variables for association or prognosis

culoskeletal conditions were identified: (1) rheumatoid arthritis,[18,45,48,50,51,56] (2) LBP,[2,7,12,16,37,49] (3) chronic pain,[14,32,42,44] (4) chronic headache,[4] (5) fibromyalgia,[55,57] (6) systemic lupus erythematosus,[17] (7) hand problems,[26] (8) chronic repetitive strain injuries,[54] (9) acute injury,[8] (10) chronic orofacial pain,[13] (11) gout,[9] and (12) osteoarthritis of the knee.[27]

For this study's research questions concerning associations of illness perceptions with pain intensity, the authors found 11 cross-sectional studies (**APPENDIX A**, available at www.jospt.org), and for illness perceptions with physical function, 11 cross-sectional studies (**APPENDIX B**, available at www.jospt.org). The study of Groarke et al[18] has a longitudinal design, but cross-sectional associations were also reported and were used to answer the questions of illness perceptions' association with pain intensity or physical function. For the prognosis of pain intensity by illness perception, the researchers found 4 longitudinal studies (**APPENDIX C**, available at www.jospt.org), and for the prognosis of physical function by illness perception, 10 longitudinal studies (**APPENDIX D**, available at www.jospt.org).

Risk of Bias

All studies but 1 had a high risk of bias (**TABLE 4**). The study by Foster et al[12] was

TABLE 3 Study Characteristics

	Patient Characteristics			Measurement Instruments			
Study	**Condition**	**Study Sample**	**Study Design**	**IP**	**Pain**	**Physical Function**	**Comments**
Scharloo et al[50]	RA, 12.0 ± 8.2 y	n = 84; 51.7 ± 12.6 y; 75% female	Cross-sectional	IPQ	...	SF-20, HAQ	SF-20: higher score, fewer limitations; HAQ: higher score, more limitations
Scharloo et al[51]	RA, 12.4 ± 8.5 y	n = 71; 52.2 ± 12.2 y; 75% female	Longitudinal	IPQ	VAS	HAQ	Time interval, 12 ± 2 mo VAS: higher score, more pain; HAQ: higher score, more limitations
Groarke et al[18]	RA, 12.6 ± 10.8 y	n = 75 at baseline and n = 52 at follow-up; 60.1 ± 12.1 y; 100% female	Cross-sectional, longitudinal	IPQ	AIMS	AIMS	Time interval, 12 and 24 mo AIMS: higher score, more pain and more limitations
Goodman et al[17]	Systemic lupus erythematosus	n = 11; age and sex unknown	Longitudinal	IPQ-R	...	SF-36	Time interval, 7 wk SF-36: higher score, fewer limitations
Stuifbergen et al[55]	Fibromyalgia, >1 y	n = 160; 18-75 y; 100% female	Cross-sectional	IPQ-R	...	SF-36	SF-36: higher score, fewer limitations
Hill et al[26]	Musculoskeletal hand problems (chronic)	n = 2113; 65.4 ± 9.6 y; 63% female	Cross-sectional	IPQ-R	AIMS	AIMS	AIMS: higher score, more pain and more limitations
Moss-Morris et al[42]	Chronic pain, 7.1 ± 6.9 y	n = 98; 42.4 ± 9.5 y; 65% female	Longitudinal	IPQ-R	...	SF-36	Time interval, 4 wk SF-36: higher score, fewer limitations
Sluiter and Frings-Dresen[54]	Chronic RSI, 5.8 ± 3.2 y	n = 1121; 40.8 ± 8.7 y; 67% female	Cross-sectional	Brief IPQ	SF-36, VAS	SF-36	VAS: higher score, more pain; SF-36: higher score, less pain and fewer limitations
Foster et al[12]	LBP, >1 mo to <3 y	n = 1591; 43.9 ± 10.3 y; 59% female	Longitudinal	IPQ-R	...	RMDQ	Time interval, 6 mo RMDQ: higher score, fewer limitations
van Wilgen et al[57]	Fibromyalgia, about 10 y	n = 51; 44.0 ± 10.0 y; 92% female	Cross-sectional	IPQ-R	NRS	...	NRS: higher score, more pain
Broadbent et al[4]	Persistent headache	n = 65; age and sex unknown	Cross-sectional	Brief IPQ	NRS	...	NRS: higher score, more pain
Chaboyer et al[8]	Injury (acute)	n = 114; <45 y, 24%; 33% female	Longitudinal	IPQ-R	...	SF-36	Time interval, 6 mo SF-36: higher score, fewer limitations
Galli et al[13]	Chronic orofacial pain, >3 mo	n = 82; 45.7 ± 16.0 y; 75% female	Longitudinal	IPQ-R	NRS	GCPS	Time interval, 3 and 6 mo NRS: higher score, more pain; GCPS: higher score, more limitations
Nicklas et al[44]	Nonmalignant chronic pain, 10 y (6 mo to 42 y)	n = 217; age and sex unknown	Cross-sectional	IPQ-R	NRS	...	NRS: higher score, more pain
Dalbeth et al[9]	Gout, <10 y	n = 132; 57 y (21-85 y); 0% female	Longitudinal	Brief IPQ	...	HAQ	Time interval, 12 mo HAQ: higher score, more limitations

Table continues on page 794.

TABLE 3 Study Characteristics (continued)

	Patient Characteristics			Measurement Instruments			
Study	**Condition**	**Study Sample**	**Study Design**	**IP**	**Pain**	**Physical Function**	**Comments**
van Os et al[56]	RA, 12.5 ± 10.7 y	n = 230; 57.3 ± 15.1 y; 76% female	Cross-sectional	IPQ-R	EQ-5D	HAQ	EQ-5D: higher score, more pain; HAQ: higher score, more limitations
Gillanders et al[14]	Chronic pain, >6 mo	n = 150; 50.8 ± 13.2 y; 67% female	Cross-sectional	IPQ-R	MPQ	RMDQ	MPQ: higher score, more pain; RMDQ: higher score, fewer limitations
Campbell et al[7]	LBP, 22.3% for >3 y	n = 488; 47.4 ± 9.0 y; 62% female	Longitudinal	IPQ-R	...	CPGS	Time interval, 6 mo and 5 y CPGS: higher score, more pain-related limitations
Glattacker et al[16]	Chronic LBP, 96% for >1 y	n = 105; 54.9 ± 11.0 y; 63% female	Longitudinal	IPQ-R	VAS	SF-36, ODI	Time interval, 1 mo and 6 mo VAS: higher score, more pain; SF-36: higher score, fewer limitations; ODI: higher score, more limitations
Norton et al[45]	RA, 5.7 ± 5.7 y	n = 227; 57.7 ± 15.0 y; 76% female	Cross-sectional	IPQ-R	EQ-5D	HAQ	EQ-5D: higher score, more pain; HAQ: higher score, more limitations
Rezaei et al[48]	RA, 12.5 ± 10.7 y	n = 100; 45.5 ± 14.0 y; 72% female	Cross-sectional	Brief IPQ	RAPS	...	RAPS: higher score, less pain
Bishop et al[2]	LBP, 12.6% for <6 wk	n = 485; 55.0 ± 15.1 y; 69% female	Longitudinal	Brief IPQ	...	RMDQ	Time interval, 2 wk, 3 and 6 mo RMDQ: higher score, fewer limitations
Hirsch[27]	OA of the knee, 6.3 ± 7.3 y	n = 141; 63.8 ± 11.2 y; 62% female	Longitudinal	IPQ-R	VAS, WOMAC	...	Time interval, 3 and 9 wk VAS: higher score, more pain
Roios et al[49]	LBP, >3 mo	CT group, n = 213 and PT group, n = 125; 46.2 ± 11.6 y and 48.0 ± 13.0 y; 51% female and 70% female	Cross-sectional	IPQ-R	...	ODI	ODI: higher score, more limitations
Järemo et al[32]	Chronic pain	n = 152; 46.3 ± 14 y; 91% female	Cross-sectional	IPQ-R	SF-36	SF-36	SF-36: higher score, fewer limitations
Leysen et al[37]	Chronic LBP, >3 mo	n = 48; 47.0 ± 15 y; 61% female	Cross-sectional	IPQ-R	...	ODI	ODI: higher score, more limitations

Abbreviations: AIMS, Arthritis Impact Measurement Scales; CPGS, Chronic Pain Grade Scale; CT, chiropractic; EQ-5D, EuroQol-5 dimensions; GCPS, Graded Chronic Pain Scale; HAQ, Health Assessment Questionnaire; IP, illness perception; IPQ, Illness Perception Questionnaire; IPQ-R, Illness Perception Questionnaire revised; LBP, low back pain; MPQ, McGill Pain Questionnaire; NRS, numeric rating scale; OA, osteoarthritis; ODI, Oswestry Disability Index; PT, physical therapy; RA, rheumatoid arthritis; RAPS, Rheumatoid Arthritis Pain Scale; RMDQ, Roland-Morris Disability Questionnaire; RSI, repetitive strain injury; SF-20, Medical Outcomes Study 20-Item Short-Form Health Survey; SF-36, Medical Outcomes Study 36-Item Short-Form Health Survey; VAS, visual analog scale; WOMAC, Western Ontario and McMaster Universities Osteoarthritis Index.

scored as low risk on all 6 domains. There was considerable variance between studies in percentages of items scored "yes," ranging from 29%[47] to 87%,[2,12,13,28,45] with an average of 66%. The QUIPS domain "study confounding" was most frequently scored as high risk of bias, and "study participation" was scored most frequently as low risk of bias. After initial assessment, there was 82% agreement on the risk-of-bias assessment of the 6 QUIPS domains between the 2 independent reviewers (E.D.R. and H.W.). Differences were resolved between the 2 assessors without the need to consult a third assessor. Funnel plots were processed and showed risk of publication bias for all illness perception dimensions.

Results of Individual Studies

The data extraction for all 26 studies is presented in tables comprising **APPENDICES A** through **D**. The authors found a total of 321 different variables for illness perceptions' association with or prognosis for pain intensity and physical function. These variables ranged from univariate, multivariate, beta, and odds ratio to relative risk. For the prognostic value of illness perceptions for pain, the researchers found only 4 studies with short- or medium-term results.[16,18,27,51] They found no studies with long-term results of more than 1 year.

Synthesis of Results

The authors considered methodological heterogeneity based on the I^2 test of more than 50% on all associations and

TABLE 4 Scores on Methodological Quality Assessment

Study	Study Participation	Study Attrition	Prognostic Factor Measurement	Outcome Measurement	Study Confounding	Statistical Analysis and Reporting
Cross-sectional designs						
Scharloo et al[50]	Moderate	Not applicable	Low	Low	Moderate	Low
Groarke et al[18]	Low	Not applicable	High	Low	High	Low
Stuifbergen et al[55]	Moderate	Not applicable	Low	Low	High	Low
Hill et al[26]	Low	Not applicable	Low	Low	High	Low
Sluiter and Frings-Dresen[54]	Moderate	Not applicable	Moderate	Low	High	Low
van Wilgen et al[57]	Moderate	Not applicable	Low	Low	High	Low
Broadbent et al[4]	High	Not applicable	Moderate	Moderate	High	Low
Nicklas et al[44]	Low	Not applicable	Moderate	Moderate	High	Low
van Os et al[56]	Moderate	Not applicable	Low	Low	Low	Low
Gillanders et al[14]	Low	Not applicable	Low	Low	High	Low
Norton et al[45]	Low	Not applicable	Low	Low	High	Low
Rezaei et al[48]	Low	Not applicable	Low	Low	High	Low
Roios et al[49]	Low	Not applicable	Low	Low	High	Low
Järemo et al[32]	Low	Not applicable	Moderate	Low	High	Low
Leysen et al[37]	Low	Not applicable	Low	Low	High	Low
Longitudinal designs						
Scharloo et al[51]	Low	High	High	Low	Low	Low
Groarke et al[18]	Low	Moderate	High	Low	High	Low
Goodman et al[17]	High	High	Moderate	Low	High	Low
Moss-Morris et al[42]	Low	Moderate	Moderate	Low	High	Low
Foster et al[12]	Low	Low	Low	Low	Low	Low
Chaboyer et al[8]	Low	High	Moderate	Low	High	Moderate
Galli et al[13]	Low	High	Low	Low	Low	Low
Dalbeth et al[9]	Low	Low	Low	Low	High	Low
Glattacker et al[16]	Low	Moderate	Low	Low	Low	Low
Campbell et al[7]	Low	Low	Low	Low	High	Low
Bishop et al[2]	Low	Moderate	Low	Low	Low	Low
Hirsch[27]	Low	Low	Moderate	Low	Low	Low

prognostic outcome scores of the included studies.[24] Clinical heterogeneity differed among studies due to the diversity of study characteristics, such as number of patients, age, musculoskeletal condition, and duration of symptoms. The measurement instruments of illness perceptions, pain intensity, and physical function were also diverse across studies. Three different versions of the IPQ were used: the IPQ,[59] the IPQ-R,[43] and the Brief IPQ.[6] For the outcome measures of pain intensity, 8 different instruments were used, and for limitations in physical function, 8 instruments were employed (**TABLE 3**).

Due to heterogeneity, the authors could not perform a meta-analysis; the data were summarized qualitatively.

TABLE 5 shows the level of evidence, according to Hayden et al,[21] for illness perception dimensions related to pain intensity or physical function in musculoskeletal pain.

Association of Illness Perceptions With Pain or Physical Function

Pain Intensity There is moderate evidence in 9 cross-sectional studies for a positive association (ie, maladaptive illness perceptions are associated with higher levels of pain), based on univariate regression, of all illness perception dimensions with pain intensity (**TABLE 5**). This positive direction of the associations of all illness perceptions with pain intensity was consistent across 8 different conditions (**APPENDIX A**). The strongest associations were found for the illness perception dimensions of consequences and identity. For instance, the study by Gillanders et al[14] reported a positive association of illness perception (consequences domain) with increased pain ($r = 0.47$) in 150 patients with chronic pain, meaning a high score on the dimension of consequences is associated with increased pain intensity.

TABLE 5 Evidence for Illness Perception Dimensions Related to Pain or Physical Function in Musculoskeletal Pain*

	Cross-sectional		Longitudinal: Pain			Longitudinal: PF		
Dimension	**Pain**	**PF**	**T1**	**T2**	**T3**	**T1**	**T2**	**T3**
Consequences	+	+	+	+		+		+
Timeline-chronic	+	+		+		+	+	
Timeline-cyclical	+	+				+		
Control beliefs-personal	+	+	+			+		+
Control beliefs-treatment	+	+	+				+	
Identity	+	+		+		+	+	+
Concern	+	+				+		+
Coherence	+	+	–			+		
Emotional representations	+	+	+			+		+

Strong *Moderate* *Limited* *Conflicting* *No evidence*

Abbreviations: PF, physical function; T1, time interval of less than 6 months; T2, time interval of 6 to 12 months; T3, time interval of greater than 12 months.
**From Hayden et al.*[21]

Physical Function The authors found moderate evidence in 10 cross-sectional studies for a positive association (ie, maladaptive illness perceptions are associated with limitations in physical function), based on univariate regression, of all illness perception dimensions with physical function (**TABLE 5**). The positive direction of the relationship of all illness perception domains with physical function was consistent across 8 different conditions. The strongest associations were found for the dimensions of consequences and identity (**APPENDIX B**). For instance, the study by Sluiter and Frings-Dresen[54] reported a positive association between illness perception (consequences domain) and increased limitation in physical function (r = 0.49) in 1122 patients with chronic repetitive strain injury, meaning that a high score on the illness perception dimension of consequences was associated with increased limitations in physical function.

Prognostic Value of Illness Perceptions for Pain Intensity or Physical Function

Pain Intensity Two longitudinal studies[16,27] with a time interval of less than 6 months found moderate evidence of illness perceptions being prognostic for greater pain intensity on the dimension of consequences, and limited evidence for maladaptive illness perceptions being prognostic for greater pain intensity on the dimensions of control beliefs (personal/treatment), coherence, and emotional representations (**APPENDIX C**).

Three longitudinal studies,[16,18,51] with time intervals of 6 to 12 months found limited evidence for illness perceptions being prognostic for more pain intensity on the illness perception dimensions of consequences, timeline-chronic, and identity (**APPENDIX C**).

None of the studies reported evidence for pain at the time interval of greater than 12 months.

Physical Function Nine longitudinal studies[7-9,12,13,16-18,42] with a time interval of less than 6 months found moderate evidence for illness perceptions being prognostic for more limitations in physical function on the illness perception dimensions of consequences, timeline (chronic/cyclical), control beliefs (personal), identity, and emotional representations, and limited evidence for illness perceptions being prognostic for more limitations in physical function on the dimensions of concern and coherence (**APPENDIX D**). The positive direction of the relationship of these illness perceptions with physical function is consistent across 8 different conditions.

One longitudinal study[16] with a time interval of 6 to 12 months found limited evidence for illness perceptions being prognostic for more limitations in physical function on the dimensions of timeline-chronic, control beliefs (treatment), and identity (**APPENDIX D**).

Two longitudinal studies,[2,7] with a time interval of greater than 12 months, found moderate evidence for illness perceptions being prognostic for more limitations in physical function on the illness perception dimensions of consequences, control beliefs (personal), and identity, and limited evidence for maladaptive illness perceptions being prognostic for more limitations in physical function on the dimensions of timeline-chronic, concern, and emotional representations (**APPENDIX D**).

DISCUSSION

The aim of this study was to systematically review the relationship between illness perceptions and pain intensity or physical function in patients with musculoskeletal pain. For cross-sectional study designs, there is moderate evidence for all illness perception dimensions being positively associated with pain intensity and physical function. Overall, the evidence for the longitudinal relationship was less evident. For pain intensity, there is moderate evidence for

the illness perception dimension of consequences to be prognostic at a time interval of less than 6 months. For physical function, there is moderate evidence that the dimensions of consequences, timeline (chronic/cyclical), control beliefs (personal), and identity are prognostic factors for physical function at a time interval of less than 6 months. In addition, there is moderate evidence that the illness perception dimensions of consequences, control beliefs (personal), and identity are prognostic factors for physical function at a time interval of greater than 12 months.

Across studies, the strength of associations and prognoses varied among all illness perception domains (**APPENDICES A** through **D**). The authors found no explanation for this variation, based on differences in number of participants, age, symptom duration, or the questionnaires used to assess illness perceptions, pain intensity, and physical function. Comparison of these findings with previous systematic reviews on illness perception and musculoskeletal pain is not possible, due to an absence of these studies in the scientific literature. Comparing the relevance of the present study's results with other reviews in the field of illness perception, the authors found their results to be in line. In 2 meta-analyses on illness perception, the same sizes of associations are reported.[10,19] One study[19] included a total of 23 illnesses (mostly nonmusculoskeletal) and outcome measures concerning physical health–related quality of life. The other study[10] included a total of 31 conditions (varying from musculoskeletal to cancer) and outcome measures on depression, anxiety, and quality of life. This means that the strength of the observed relations of illness perceptions with pain intensity and physical function in this review is comparable to those found in other studies on this topic.

Prognosis is the probable course and outcome of a health condition over time, and in explanatory prognostic research, 3 phases can be identified[20]: phase 1, identifying associations; phase 2, testing independent associations; and phase 3, understanding prognostic pathways. The authors identified no phase 3 studies, 9 phase 2 studies,[2,10,12,13,16,27,28,31,51] and 20 phase 1 studies.[4,7,8,9,14,17,18,25,32,37,42,44,45,48,49,50,54,55,56,57] This means for phase 1 studies that illness perceptions, as prognostic factors, were reported without controlling for other prognostic factors. Therefore, the interpretation of reported associations and prognoses should be treated with caution.

This is the first review to the authors' knowledge that focuses on the relationship between illness perceptions and pain intensity and physical function in individuals with musculoskeletal pain. The search strategy was designed in collaboration with a librarian information specialist (J.M.). It is known that the contribution of a librarian information specialist in designing a search strategy for systematic reviews is highly correlated with the quality of the reported search strategy.[47] Therefore, the authors consider their search strategy a strong element of the study. The risk-of-bias assessment was performed according to the recommendation of Hayden et al,[22] and led to determination of high risk of bias for all studies but 1.

The quality of the studies included in this review is not in line with the reported study quality found in another review on prognostic factors of musculoskeletal pain.[1] After performing a sensitivity analysis, that study used a cutoff point of 9 on a 15-item scale (60%) as indicating a low-risk-of-bias study, while the present study did not employ a total score to indicate overall study quality. As a result, this assessment of risk of bias may be called strict, a characteristic that should be considered when interpreting conclusions about the quality of each individual study included in this review.

Limitations

There are several limitations of this systematic review to be considered. First, the diversity of musculoskeletal pain conditions included may have influenced this synthesis. However, despite this diversity, the direction of the association is consistent throughout the included studies. Second, the strength of the association could not be assessed in a meta-analysis; therefore, a qualitative analysis of the data was performed. Because of this limitation, the authors cannot report on the strength of the pooled association or prognostic factor.

The association of illness perceptions with prognosis for pain intensity and physical function, though small in strength for cross-sectional studies and limited in evidence for longitudinal studies overall, seems to be independent of the nature of the musculoskeletal condition. This finding aligns with another systematic review that focused on generic prognostic factors for musculoskeletal pain.[1] The authors found 15 possible relevant prognostic factors identified in studies of patients with at least 2 different pain sites. Regardless of the location of the musculoskeletal pain, generic factors such as pain intensity, widespread pain, high functional disability, somatization, and movement restriction were reported as prognostic factors for pain. The authors see the same pattern in the present study; regardless of the musculoskeletal pain condition, the direction of the relationship was consistent. As a result, this study provides supplementary information for understanding the role of illness perception in musculoskeletal pain.

The authors considered higher scores on the illness perception domains of consequences, timeline, identity, concern, and emotional representations, and lower scores on the domains of control beliefs (personal/treatment) and coherence, to be maladaptive, because they are positively associated with, or prognostic for, increased pain intensity and increased limitations in physical function. The consistency of these findings, independent of musculoskeletal pain condition, contributes to understanding the role of illness perception in musculoskeletal pain. For instance, baseline assessment of maladaptive illness perceptions in patients

with musculoskeletal pain provides some insight in how patients themselves think about their pain or physical function.

For clinicians, addressing patients' illness perceptions may open new possibilities for management. In this review, the authors found 3 validated questionnaires for illness perception assessment. These findings show no real differences of strength of association between illness perceptions and pain or physical function among these questionnaires. The most used questionnaire was the IPQ-R, which consists of 71 items. The Brief IPQ has 9 items. The latter might have less patient burden and so may be preferred for use in daily practice.

Changing maladaptive illness perceptions may have a positive influence on pain and physical function. The authors found 1 randomized controlled trial that addressed maladaptive illness perceptions of patients with chronic LBP with a cognitive treatment protocol, which showed promising results in a better level of patient-specific physical function after 18 weeks.[14] The cognitive treatment targeted maladaptive illness perceptions of patients about their back pain and aimed to alter these perceptions. In 10 to 14 one-hour treatment sessions, maladaptive illness perceptions were assessed and challenged, and alternative illness perceptions were formulated. This was done by mapping the illness perceptions with the IPQ-R and further exploring these perceptions with a Socratic style of dialog. More research on incorporating maladaptive illness perceptions in intervention studies for the management of musculoskeletal pain is recommended, as well as research on the prognoses of illness perceptions for pain and physical function.

CONCLUSION

There is limited to moderate evidence for the cross-sectional relationship between illness perceptions and various musculoskeletal pain conditions. The prognostic value of these relationships, however, remains unclear. For future research, the authors suggest investigating the longitudinal relationship between illness perception domains and outcome in more detail. In addition, studies on the feasibility and impact of incorporating illness perceptions in interventions for the management of musculoskeletal pain are recommended. ◉

KEY POINTS

FINDINGS: There is limited to moderate evidence from cross-sectional studies showing that illness perceptions are related to pain intensity and physical function in individuals with a range of musculoskeletal pain conditions. The authors of this systematic review recommend investigating the longitudinal relationship between illness perception domains and outcome in more detail.

IMPLICATIONS: Addressing patients' illness perceptions opens new possibilities for clinical management. Studies on the feasibility and impact of incorporating illness perceptions in interventions for the management of musculoskeletal pain are recommended.

CAUTION: Due to the heterogeneity of included studies, these findings are not robust and should be interpreted with caution.

ACKNOWLEDGMENTS: *The authors thank Cas Kruitwagen, MSc, for statistical advice.*

REFERENCES

1. Artus M, Campbell P, Mallen CD, Dunn KM, van der Windt DA. Generic prognostic factors for musculoskeletal pain in primary care: a systematic review. *BMJ Open*. 2017;7:e012901. https://doi.org/10.1136/bmjopen-2016-012901
2. Bishop FL, Yardley L, Prescott P, Cooper C, Little P, Lewith GT. Psychological covariates of longitudinal changes in back-related disability in patients undergoing acupuncture. *Clin J Pain*. 2015;31:254-264. https://doi.org/10.1097/AJP.0000000000000108
3. Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D. Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. *Eur J Pain*. 2006;10:287-333. https://doi.org/10.1016/j.ejpain.2005.06.009
4. Broadbent E, Niederhoffer K, Hague T, Corter A, Reynolds L. Headache sufferers' drawings reflect distress, disability and illness perceptions. *J Psychosom Res*. 2009;66:465-470. https://doi.org/10.1016/j.jpsychores.2008.09.006
5. Broadbent E, Petrie KJ, Main J, Weinman J. The Brief Illness Perception Questionnaire. *J Psychosom Res*. 2006;60:631-637. https://doi.org/10.1016/j.jpsychores.2005.10.020
6. Broadbent E, Wilkes C, Koschwanez H, Weinman J, Norton S, Petrie KJ. A systematic review and meta-analysis of the Brief Illness Perception Questionnaire. *Psychol Health*. 2015;30:1361-1385. https://doi.org/10.1080/08870446.2015.1070851
7. Campbell P, Foster NE, Thomas E, Dunn KM. Prognostic indicators of low back pain in primary care: five-year prospective study. *J Pain*. 2013;14:873-883. https://doi.org/10.1016/j.jpain.2013.03.013
8. Chaboyer W, Lee BO, Wallis M, Gillespie B, Jones C. Illness representations predict health-related quality of life 6 months after hospital discharge in individuals with injury: a predictive survey. *J Adv Nurs*. 2010;66:2743-2750. https://doi.org/10.1111/j.1365-2648.2010.05431.x
9. Dalbeth N, Petrie KJ, House M, et al. Illness perceptions in patients with gout and the relationship with progression of musculoskeletal disability. *Arthritis Care Res (Hoboken)*. 2011;63:1605-1612. https://doi.org/10.1002/acr.20570
10. Dempster M, Howell D, McCorry NK. Illness perceptions and coping in physical health conditions: a meta-analysis. *J Psychosom Res*. 2015;79:506-513. https://doi.org/10.1016/j.jpsychores.2015.10.006
11. Executive Agency for Health and Consumers. Musculoskeletal Health in Europe. Truro, UK: European Musculoskeletal Conditions Surveillance and Information Network; 2012.
12. Foster NE, Bishop A, Thomas E, et al. Illness perceptions of low back pain patients in primary care: what are they, do they change and are they associated with outcome? *Pain*. 2008;136:177-187. https://doi.org/10.1016/j.pain.2007.12.007
13. Galli U, Ettlin DA, Palla S, Ehlert U, Gaab J. Do illness perceptions predict pain-related disability and mood in chronic orofacial pain patients? A 6-month follow-up study. *Eur J Pain*. 2010;14:550-558. https://doi.org/10.1016/j.ejpain.2009.08.011
14. Gillanders DT, Ferreira NB, Bose S, Esrich T. The relationship between acceptance, catastrophizing and illness representations in chronic pain. *Eur J Pain*. 2013;17:893-902. https://doi.org/10.1002/j.1532-2149.2012.00248.x
15. Glattacker M, Heyduck K, Meffert C. [Development of a questionnaire to assess treatment representations in rehabilitation patients]. *Rehabilitation (Stuttg)*. 2009;48:345-353. https://doi.org/10.1055/s-0029-1234052
16. Glattacker M, Heyduck K, Meffert C. Illness beliefs and treatment beliefs as predictors of short-term

and medium-term outcome in chronic back pain. *J Rehabil Med*. 2013;45:268-276. https://doi.org/10.2340/16501977-1104

17. Goodman D, Morrissey S, Graham D, Bossingham D. The application of cognitive–behaviour therapy in altering illness representations of systemic lupus erythematosus. *Behav Change*. 2005;22:156-171. https://doi.org/10.1375/bech.2005.22.3.156
18. Groarke AM, Curtis R, Coughlan R, Gsel A. Predictors of physical, psychological and social adjustment in women with rheumatoid arthritis. *Irish J Psychol*. 2005;26:161-176. https://doi.org/10.1080/03033910.2005.10446218
19. Hagger MS, Orbell S. A meta-analytic review of the common-sense model of illness representations. *Psychol Health*. 2003;18:141-184. https://doi.org/10.1080/088704403100081321
20. Hayden JA, Côté P, Steenstra IA, Bombardier C, QUIPS-LBP Working Group. Identifying phases of investigation helps planning, appraising, and applying the results of explanatory prognosis studies. *J Clin Epidemiol*. 2008;61:552-560. https://doi.org/10.1016/j.jclinepi.2007.08.005
21. Hayden JA, Tougas ME, Riley R, Iles R, Pincus T. Individual recovery expectations and prognosis of outcomes in non-specific low back pain: prognostic factor exemplar review. *Cochrane Database Syst Rev*. 2014:CD011284. https://doi.org/10.1002/14651858.CD011284
22. Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. *Ann Intern Med*. 2013;158:280-286. https://doi.org/10.7326/0003-4819-158-4-201302190-00009
23. HealthMeasures. Intro to PROMIS®. Available at: http://www.healthmeasures.net/explore-measurement-systems/promis/intro-to-promis. Accessed December 31, 2017.
24. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327:557-560. https://doi.org/10.1136/bmj.327.7414.557
25. Hill JC, Lewis M, Sim J, Hay EM, Dziedzic K. Predictors of poor outcome in patients with neck pain treated by physical therapy. *Clin J Pain*. 2007;23:683-690. https://doi.org/10.1097/AJP.0b013e3181468e67
26. Hill S, Dziedzic K, Thomas E, Baker SR, Croft P. The illness perceptions associated with health and behavioural outcomes in people with musculoskeletal hand problems: findings from the North Staffordshire Osteoarthritis Project (NorStOP). *Rheumatology (Oxford)*. 2007;46:944-951. https://doi.org/10.1093/rheumatology/kem015
27. Hirsch G. *Somatic and psychological predictors of response to intra-articular corticosteroid injection in knee osteoarthritis* [thesis]. Manchester, UK: University of Manchester; 2015.
28. Hirsch G, Hale E, Kitas G, Klocke R. Psychological factors including illness perceptions may predict response to corticosteroid injection in osteoarthritis [abstract]. *Rheumatology (Oxford)*. 2012;51:iii100. https://doi.org/10.1093/rheumatology/kes109
29. The Illness Perception Questionnaire. Available at: https://www.uib.no/ipq/index.html. Accessed March 9, 2017.
30. International Association for the Study of Pain. Pain terms. Available at: http://www.iasp-pain.org/Education/Content.aspx?ItemNumber=1698#Pain. Accessed August 16, 2018.
31. Israel JI, White KS, Gervino EV. Illness perceptions, negative emotions, and pain in patients with noncardiac chest pain. *J Clin Psychol Med Settings*. 2015;22:77-89. https://doi.org/10.1007/s10880-015-9419-6
32. Järemo P, Arman M, Gerdle B, Larsson B, Gottberg K. Illness beliefs among patients with chronic widespread pain - associations with self-reported health status, anxiety and depressive symptoms and impact of pain. *BMC Psychol*. 2017;5:24. https://doi.org/10.1186/s40359-017-0192-1
33. Leventhal H, Brissette I, Leventhal EA. The common-sense model of self-regulation of health and illness. In: Cameron LD, Leventhal H, eds. *The Self-Regulation of Health and Illness Behaviour*. London, UK: Taylor & Francis/Routledge; 2003:42-65.
34. Leventhal H, Meyer D, Nerenz D. The common sense representation of illness danger. In: Rachman SJ, ed. *Contributions to Medical Psychology: Volume 2*. Oxford, UK: Pergamon; 1980:7-30.
35. Leventhal H, Phillips LA, Burns E. The Common-Sense Model of Self-Regulation (CSM): a dynamic framework for understanding illness self-management. *J Behav Med*. 2016;39:935-946. https://doi.org/10.1007/s10865-016-9782-2
36. Leysen M, Nijs J, Meeus M, et al. Clinimetric properties of Illness Perception Questionnaire revised (IPQ-R) and Brief Illness Perception Questionnaire (Brief IPQ) in patients with musculoskeletal disorders: a systematic review. *Man Ther*. 2015;20:10-17. https://doi.org/10.1016/j.math.2014.05.001
37. Leysen M, Nijs J, Van Wilgen CP, et al. Illness perceptions explain the variance in functional disability, but not habitual physical activity, in patients with chronic low back pain: a cross-sectional study. *Pain Pract*. 2018;18:523-531. https://doi.org/10.1111/papr.12642
38. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol*. 2009;62:e1-e34. https://doi.org/10.1016/j.jclinepi.2009.06.006
39. Linton SJ, Shaw WS. Impact of psychological factors in the experience of pain. *Phys Ther*. 2011;91:700-711. https://doi.org/10.2522/ptj.20100330
40. Macfarlane GJ. Changing patient perceptions of their illness: can they contribute to an improved outcome for episodes of musculoskeletal pain? *Pain*. 2008;136:1-2. https://doi.org/10.1016/j.pain.2008.02.014
41. March L, Smith EU, Hoy DG, et al. Burden of disability due to musculoskeletal (MSK) disorders. *Best Pract Res Clin Rheumatol*. 2014;28:353-366. https://doi.org/10.1016/j.berh.2014.08.002
42. Moss-Morris R, Humphrey K, Johnson MH, Petrie KJ. Patients' perceptions of their pain condition across a multidisciplinary pain management program: do they change and if so does it matter? *Clin J Pain*. 2007;23:558-564. https://doi.org/10.1097/AJP.0b013e318093fcab
43. Moss-Morris R, Weinman J, Petrie K, Horne R, Cameron L, Buick D. The revised Illness Perception Questionnaire (IPQ-R). *Psychol Health*. 2002;17:1-16. https://doi.org/10.1080/08870440290001494
44. Nicklas LB, Dunbar M, Wild M. Adherence to pharmacological treatment of non-malignant chronic pain: the role of illness perceptions and medication beliefs. *Psychol Health*. 2010;25:601-615. https://doi.org/10.1080/08870440902783610
45. Norton S, Hughes LD, Chilcot J, et al. Negative and positive illness representations of rheumatoid arthritis: a latent profile analysis. *J Behav Med*. 2014;37:524-532. https://doi.org/10.1007/s10865-013-9506-9
46. Radat F, Lantéri-Minet M, Nachit-Ouinekh F, et al. The GRIM2005 study of migraine consultation in France. III: psychological features of subjects with migraine. *Cephalalgia*. 2009;29:338-350. https://doi.org/10.1111/j.1468-2982.2008.01718.x
47. Rethlefsen ML, Farrell AM, Osterhaus Trzasko LC, Brigham TJ. Librarian co-authors correlated with higher quality reported search strategies in general internal medicine systematic reviews. *J Clin Epidemiol*. 2015;68:617-626. https://doi.org/10.1016/j.jclinepi.2014.11.025
48. Rezaei F, Doost HT, Molavi H, Abedi MR, Karimifar M. Depression and pain in patients with rheumatoid arthritis: mediating role of illness perception. *Egypt Rheumatol*. 2014;36:57-64. https://doi.org/10.1016/j.ejr.2013.12.007
49. Roios E, Paredes AC, Alves AF, Pereira MG. Cognitive representations in low back pain in patients receiving chiropractic versus physiotherapy treatment. *J Health Psychol*. 2017;22:1012-1024. https://doi.org/10.1177/1359105315621781
50. Scharloo M, Kaptein AA, Weinman J, et al. Illness perceptions, coping and functioning in patients with rheumatoid arthritis, chronic obstructive pulmonary disease and psoriasis. *J Psychosom Res*. 1998;44:573-585. https://doi.org/10.1016/S0022-3999(97)00254-7
51. Scharloo M, Kaptein AA, Weinman JA, Hazes JM, Breedveld FC, Rooijmans HG. Predicting functional status in patients with rheumatoid arthritis. *J Rheumatol*. 1999;26:1686-1693.
52. Shea BJ, Hamel C, Wells GA, et al. AMSTAR is a

reliable and valid measurement tool to assess the methodological quality of systematic reviews. *J Clin Epidemiol*. 2009;62:1013-1020. https://doi.org/10.1016/j.jclinepi.2008.10.009

53. Siemonsma PC, Stuive I, Roorda LD, et al. Cognitive treatment of illness perceptions in patients with chronic low back pain: a randomized controlled trial. *Phys Ther*. 2013;93:435-448. https://doi.org/10.2522/ptj.20110150

54. Sluiter JK, Frings-Dresen MH. Quality of life and illness perception in working and sick-listed chronic RSI patients. *Int Arch Occup Environ Health*. 2008;81:495-501. https://doi.org/10.1007/s00420-007-0222-z

55. Stuifbergen AK, Phillips L, Voelmeck W, Browder R. Illness perceptions and related outcomes among women with fibromyalgia syndrome. *Womens Health Issues*. 2006;16:353-360. https://doi.org/10.1016/j.whi.2006.05.003

56. van Os S, Norton S, Hughes LD, Chilcot J. Illness perceptions account for variation in positive outlook as well as psychological distress in rheumatoid arthritis. *Psychol Health Med*. 2012;17:427-439. https://doi.org/10.1080/13548506.2011.626432

57. van Wilgen CP, van Ittersum MW, Kaptein AA, van Wijhe M. Illness perceptions in patients with fibromyalgia and their relationship to quality of life and catastrophizing. *Arthritis Rheum*. 2008;58:3618-3626. https://doi.org/10.1002/art.23959

58. Vos T, Flaxman AD, Naghavi M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380:2163-2196. https://doi.org/10.1016/S0140-6736(12)61729-2

59. Weinman J, Petrie KJ, Moss-Morris R, Horne R. The Illness Perception Questionnaire: a new method for assessing the cognitive representation of illness. *Psychol Health*. 1996;11:431-445. https://doi.org/10.1080/08870449608400270

60. Woolf AD, Pfleger B. Burden of major musculoskeletal conditions. *Bull World Health Organ*. 2003;81:646-656.

@ MORE INFORMATION
WWW.JOSPT.ORG

APPENDIX A

CROSS-SECTIONAL ASSOCIATIONS BETWEEN ILLNESS PERCEPTION DOMAINS AND MUSCULOSKELETAL PAIN AT BASELINE

Illness/Study/Measure	Consequences	Timeline-Chronic	Timeline-Cyclical	Control Beliefs-Personal	Control Beliefs-Treatment	Identity	Concern	Coherence	Emotional Representations
Rheumatoid arthritis									
Groarke et al[18]									
AIMS*	ß = 0.21	ß = −0.05			r = 0.38† ß = 0.31†	ß = 0.23			
van Os et al[56]									
EQ-5D‡	r = 0.34†	r = 0.09	r = 0.08	r = 0.15§	r = 0.19†	r = 0.32†		r = 0.01	
Norton et al[45]									
EQ-5D‡	r = 0.54	r = 0.15	r = 0.20	r = 0.35	r = 0.35	r = 0.54		r = 0.12	r = 0.29
Rezaei et al[48]									
RAPS‖	r = 0.54†	r = 0.41†		r = −0.34†	r = −0.11	r = 0.34‡	r = 0.25§	r = −0.18	r = 0.39†
Chronic pain									
Nicklas et al[44]									
NRS‡	r = 0.33†	r = 0.23†	r = −0.14§	r = 0.21†	r = 0.24	r = 0.06		r = 0.14§	r = 0.27†
Gillanders et al[14]									
MPQ‡	r = 0.47†	r = 0.34†	r = 0.02	r = 0.07	r = 0.22	r = 0.50†		r = 0.12	r = 0.29†
Järemo et al[32]									
SF-36‡	r = 0.47†	r = 0.13	r = 0.06	r = 0.23†	r = 0.17§	r = 0.41†		r = 0.11	r = 0.24†
Fibromyalgia									
van Wilgen et al[57]									
NRS‡	r = 0.28	r = −0.23	r = 0.23	r = −0.01	r = 0.02	r = 0.10		r = 0.20	r = 0.09
Musculoskeletal hand problems									
Hill et al[26]									
AIMS‡	1.29 (1.25, 1.32)¶	2.51 (2.07, 3.04)¶							
Chronic RSI									
Sluiter and Frings-Dresen[54]									
VAS-intensity‖	r = 0.51	r = 0.18		r = 0.36	r = 0.19	r = 0.64	r = 0.46	r = 0.16	r = 0.35
SF-36‖	r = 0.62	r = 0.15		r = 0.42	r = 0.21	r = 0.67	r = 0.49	r = 0.19	r = 0.41
Headache									
Broadbent et al[4]									
NRS-average‖	r = 0.34†	r = 0.25§		r = 0.30§	r = 0.05	r = 0.32§	r = 0.45†	r = 0.10	r = 0.14
NRS-worst‖	r = 0.37†	r = 0.30§		r = 0.15§	r = −0.06	r = 0.36§	r = 0.27§	r = 0.12	r = 0.12

Abbreviations: AIMS, Arthritis Impact Measurement Scales; EQ-5D, EuroQol-5 dimensions; MPQ, McGill Pain Questionnaire; NRS, numeric rating scale; RAPS, Rheumatoid Arthritis Pain Scale; RSI, repetitive strain injury; SF-36, Medical Outcomes Study 36-Item Short-Form Health Survey; VAS, visual analog scale.

**Correlated with the Illness Perception Questionnaire.*

†P<.01.

‡Correlated with the Illness Perception Questionnaire revised.

§P<.05.

‖Correlated with the Brief Illness Perception Questionnaire.

¶Values are adjusted odds ratio (95% confidence interval).

APPENDIX B

CROSS-SECTIONAL ASSOCIATIONS BETWEEN ILLNESS PERCEPTIONS AND MUSCULOSKELETAL PHYSICAL FUNCTION

Illness/Study/Measure	Consequences	Timeline-Chronic	Timeline-Cyclical	Control Beliefs-Personal	Control Beliefs-Treatment	Identity	Concern	Coherence	Emotional Representations
Rheumatoid arthritis									
Scharloo et al[50]									
SF-20* (T0)					$\beta = 0.25^{\dagger\ddagger}$	$\beta = -0.33^{\dagger\S}$			
HAQ* (T0)					$\beta = -0.24^{\dagger\ddagger}$	$\beta = 0.50^{\dagger\ddagger}$			
Groarke et al[18]									
AIMS* (T0)	$r = 0.35^{\ddagger}$ $\beta = 0.23^{\|}$	$\beta = 0.05^{\|}$			$\beta = -0.20^{\|\P}$	$r = 0.38^{\S}$ $\beta = 0.29^{\|\P}$			
AIMS* (T1)	$r = 0.55^{\ddagger}$ $\beta = 0.14^{\|}$	$r = 0.03$ $\beta = 0.10^{\|}$			$\beta = -0.13^{\|}$	$r = 0.34^{\S}$ $\beta = 0.14^{\|}$			
AIMS* (T2)	$r = 0.43^{\ddagger}$ $\beta = 0.25^{\|}$	$\beta = 0.07^{\|}$			$\beta = -0.06^{\|}$	$\beta = 0.10^{\|}$			
van Os et al[56]									
HAQ# (T0)	$r = 0.55^{\S}$	$r = 0.10$	$r = 0.03$	$r = 0.22^{\S}$	$r = -0.18^{\S}$	$r = 0.34^{\S}$		$r = 0.03$	
Norton et al[45]									
HAQ# (T0)	$r = 0.56$	$r = 0.12$		$r = 0.30$	$r = -0.32$	$r = 0.42$		$r = -0.05$	$r = 0.23$
Chronic pain									
Gillanders et al[14]									
RMDQ# (T0)	$r = 0.60^{\P}$	$r = 0.40^{\P}$	$r = -0.07$	$r = 0.13$	$r = 0.30^{\P}$	$r = 0.34^{\P}$		$r = 0.04$	$r = 0.37^{\P}$
Järemo et al[32]									
SF-36 PF# (T0)	$r = 0.36^{\S}$ $\beta = 0.49^{\dagger\S}$	$r = 0.13$	$r = 0.11$	$r = 0.33^{\S}$ $\beta = 0.66^{\dagger\S}$	$r = 0.22^{\P}$	$r = 0.28^{\P}$ $\beta = 0.83^{\dagger\S}$		$r = 0.07$	$r = 0.04$
Fibromyalgia									
Stuifbergen et al[55]									
SF-36 PF#				$\beta = 0.33^{\|}$					
Musculoskeletal hand problems									
Hill et al[26]									
AIMS# (T0)	1.26 (1.23, 1.29)** 1.26 (1.22, 1.29)†† 1.18 (1.14, 1.23)‡‡					5.34 (4.29, 6.64)** 5.34 (4.19, 6.81)†† 2.32 (1.73, 3.12)‡‡			
Chronic RSI									
Sluiter and Frings-Dresen[54]									
SF-36 PF§§ (T0)	$r = 0.49$	$r = 0.16$		$r = 0.34$	$r = 0.15$	$r = 0.49$	$r = 0.30$	$r = 0.12$	$r = 0.25$
Chronic low back pain									
Roios et al[49]									
CT: ODI# (T0)						$\beta = 0.18^{\dagger}$			
PT: ODI# (T0)				$\beta = -0.10^{\dagger}$		$\beta = 0.18^{\dagger}$			
Leysen et al[37]									
ODI# (T0)	$\beta = 1.25^{\P}$		$\beta = 0.64^{\P}$	$\beta = 0.92^{\P}$					$\beta = 0.64^{\P}$

Abbreviations: AIMS, Arthritis Impact Measurement Scales; CT, chiropractic; GCPS, Graded Chronic Pain Scale; HAQ, Health Assessment Questionnaire; ODI, Oswestry Disability Index; PF, physical functioning; PT, physical therapy; RMDQ, Roland-Morris Disability Questionnaire; RSI, repetitive strain injury; SF-12, Medical Outcomes Study 12-Item Short-Form Health Survey; SF-20, Medical Outcomes Study 20-Item Short-Form Health Survey; SF-36, Medical Outcomes Study 36-Item Short-Form Health Survey; T0, baseline; T1, time interval of less than 6 months; T2, time interval of 6 to 12 months.

Legend continues on page E3.

APPENDIX B

**Correlated with the Illness Perception Questionnaire.*
†Values are unadjusted.
‡P<.001.
§P<.01.
‖Values are adjusted.
¶P<.05.
#Correlated with the Illness Perception Questionnaire revised.
***Values are unadjusted odds ratio (95% confidence interval).*
††Values are odds ratio (95% confidence interval) adjusted for age/sex/perceived diagnoses.
‡‡Values are odds ratio (95% confidence interval) adjusted for all other variables.
§§Correlated with the Brief Illness Perception Questionnaire.

APPENDIX C

LONGITUDINAL ASSOCIATIONS BETWEEN ILLNESS PERCEPTIONS AND MUSCULOSKELETAL PAIN

Illness/Study/Measure	Consequences	Timeline-Chronic	Timeline-Cyclical	Control Beliefs-Personal	Control Beliefs-Treatment	Identity	Concern	Coherence	Emotional Representations
Rheumatoid arthritis									
Scharloo et al[51]									
VAS* (T2)						$\beta = 0.37^{\ddagger}$			
Groarke et al[18]									
AIMS* (T2)	$r = 0.39^{\S}$								
Low back pain									
Glattacker et al[16]									
VAS‖ (T1)	$\beta = 0.24^{\dagger}$				$\beta = 0.17^{\dagger}$				
SF-36 BP‖ (T1)	$\beta = 0.33^{\dagger}$							$\beta = -0.20^{\dagger}$	
VAS‖ (T2)		$\beta = 0.28^{\dagger}$							
SF-36 BP‖ (T2)		$\beta = 0.38^{\dagger}$							
Knee osteoarthritis									
Hirsch[27]									
VAS‖ (3 wk)	0.90 (0.84, 0.97)¶ 0.93 (0.86, 1.01)#				1.25 (1.1, 1.4)¶ 1.40 (1.1, 1.8)#				
VAS‖ (9 wk)	0.82 (0.75, 0.90)¶ 0.83 (0.75, 0.93)#				1.25 (1.08, 1.44)¶ 1.24 (1.05, 1.47)#				0.91 (0.84, 0.98)¶ 0.99 (0.90, 1.09)#
POA subset									
VAS‖ (3 wk)	0.81 (0.73, 0.91)¶ 0.92 (0.82, 1.02)#			1.14 (1.01, 1.28)¶	1.30 (1.09, 1.50)¶ 1.20 (1.0, 1.45)#				
VAS‖ (9 wk)	0.81 (0.73, 0.90)¶ 0.82 (0.75, 0.93)#			1.20 (1.05, 1.40)¶ 1.12 (0.94, 1.33)#	1.30 (1.09, 1.50)¶ 1.27 (1.01, 1.59)#				0.90 (0.83, 0.99)¶ 0.96 (0.86, 1.07)#

Abbreviations: AIMS, Arthritis Impact Measurement Scales; BP, bodily pain; POA, primary osteoarthritis; SF-36, Medical Outcomes Study 36-Item Short-Form Health Survey; T1, time interval of less than 6 months; T2, time interval of 6 to 12 months; VAS, visual analog scale.

**Correlated with the Illness Perception Questionnaire.*

†Values are adjusted.

‡P<.01.

§P<.05.

‖Correlated with the Illness Perception Questionnaire revised.

¶Values are unadjusted odds ratio (95% confidence interval).

#Values are adjusted odds ratio (95% confidence interval).

APPENDIX D

LONGITUDINAL ASSOCIATIONS BETWEEN ILLNESS PERCEPTIONS AND MUSCULOSKELETAL PHYSICAL FUNCTION

Illness/Study/ Measure	Consequences	Timeline-Chronic	Timeline-Cyclical	Control Beliefs-Personal	Control Beliefs-Treatment	Identity	Concern	Coherence	Emotional Representations
Rheumatoid arthritis									
Groarke et al[18]									
AIMS* (T1)	$r = 0.35^{\dagger}$								
Low back pain									
Foster et al[12]									
RMDQ‡ (T1)	1.61 (1.2, 2.1)§ 1.38 (1.0, 1.9)‖	2.00 (1.5, 2.6)§ 1.83 (1.4, 2.5)‖	0.97 (0.7, 1.3)§ 1.09 (0.8, 1.4)‖	1.59 (1.2, 2.1)§ 1.49 (1.1, 2.0)‖	1.52 (1.2, 2.0)§ 1.40 (1.1, 1.8)‖	1.19 (0.9, 1.5)§ 1.04 (0.8, 1.4)‖		1.11 (0.8, 1.5)§ 1.08 (0.8, 1.5)‖	1.34 (1.0, 1.8)§ 1.23 (0.9, 1.7)‖
Campbell et al[7]									
CPGS‡ (T1)	1.09 (1.1, 1.1)§ 1.09 (1.1, 1.1)‖	1.07 (1.0, 1.1)§ 1.09 (1.1, 1.1)‖ 1.04 (1.0, 1.1)¶#	1.00 (0.96, 1.0)§ 1.01 (0.97, 1.1)‖	0.93 (0.90, 0.97)§ 0.91 (0.88, 0.95)‖	0.93 (0.90, 0.97)§ 0.90 (0.86, 0.94)‖	1.15 (1.1, 1.2)§ 1.16 (1.1, 1.2)‖		1.02 (0.99, 1.0)§ 1.03 (1.0, 1.1)‖	1.07 (1.1, 1.1)§ 1.06 (1.0, 1.1)‖
CPGS‡ (T3)	1.01 (0.97, 1.1)§ 1.03 (0.97, 1.1)‖	1.04 (1.0, 1.1)§ 1.06 (1.0, 1.1)‖ 1.06 (1.0, 1.1)†¶		0.97 (0.93, 1.0)§ 0.98 (0.97, 1.0)‖	1.03 (0.97, 1.1)§ 1.01 (0.95, 1.1)‖	1.02 (0.95, 1.1)§ 0.98 (0.90, 1.1)‖		1.01 (0.98, 1.1)‖	1.07 (0.97, 1.1)§ 0.95 (0.91, 0.99)‖ 0.98 (0.94, 1.0)¶
Glattacker et al[16]									
SF-36 PF‡ (T1)								$\beta = 0.12^{\#**}$	
SF-36 SF‡ (T1)	$\beta = 0.26^{\#**}$							$\beta = 0.17^{\#**}$	
ODI‡ (T1)								$\beta = 0.22^{\#**}$	
SF-36 PF‡ (T2)		$\beta = 0.20^{\#**}$							
SF-36 SF‡ (T2)					$\beta = 0.26^{\#**}$				
ODI‡ (T2)						$\beta = 0.20^{\#**}$			
Bishop et al[2]									
RMDQ††: BPE (T3)	$\beta = 1.04^{\dagger**}$	$\beta = -0.03^{**}$		$\beta = 0.24^{\#**}$	$\beta = -0.02^{**}$	$\beta = 0.39^{\#**}$	$\beta = 0.13^{**}$	$\beta = -0.12^{**}$	$\beta = 0.16^{**}$
RMDQ††: WPE (T3)	$\beta = 0.35^{\dagger**}$	$\beta = -0.02^{**}$		$\beta = 0.24^{\dagger**}$	$\beta = 0.13^{**}$	$\beta = 0.34^{\dagger**}$	$\beta = 0.28^{\dagger**}$	$\beta = 0.13^{\#**}$	$\beta = 0.26^{\dagger**}$
Orofacial pain									
Galli et al[13]									
GCPS‡ (3 mo)	$r = 0.30^{\#}$	$r = -0.04$	$r = 0.05$	$r = 0.04$				$r = 0.004$	$r = -0.01$
GCPS‡ (6 mo)	$r = 0.16$	$r = -0.07$	$r = 0.12$	$r = 0.15$				$r = -0.04$	$r = -0.11$
Chronic pain									
Moss-Morris et al[42]									
SF-36 PF‡ (T1)	$\beta = 0.29^{\dagger**}$	$\beta = 0.01^{**}$		$\beta = -0.05^{**}$				$\beta = 0.07^{**}$	$\beta = 0.07^{**}$
Systemic lupus erythematosus									
Goodman et al[17]									
SF-36‡ (T1)	$r = 0.26$	$r = 0.29$	$r = 0.39^{\#}$	$r = 0.16$	$r = 0.16$	$r = 0.58^{\ddagger}$		$r = 0.28$	$r = 0.20$
Injury									
Chaboyer et al[8]									
SF-36 PF‡ (T1)		$\beta = -11.19$ $\beta = -0.51$				$\beta = -3.30$ $\beta = -0.27$			

Table continues on page E6.

APPENDIX D

Illness/Study/ Measure	Consequences	Timeline-Chronic	Timeline-Cyclical	Control Beliefs-Personal	Control Beliefs-Treatment	Identity	Concern	Coherence	Emotional Representations
Gout									
Dalbeth et al[9]									
HAQ[††] (T1)	ß = 0.29**			ß = 0.23**	ß = 0.24**	ß = 0.22**			

Abbreviations: AIMS, Arthritis Impact Measurement Scales; BPE, between-persons effect; CPGS, Chronic Pain Grade Scale; GCPS, Graded Chronic Pain Scale; HAQ, Health Assessment Questionnaire; ODI, Oswestry Disability Index; PF, physical functioning; RMDQ, Roland-Morris Disability Questionnaire; SF, social functioning; SF-36, Medical Outcomes Study 36-Item Short-Form Health Survey; T1, time interval of less than 6 months; T2, time interval of 6 to 12 months; T3, time interval of greater than 12 months; WPE, within-persons effect.

**Correlated with the Illness Perception Questionnaire.*

[†]*P<.01.*

[‡]*Correlated with the Illness Perception Questionnaire revised.*

[§]*Values are unadjusted relative risk (95% confidence interval).*

[‖]*Values are adjusted relative risk (95% confidence interval).*

[¶]*Values are final-model relative risk (95% confidence interval).*

[#]*P<.05.*

***Values are adjusted.*

[††]*Correlated with the Brief Illness Perception Questionnaire.*

[‡‡]*P<.001.*

ELLIOT M. GREENBERG, PT, PhD[1] • **ERIC T. GREENBERG,** PT[2] • **JEFFREY ALBAUGH,** PT[1]
EILEEN STOREY, BS[3] • **THEODORE J. GANLEY,** MD[1,3,4]

Rehabilitation Practice Patterns Following Anterior Cruciate Ligament Reconstruction: A Survey of Physical Therapists

BACKGROUND: Recovery from anterior cruciate ligament reconstruction (ACLR) requires an intensive course of postoperative rehabilitation. Although guidelines outlining evidence-based rehabilitation recommendations have been published, actual practice patterns of physical therapists are unknown.

OBJECTIVES: To analyze the current landscape of clinical practice as it pertains to rehabilitation progression and the use of time and objective criteria in rehabilitation following ACLR.

METHODS: In this cross-sectional study, an online survey was distributed to members of the Academy of Orthopaedic Physical Therapy, the American Academy of Sports Physical Therapy, and the Private Practice Section of the American Physical Therapy Association between January and March 2017.

RESULTS: The study analyzed a sample of 1074 responses. Supervised physical therapy was reported to last 5 months or less by 56% of survey respondents. The most frequent time frames for activity progression were 3 to 4 months (58%) for jogging, 4 to 5 months (50%) for modified sports activity, and 9 to 12 months (40%) for unrestricted sports participation. More than 80% of respondents reported using strength and functional measures during rehabilitation. Of those physical therapists who assessed strength, 56% used manual muscle testing as their only means of strength testing. Single-limb hop testing (89%) was the most frequently reported measure used to allow patients to begin modified sports activity following ACLR. Performance criteria for strength and functional tests varied significantly across all phases of rehabilitation. The 45% of respondents who reported using patient-reported outcome measures indicated that just under 10% of those measures involved fear or athletic confidence scales.

CONCLUSION: Considerable variation in practice exists among American Physical Therapy Association members regarding rehabilitation following ACLR. This variability in practice may contribute to suboptimal outcomes and confusion among practitioners and patients. *J Orthop Sports Phys Ther 2018;48(10):801-811. Epub 22 May 2018. doi:10.2519/jospt.2018.8264*

KEY WORDS: *ACL, anterior cruciate ligament, physical therapy, physical therapy survey, postoperative rehabilitation*

Anterior cruciate ligament (ACL) injuries commonly occur during sports requiring jumping, cutting, and pivoting. Although select individuals may attempt conservative management,[25] the vast majority undergo reconstructive surgery, with some 300 000 anterior cruciate ligament reconstructions (ACLRs) performed annually in the United States alone.[12] Despite continued advancements in surgical techniques and rehabilitation, outcomes following ACLR may be less than desirable, with only 33% of athletes returning to sports within 1 year after surgery[4] and 37% never returning to their prior levels of sports participation.[5] Additionally, and perhaps more alarmingly, up to 30% of individuals may incur a second ACL injury,[23] resulting in higher health care costs and increased disability.

Postoperative rehabilitation can play a vital role in successful recovery following ACLR by optimizing function and reducing the risk of a second ACL injury.[21,38] Historically, rehabilitation progression following ACLR relied heavily on time-based standards, respecting the processes of graft maturation and physiological healing.[17] However, rehabilitation recommendations have evolved over time, and most contemporary protocols recommend a more comprehensive decision-making framework, using a synthesis of time and objective functional performance criteria to guide postoperative progression.[1,7,26,27] Nonetheless, a recent systematic review found that more than 70% of published studies excluded functional measures in return-to-play decision making, revealing a discrepancy between current recommendations for best practice and published literature.[6]

Further, literature analysis reveals significant variation in published rehabilitation protocols and inconsistent recommendations of specific performance measures or criteria for decision making

[1]Sports Medicine and Performance Center at The Children's Hospital of Philadelphia, Philadelphia, PA. [2]Department of Physical Therapy, New York Institute of Technology, Old Westbury, NY. [3]Division of Orthopaedic Surgery, The Children's Hospital of Philadelphia, Philadelphia, PA. [4]Department of Orthopaedic Surgery, University of Pennsylvania, Philadelphia, PA. The Institutional Review Board of The Children's Hospital of Philadelphia approved this study (protocol 16-013163). The authors certify that they have no affiliations with or financial involvement in any organization or entity with a direct financial interest in the subject matter or materials discussed in the article. Address correspondence to Dr Elliot M. Greenberg, Sports Medicine and Performance Center at The Children's Hospital of Philadelphia, Pediatric and Adolescent Specialty Care Center Bucks County, 500 West Butler Avenue, Chalfont, PA 18914. E-mail: greenberge@email.chop.edu

regarding activity progression throughout all phases of rehabilitation.[32,43] This discordance contributes to a complicated practice environment, which may lead to confusion or inconsistent clinical practice patterns among physical therapists treating patients after ACLR.

The purpose of this study was to understand the current landscape of clinical practice among members of the American Physical Therapy Association (APTA), as it pertains to rehabilitation decision making and the use of objective tests to guide activity progression following ACLR. The findings are analyzed within the context of current literature to determine whether clinical practice reflects up-to-date recommendations and scientific evidence. Additionally, the relationship of clinical practice patterns to clinician characteristics was analyzed to understand some of the underlying factors that may be related to individual clinical decision making.

METHODS

Survey Development

A team of clinicians (3 physical therapists and 1 orthopaedic surgeon) highly experienced in managing patients following ACLR collaborated to develop an electronic survey using the Research Electronic Data Capture (REDCap) tool (Vanderbilt University, Nashville, TN) hosted at the Children's Hospital of Philadelphia.[20] The initial phase of survey development consisted of identifying key rehabilitation transitional phases and writing preliminary questions based on previously published reports and clinical expertise.[32-34] Due to differences that may exist in rehabilitation and activity progression based on patient characteristics such as age, activity level, surgery type, graft type, and concomitant injuries, the researchers determined that the questionnaire should be grounded in a standardized case vignette that represent typical patients encountered in a sports or orthopaedic setting. Survey participants were asked to answer the survey items based on their typical treatment in the following case: "Your patient is a 17-year-old female soccer player who underwent ACL reconstruction using a hamstring autograft. There were no concomitant injuries and she is having an uncomplicated postoperative recovery. Her goal is to return to soccer competition at the collegiate level upon full recovery."

Each survey participant was instructed to answer all questions that related to his or her clinical practice of treating patients after ACLR. To be time efficient, the electronic survey incorporated the use of branching logic, which prompted specific follow-up questions only if pertinent responses were selected in previous questions. Thus, the total number of questions answered by each participant varied based on respondents' individual practice patterns.

The survey consisted of 6 sections: (1) clinician demographics and clinical practice information, (2) clinical decisions related to jogging, (3) clinical decisions related to modified sports activity (eg, agilities, sport-specific drills/skills), (4) return to unrestricted sports, (5) use of injury prevention programs, and (6) use of functional bracing upon return to sport.

The initial development team reviewed and tested the survey among themselves for format, inclusivity of content, clarity, and survey functionality. After all initial revisions were made, the survey was pilot tested among a group of 5 physical therapists and 3 orthopaedic surgeons. All suggestions were considered, and modifications to the survey were made after consultation among all authors. During pilot testing, the survey took approximately 4 to 7 minutes to complete. The complete survey instrument is available in the **APPENDIX** (available at www.jospt.org).

Survey Distribution

Physical therapists were recruited through e-mail invitations sent to members of the APTA Academy of Orthopaedic Physical Therapy and American Academy of Sports Physical Therapy. In addition, participation was solicited from members of the APTA Private Practice Section via an embedded link in that section's electronic newsletter. These groups were selected based on the likelihood that their members would treat the intended patient population. A single reminder e-mail was sent near the halfway point in the survey time frame to American Academy of Sports Physical Therapy members only. All other possible participants only received a single invitation at the study's outset.

The invitation provided a brief study description and encouraged physical therapists who actively treat patients after ACLR to participate. Interested participants clicked the electronic link connecting them to a more detailed study description, which included eligibility criteria. Access to the survey was granted after selecting "yes" to the question indicating their informed consent to participate. No identifying information was collected on any of the participants, thus participation was completely anonymous.

Survey responses were collected over 2 months between January and March 2017. Before it began, this study received approval from the Institutional Review Board of The Children's Hospital of Philadelphia (protocol 16-013163). The Checklist for Reporting Results of Internet E-Surveys was used to ensure the quality of reporting the findings of this study.[16]

Data Analysis

Data were analyzed using SPSS Version 24.0 (IBM Corporation, Armonk, NY). The primary analysis involved the use of descriptive statistics to summarize the distribution, frequency, and dispersion of respondents' responses. A secondary analysis using the chi-square statistic was conducted to determine whether relationships existed between clinician characteristics that indicated advanced clinical proficiency (years of experience, volume of patients treated following ACLR, and board-certified specialist certification) and rehabilitation progression after ACLR. These groups were

operationally defined as follows: (1) experienced versus less experienced practitioners were those with either 16 or more years or those with 0 to 4 years of clinical experience, respectively; (2) high- versus low-volume practitioners were those treating more than 10 or those treating 5 or fewer patients following ACLR per year; and (3) board-certified versus non–board-certified practitioners were divided between those who possessed orthopaedic clinical specialist (OCS) or sports clinical specialist (SCS) certifications and those who did not.

For dichotomous analysis, significance values were set at $P<.05$. For analysis within which multiple comparisons were made, appropriate Bonferroni correction was used to determine statistical significance.

RESULTS

Survey Response

A total of 1084 survey responses were recorded; 10 responses were excluded from data analysis (7 failed to consent and 3 were not licensed physical therapists). Therefore, a total of 1074 responses were included in the analyses. Of these, 593 (55.2%) respondents accessed the survey from the Academy of Orthopaedic Physical Therapy e-mail invitation, while 403 (37.5%) accessed the survey from the American Academy of Sports Physical Therapy invitation. The remaining 78 (7.3%) respondents accessed the survey through other means, including the Private Practice Section newsletter or word of mouth.

Respondents' Profile

Demographic and professional characteristics of respondents are presented in **TABLE 1**. All states but 1 (Rhode Island) were represented in the sample. Respondents were well distributed across various years of clinical practice and volume of patients post ACLR treated annually. Most respondents treated patients primarily in a private practice (42.8%) or hospital-based outpatient facility (35.8%). Though nearly all respondents were members of the APTA (92.5%), half held either OCS or SCS board certifications (52.5%).

Decision Making Regarding Activity Progression

A large proportion of this sample (80.1%) indicated that progression of activity after ACLR was largely a collaborative process, with shared decision making between the orthopaedic surgeon and the physical therapist.

Time Criterion

"For the patient described, how long would your typical course of rehabilitation be?

TABLE 1 Demographics of the Study Respondents

Characteristic	Value, n (%)
Time in practice as a physical therapist, y	
0-4	246 (22.9)
5-10	218 (20.3)
11-15	157 (14.6)
≥16	451 (42.0)
Primary practice setting*	
Private practice	460 (42.8)
Hospital-based outpatient	384 (35.8)
Corporate-owned outpatient practice	118 (11.0)
Academic/collegiate facility	73 (6.8)
Other	37 (3.4)
Region of practice†	
South Atlantic (DE, FL, GA, MD, NC, SC, VA, WV)	200 (18.6)
Middle Atlantic (NJ, NY, PA)	157 (14.6)
East North Central (IL, IN, MI, OH, WI)	162 (15.1)
West North Central (IA, KS, MN, MO, NE, ND, SD)	107 (10.0)
East South Central (AL, KY, MS, TN)	37 (3.5)
West South Central (AR, LA, OK, TX)	85 (7.9)
New England (CT, ME, MA, NH, RI, VT)	144 (13.4)
Pacific (AK, CA, HI, OR, WA)	149 (13.9)
Mountain (AZ, CO, ID, MT, NV, NM, UT, WY)	95 (8.9)
ACLs treated per year, n	
0	11 (1.0)
1-5 (low volume)	347 (32.3)
6-10 (medium volume)	309 (28.8)
>10 (high volume)	407 (37.9)
ABPTS certification (OCS or SCS)‡	
Yes	564 (52.5)
No	508 (47.3)
Current APTA member§	
Yes	993 (92.5)
No	69 (6.4)

Abbreviations: ABPTS, American Board of Physical Therapy Specialties; ACL, anterior cruciate ligament; APTA, American Physical Therapy Association; OCS, orthopaedic clinical specialist; SCS, sports clinical specialist.
**Two (0.2%) respondents did not identify primary practice setting.*
†Twenty-eight (2.6%) respondents did not identify state of practice.
‡Two (0.2%) respondents did not identify whether they were ABPTS certified in orthopaedics or sports.
§Twelve (1.1%) respondents did not identify whether they were a current APTA member.

(ie, How long would you treat them within an office setting?)"

The length of supervised rehabilitation spanned from 1 to 3 months (15.6%) to 12 months (11.2%), although the majority reported 4 to 5 months (40.6%) and 6 to 8 months (32.1%). Significant associations between length of rehabilitation and clinician characteristics are detailed in **TABLE 2**. Clinicians with less clinical experience, higher volumes of patients post ACLR, and OCS or SCS certification indicated a longer overall duration of clinical care.

"I would typically allow the athlete in this example to begin [jogging, modified sports activity, or unrestricted return to sports] at _____ months postsurgery."

Response frequencies pertaining to transitional time points of jogging, modified sports activity, and unrestricted return to sports are presented in **TABLE 3**. Nearly all physical therapists indicated that they would initiate jogging between 2 and 5 months post surgery, with the majority (58%) reporting the introduction of jogging at 3 to 4 months. Modified sports activity (eg, agility and coordination drills) was most often initiated at 4 to 5 months (50.1%) and 6 to 7 months (31.4%). Progression to unrestricted sports was reported to occur most frequently between 9 and 12 months post surgery (39.8%); however, responses within this phase were more widely distributed compared to the other 2 transition points.

TABLE 2 Associations Between Length of Treatment Course and Physical Therapist Characteristics

	<6 mo of Treatment, n	>6 mo of Treatment, n	χ^2	*P* Value	Effect Size
Experience, y					
0-4	102	143	34.253	<.001*	0.222
≥16	291	159			
ACLs treated per year, n					
1-5	212	134	15.200	<.001*	0.142
>10	190	214			
OCS or SCS certification					
No	309	198	8.909	.003*	0.091
Yes	291	270			

Abbreviations: ACL, anterior cruciate ligament; OCS, orthopaedic clinical specialist; SCS, sports clinical specialist.
**Significant (P<.05).*

TABLE 3 Time-Based Responses to Initiate Functional Milestones Following ACLR*

Time, mo	Jogging	Modified Sports Activity	Unrestricted Return to Sport
<3		66 (6.2)	
2-3	200 (18.6)		
3-4	622 (57.9)		
4-5	204 (19.0)	538 (50.1)	22 (2.1)
>6	38 (3.5)		
6-7		337 (31.4)	251 (23.4)
7-8			157 (14.7)
8-9		46 (4.3)	
9-12		27 (2.5)	426 (39.8)
>12		3 (0.3)	57 (5.3)

Abbreviation: ACLR, anterior cruciate ligament reconstruction.
**Values are n (%). Those who answered "typically do not treat patients at this functional milestone": jogging, 9 (0.8%); modified sports activity, 47 (4.4%); and unrestricted return to sport, 158 (14.7%).*

Criterion-Based Measures

Progression to Jogging, Modified Sports Activity, and Unrestricted Return to Sports *"Are there specific physical tests, examination findings, or criteria that you utilize in order to assist in the decision to progress to [jogging, modified sports activity, or unrestricted return to sports]?"* The most often reported criteria to initiate jogging and modified sports activity were knee strength (91.6% and 80%, respectively), functional/balance tests (86.9% and 82.5%, respectively), knee range of motion (80.3% and 61.9%, respectively), and degree of knee effusion (70.6% and 59.6%, respectively). In regard to unrestricted return to sports, a majority of respondents (54.7%) indicated that they did not require any additional testing to progress patients following ACLR to this stage of rehabilitation (**FIGURE 1**).

Knee Strength Knee strength was often reported as key to determining readiness for activity progression throughout all phases of rehabilitation. If a respondent included knee strength as part of the assessment, then additional data were gathered about the method of testing, which are represented in **FIGURE 2**. Of those who relied on knee strength to initiate jogging and modified sports activity, manual muscle testing (MMT) was the most common assessment method used, by 80.6% (n = 793/984) and 74.3% (n = 638/859) of respondents, respectively. Further, 54.9% (n = 472/859) used at least 1 method of objective measure (isometric dynamometry or handheld dynamometry [HHD], isokinetics, or repetition-maximum testing). Of those who relied on MMT

to begin modified sports activity, 56.1% (n = 358/638) reported using MMT as their only means of strength assessment, while the remaining 43.9% (n = 280/638) used it in conjunction with more objective measures.

High-volume practitioners and certified specialists were more likely to use objective strength measures, while low-volume practitioners and non–certified specialists were more likely to rely solely on MMT (χ^2 = 22.088, P<.001 and χ^2 = 7.804, P = .005). There was no association between the amount of clinician experience and type of strength measure (χ^2 = 0.264, P = .608).

Quadriceps limb symmetry index (LSI) standards were recorded from those respondents who used isometric dynamometry or HHD, isokinetics, or repetition-maximum testing for both jogging and modified sports activity phases of rehabilitation (**FIGURE 3**). Although there was significant variation in responses, a quadriceps LSI of greater than 80% was the most commonly cited criterion to initiate jogging, regardless of testing mode. However, when progressing to modified sports activity, those who performed repetition-maximum, isometric dynamometry, or HHD testing more often required a more stringent criterion (greater than 90% LSI) than those who employed isokinetic assessments (greater than 85% LSI).

Functional Testing The lateral step-down test (70%), Y Balance Test (YBT) or Star Excursion Balance Test (SEBT) (55.6%), and Functional Movement Screen (31.4%) were the most frequently cited functional tests used to initiate jogging. To determine readiness for modified sports activity, single-limb hop tests were used by the majority of respondents (89.2%), followed by the YBT or SEBT (48.8%) and the drop vertical jump (39.5%) (**TABLE 4**). Of those who used single-limb hop testing, 79.4% reported using at least 2 types of hop tests, with the single hop (89.4%) and triple hop (80%) used most frequently (**TABLE 5**). Approximately 60% of the sample required an LSI of 90% or greater for progression to modified sports activity, while the remainder of respondents reported that they considered an LSI of 75% to 85% acceptable.

Though one of the more common responses, cutoff criteria varied for the YBT, SEBT, and Functional Movement Screen. Of those respondents who used the YBT or SEBT to initiate jogging, 42% said they require an anterior reach side-to-side difference of less than 4 cm, while 72.6% require a between-limb composite reach score of greater than 90%. Similarly, of those who reported using the Functional Movement Screen, 51.9% consider the overall score, 61.1% stress the performance on isolated movements, and 82.3% rely on side-to-side movement symmetry during unilateral movements.

High-volume practitioners were more likely to use the YBT or SEBT (χ^2 = 10.895, P = .001) and drop vertical jump (χ^2 = 14.576, P<.001). Less experienced clinicians were also more likely to employ the YBT or SEBT (χ^2 = 17.46, P<.001) (**FIGURE 4**).

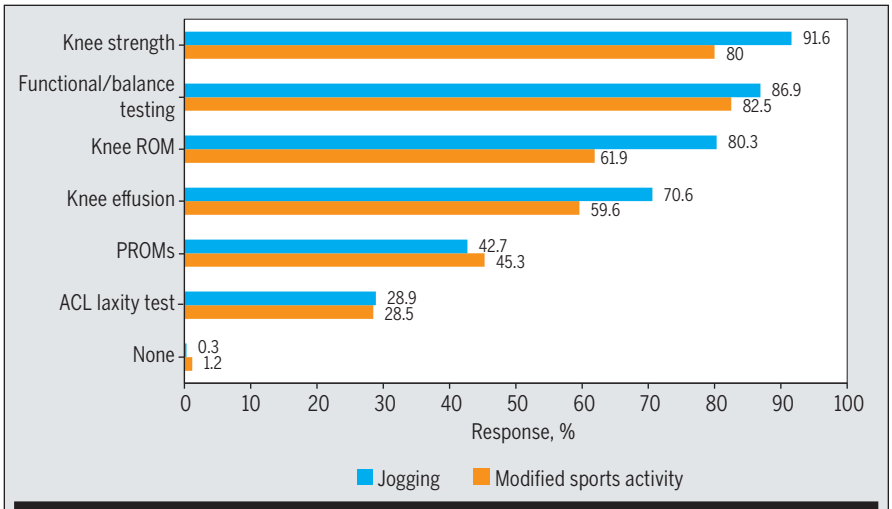


FIGURE 1. Criterion-based responses for initiating jogging and modified sports activity following ACL reconstruction (n = 1074). Abbreviations: ACL, anterior cruciate ligament; PROM, patient-reported outcome measure; ROM, range of motion.

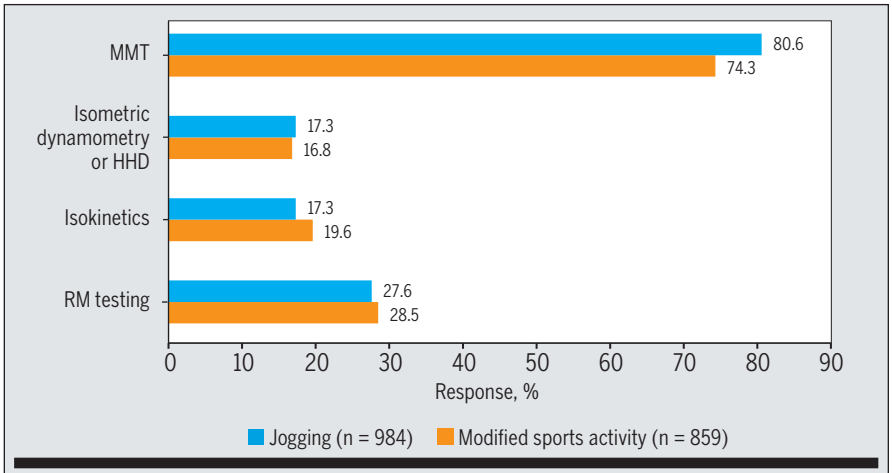


FIGURE 2. Mode of strength testing from those who responded that they use strength testing in decision making to initiate running and modified sports activity. Abbreviations: HHD, handheld dynamometry; MMT, manual muscle testing; RM, repetition maximum.

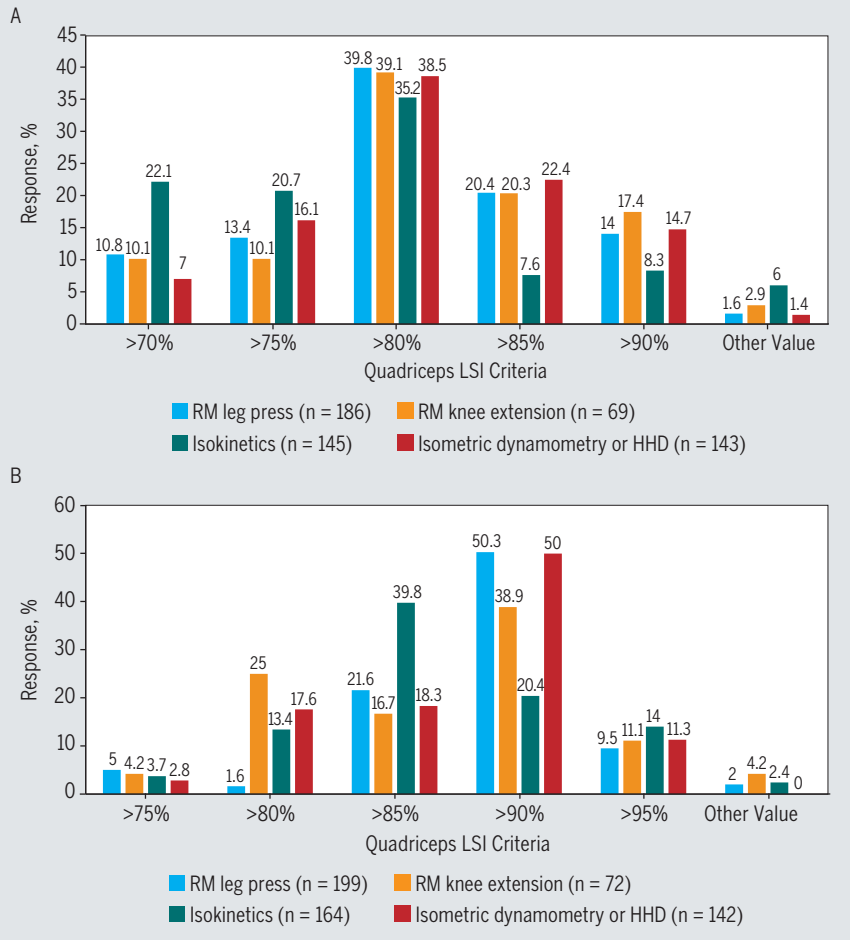


FIGURE 3. Requirements of quadriceps LSIs for various modes of strength testing to initiate (A) jogging and (B) modified sports activity. Abbreviations: HHD, handheld dynamometry; LSI, limb symmetry index; RM, repetition maximum.

TABLE 4 Functional Test Selection Used to Initiate Jogging and Modified Sports Activity*

Test	Jogging (n = 933)	Modified Sports Activity (n = 886)
Lateral step-down test	653 (70)	...
YBT or SEBT	519 (55.6)	432 (48.8)
FMS	293 (31.4)	276 (31.2)
Straight leg raise	262 (28.1)	...
Balance assessment tool	130 (13.9)	107 (12.1)
Single-limb hop test	...	790 (89.2)
Drop vertical jump	...	350 (39.5)
LESS	...	94 (10.6)
Vail Sport Test	...	66 (7.4)
Other test not listed	233 (25)	112 (12.6)

Abbreviations: FMS, Functional Movement Screen; LESS, Landing Error Scoring System; SEBT, Star Excursion Balance Test; YBT, Y Balance Test.
**Values are n (%).*

Patient-Reported Outcome Measures

Patient-reported outcome measures were used by 45.3% of physical therapists to progress patients to modified sports activity. Among clinicians using these tools, the Lower Extremity Functional Scale was most widely reported (39.2%), while scales related to fear or athletic confidence were less commonly cited (9.7%) (**FIGURE 5**). There were no significant associations between clinician characteristics and the use of patient-reported outcome measures.

Postrehabilitation Factors: Injury Prevention Programs and Functional Bracing

Although most physical therapists (74.9%) recommend injury prevention programs after ACLR, high-volume practitioners (χ^2 = 20.266, P<.001) and certified specialists (χ^2 = 4.007, P = .045) were more likely to incorporate them into their plan of care. There was no clear consensus around program preference; most used the Prevent Injury and Enhance Performance program (31.3%), Fédération Internationale de Football Association 11+ (21.4%), or an individually adapted program (29.7%). Overall, 41.1% of physical therapists favored the use of functional bracing on return to sports, with certified specialists being less likely to recommend it (χ^2 = 4.767, P = .029).

DISCUSSION

The results of this survey provide a detailed description of physical therapy practice patterns for postoperative care of young athletes following ACLR. One of the most noticeable findings was the degree of variability in clinical testing and decision making, particularly within the later phases of rehabilitation, during the transition back to sports activity. Although surprising, this result may reflect the lack of well-defined clinical evidence to guide practice, as currently there is no consensus about the ideal postoperative rehabilitation program.[15,43]

The incorporation of time-based criteria into ACLR rehabilitation protocols

has been advocated, based on biological features of graft strength, stiffness, and strength of fixation.[44] The results of this survey show that agreement regarding time-based criteria decreased as the rehabilitation course progressed. While a majority (defined as greater than 50% of the sample) of physical therapists agreed that jogging and return to modified sports activity should occur from 3 to 4 and 4 to 5 months post surgery, respectively, agreement about progression to unrestricted return to sports was less apparent. These results likely reflect the complexity of decision making in the later phases of rehabilitation, which may include type of sport and patient-specific factors. However, this finding may also reflect the variability in guidelines, with published reports demonstrating similar variation, calling for this transition as early as 4 or greater than 9 months post surgery.[1,24,44]

Interestingly, 88.3% of the sample indicated a typical duration of supervised rehabilitation of 8 months or less, while 45.1% indicated that they do not recommend unrestricted return to sports until 9 to 12 or more months after surgery. These findings imply that there may be a long gap between the discontinuation of supervised rehabilitation and return to activity. While other rehabilitation professionals, such as athletic trainers, may be able to advise athletes during this period, a recent survey demonstrated that only 37% of public secondary schools provide full-time athletic training services.[37] As a result, most patients would be responsible for self-managing this advanced phase of recovery without any professional supervision.

The subgroup analysis revealed that less experienced clinicians, high-volume practitioners, and certified specialists advocated for a longer duration of supervised care. One possible explanation for this finding is that these groups are more cognizant of contemporary ACLR rehabilitation models, which call for more prolonged time frames prior to returning to sports,[8,15,24] resulting in the desire for a longer duration of supervised rehabilitation. Alternatively, this finding may also be explained by more experienced clinicians having established a network of community-based alternative practitioners (eg, athletic trainers, personal trainers, coaches) to entrust supervision of late-phase rehabilitation for these athletes. Future research is necessary to explore these observations and understand the driving force behind time-based decisions among these groups of practitioners.

More than 80% of respondents agreed on the importance of a multidimensional approach to informed decision making after ACLR, using physical measures such as strength, lower extremity function, and dynamic stability. This response aligns with recommendations in published literature.[3,15,43]

Despite significant agreement on these principles, physical therapists varied in their mode and interpretation of these measures. For example, while

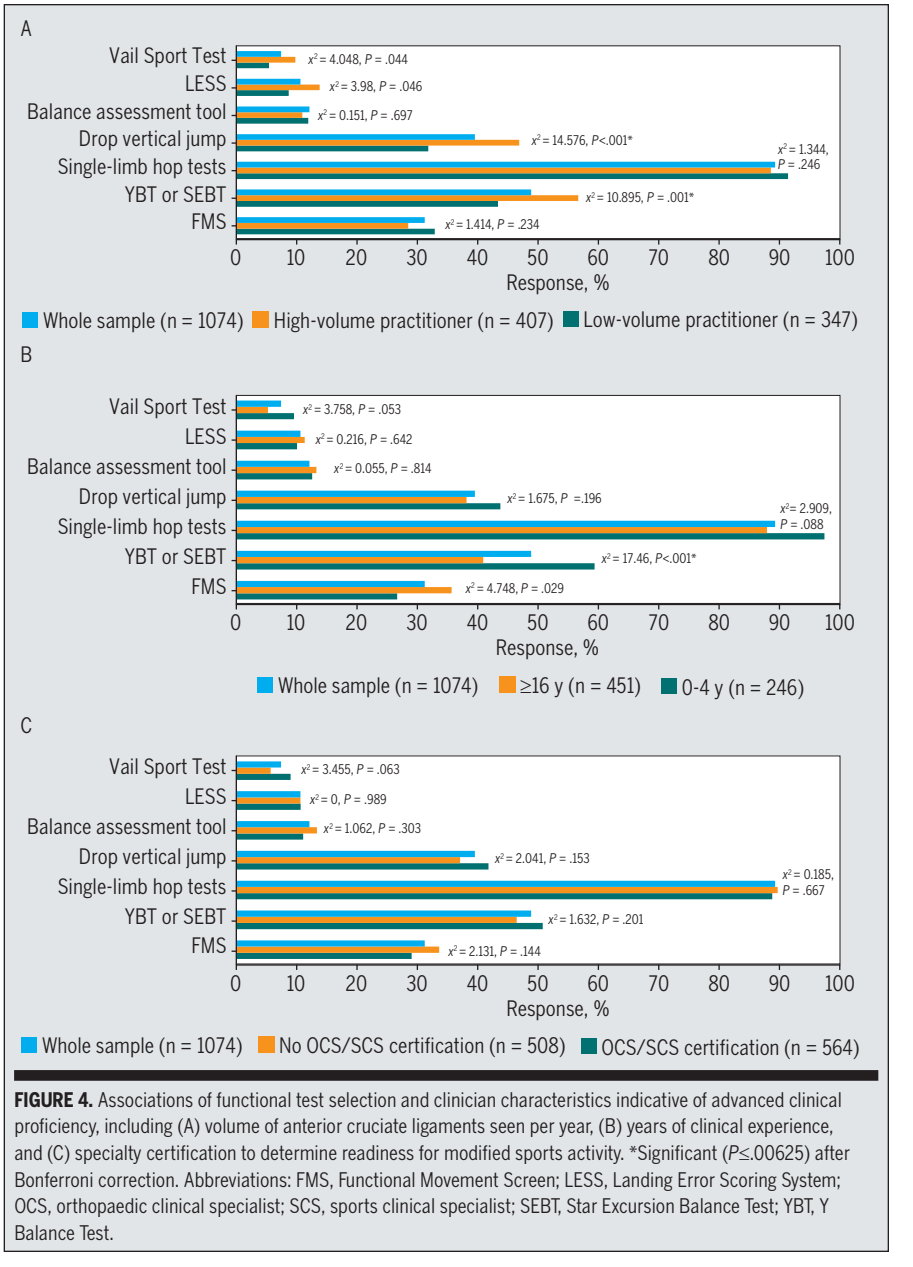


FIGURE 4. Associations of functional test selection and clinician characteristics indicative of advanced clinical proficiency, including (A) volume of anterior cruciate ligaments seen per year, (B) years of clinical experience, and (C) specialty certification to determine readiness for modified sports activity. *Significant ($P \leq .00625$) after Bonferroni correction. Abbreviations: FMS, Functional Movement Screen; LESS, Landing Error Scoring System; OCS, orthopaedic clinical specialist; SCS, sports clinical specialist; SEBT, Star Excursion Balance Test; YBT, Y Balance Test.

more than 90% of the sample reported incorporating thigh muscle strength assessment, testing procedures and LSI criterion values varied considerably across respondents. Of note, more than half of the sample indicated that they used MMT as the only strength measure to progress their patients to modified sports activity. Although MMT is a basic skill universally applied across all areas of rehabilitation, this test may lack the sensitivity to detect residual strength deficits that may be present at this phase of recovery,[9] leading to poorly informed decision making.

The subgroup analysis indicated that high-volume practitioners and certified specialists were more likely to incorporate objective strength measures. This may reflect a higher level of training or a greater appreciation for the precision offered by these more sensitive measures of strength. Conversely, less use of these measures among physical therapists who treat a low volume of patients after ACLR may be explained by an economic cost-benefit analysis, as additional equipment costs, additional training, and time requirements spent on performing more involved testing procedures may not be justified from a business perspective.

Variability continues to be perpetuated when analyzing LSI criteria identified to progress patients through the various phases of rehabilitation. Independent of testing mode, physical therapists were unable to reach a majority (greater than 50% agreement) on the required strength LSI for functional advancement. This finding may reflect a lack of clear evidence to guide practice, as suggested by the large variability of LSI thresholds found in the published literature.[1,32,39–41,44] Alternatively, research has shown that a significant time lag exists before new evidence trickles down to routine clinical practice, and that an individual clinician's willingness to adapt clinical practice to this evidence varies.[11,13,35] Thus, these findings could be explained by variability in individual adoption of contemporary strength LSI recommendations. This hypothesis may explain why a large proportion of physical therapists indicated using cutoff values of less than 90% LSI to return to sports-related activities, while many recently published reports advocate for greater than 90% LSI.[1,6,39,40,42,43]

Understanding the driving forces behind these findings is important, as this variation in clinical practice may contribute to substandard outcomes following ACLR. While optimal thresholds for strength requirements are unknown, evidence suggests that an LSI of less than 90% may increase the risk of reinjury upon resumption of level 1 sports.[19] Improving the use of objective strength testing, along with implementing strategies to facilitate the adoption of standardized LSI requirements among treating clinicians, will lead to more empowered decision making at the time of return to sport and improved ability to conduct comparative outcomes research.

Practice variation diminishes when it comes to the use of functional testing procedures, with nearly 90% of the sample reporting the use of single-leg hop testing as part of their practice to determine a patient's readiness to begin modified sports activity. Since first appearing in the literature in the early 1990s, the battery of single-limb hop tests described by Noyes et al[36] have been almost universally adopted as a necessary performance test for return-to-play decision making after ACLR.[15,31] These hop tests are reliable, easy to administer, require minimal equipment or physical space, and have demonstrated good discriminative accuracy and predictive abilities.[10,29,36] The consistency of this recommendation along with the simplicity of testing procedure may be the driving forces behind the

TABLE 5 Hop Test Selection for Decision Making to Return to Modified Sports Activity*

Utilized Hop Test Battery† (n = 790)	Value
One hop test only	154 (19.5)
At least 2 types of hop tests	627 (79.4)
At least 3 types of hop tests	426 (53.9)
Four types of hop tests	214 (27.1)

**Values are n (%).*
†Hop tests included the single hop for distance, triple hop for distance, triple crossover hop for distance, and timed 6-m hop.

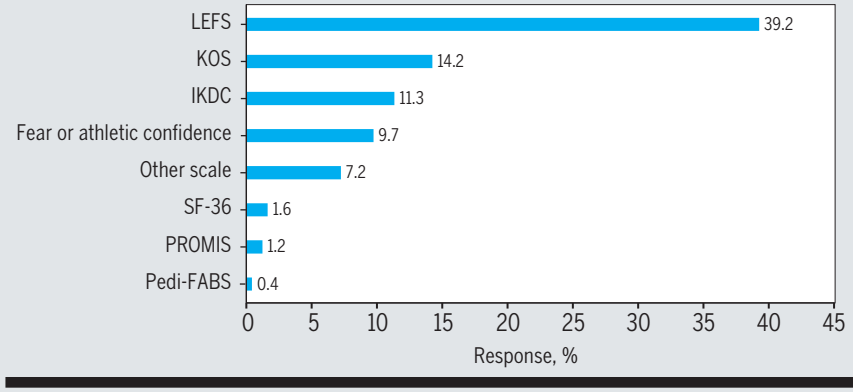


FIGURE 5. Frequencies of type of patient-reported outcome measures used from those respondents who reported their use in decision making to initiate modified sports activity. Abbreviations: IKDC, International Knee Documentation Committee Subjective Knee Evaluation Form; KOS, Knee Outcome Survey; LEFS, Lower Extremity Functional Scale; Pedi-FABS, Pediatric Functional Activity Brief Scale; PROMIS, Patient-Reported Outcome Measure Information System; SF-36, Medical Outcomes Study 36-Item Short-Form Health Survey.

level of agreement seen within the survey results.

Interestingly, high-volume practitioners and less experienced clinicians reported including more novel measures of limb function, such as the YBT or SEBT or the drop vertical jump, in their test batteries. This result may reflect a deeper appreciation or early adoption of recommendations in current literature that seek to include measures of functional limb control and movement quality to improve recognition of performance deficits or risk factors for reinjury.[18,22,43]

Use of knee-specific patient-reported outcome measures, such as the Knee Outcome Survey or International Knee Documentation Committee Subjective Knee Evaluation Form, is often advocated to quantify functional deficits that may impact a patient's successful return to activity following ACLR.[1,14] Regrettably, less than half of physical therapists in this sample reported using patient-reported outcome measures as part of their decision-making criteria to progress patients to modified sports activity. Moreover, it has become clear that physical recovery alone is not sufficient to ensure successful return to sports, and many authors have emphasized the importance of assessing psychological readiness and fear of reinjury.[2,14,15,28] Despite these recommendations, just under 10% of this sample indicated incorporation of patient-reported outcome measures related to fear or athletic confidence, neglecting the holistic framework highlighted within the biopsychosocial approach to patient management.

Limitations

There are several limitations to this study that should be recognized. The survey questionnaire was not previously validated, and although efforts were made to ensure clarity and accurate interpretation during development and pilot testing, individual variations in the interpretation of questions may exist. Nearly all of the sample were members of the APTA, which may limit the generalizability of these results to the larger population of nonmember physical therapists. Due to this relative homogeneous trait of the surveyed population, it is possible that these findings may underestimate the true degree of variability that does exist.

Due to the electronic distribution methods and anonymous nature of the survey, the authors were unable to account for e-mails that were undelivered, unopened, or received in duplicate by members of more than 1 APTA section; thus, they are unable to determine a true response rate. To ensure honesty in respondent answers, this survey was anonymous; no information was collected regarding the individual, which may have allowed some to access the survey more than once if they wished to do so.

Last, while the authors attempted to understand some of the driving forces behind clinical practice patterns identified within the survey results, there were no questions related to the effects of third-party payer regulations or other external influences on practice, thus the authors cannot account for these confounding variables.

CONCLUSION

This survey is the first to characterize the clinical practice patterns of physical therapists responsible for the treatment of patients after ACLR. The results indicate that there is a large degree of variation in rehabilitation progression among APTA members, particularly with regard to timing of activity progression, strength assessment, and use of patient-reported outcome measures. This pattern of inconsistency escalated as the time from surgery increased. Physical therapists who treated a larger volume of patients after ACLR, more recent graduates, and those with specialty certifications generally reported clinical practice patterns that were more consistent with current best evidence.[30,31] Future research should be directed toward understanding which factors contribute to this variability in clinical approach, as such inconsistency may foster feelings of confusion among patients and impact outcomes following ACLR. ◉

KEY POINTS

FINDINGS: Physical therapists report a large degree of variation in rehabilitation practice patterns after anterior cruciate ligament reconstruction, particularly with regard to time, strength assessment, and use of standardized outcome measures.

IMPLICATIONS: This variability in clinical practice standards may contribute to suboptimal outcomes and cause confusion among patients and families.

CAUTION: These results should be interpreted with caution, as this sample represents only a small portion of all licensed physical therapists who may be treating individuals following anterior cruciate ligament reconstruction.

REFERENCES

1. Adams D, Logerstedt DS, Hunter-Giordano A, Axe MJ, Snyder-Mackler L. Current concepts for anterior cruciate ligament reconstruction: a criterion-based rehabilitation progression. *J Orthop Sports Phys Ther*. 2012;42:601-614. https://doi.org/10.2519/jospt.2012.3871
2. Ardern CL. Anterior cruciate ligament reconstruction—not exactly a one-way ticket back to the preinjury level: a review of contextual factors affecting return to sport after surgery. *Sports Health*. 2015;7:224-230. https://doi.org/10.1177/1941738115578131
3. Ardern CL, Ekås G, Grindem H, et al. 2018 International Olympic Committee consensus statement on prevention, diagnosis and management of paediatric anterior cruciate ligament (ACL) injuries. *Knee Surg Sports Traumatol Arthrosc*. 2018;26:989-1010. https://doi.org/10.1007/s00167-018-4865-y
4. Ardern CL, Webster KE, Taylor NF, Feller JA. Return to the preinjury level of competitive sport after anterior cruciate ligament reconstruction surgery: two-thirds of patients have not returned by 12 months after surgery. *Am J Sports Med*. 2011;39:538-543. https://doi.org/10.1177/0363546510384798
5. Ardern CL, Webster KE, Taylor NF, Feller JA. Return to sport following anterior cruciate ligament reconstruction surgery: a systematic review and meta-analysis of the state of play. *Br J Sports Med*. 2011;45:596-606. https://doi.org/10.1136/bjsm.2010.076364

6. Barber-Westin SD, Noyes FR. Factors used to determine return to unrestricted sports activities after anterior cruciate ligament reconstruction. *Arthroscopy*. 2011;27:1697-1705. https://doi.org/10.1016/j.arthro.2011.09.009
7. Barber-Westin SD, Noyes FR. Objective criteria for return to athletics after anterior cruciate ligament reconstruction and subsequent reinjury rates: a systematic review. *Phys Sportsmed*. 2011;39:100-110. https://doi.org/10.3810/psm.2011.09.1926
8. Beischer S, Hamrin Senorski E, Thomeé C, Samuelsson K, Thomeé R. Young athletes return too early to knee-strenuous sport, without acceptable knee function after anterior cruciate ligament reconstruction. *Knee Surg Sports Traumatol Arthrosc*. 2018;26:1966-1974. https://doi.org/10.1007/s00167-017-4747-8
9. Bohannon RW. Manual muscle testing: does it meet the standards of an adequate screening test? *Clin Rehabil*. 2005;19:662-667. https://doi.org/10.1191/0269215505cr873oa
10. Bolgla LA, Keskula DR. Reliability of lower extremity functional performance tests. *J Orthop Sports Phys Ther*. 1997;26:138-142. https://doi.org/10.2519/jospt.1997.26.3.138
11. Choudhry NK, Fletcher RH, Soumerai SB. Systematic review: the relationship between clinical experience and quality of health care. *Ann Intern Med*. 2005;142:260-273. https://doi.org/10.7326/0003-4819-142-4-200502150-00008
12. Cohen SB, Sekiya JK. Allograft safety in anterior cruciate ligament reconstruction. *Clin Sports Med*. 2007;26:597-605. https://doi.org/10.1016/j.csm.2007.06.003
13. Cook C. Emotional-based practice. *J Man Manip Ther*. 2011;19:63-65. https://doi.org/10.1179/106698111X12973307659582
14. Czuppon S, Racette BA, Klein SE, Harris-Hayes M. Variables associated with return to sport following anterior cruciate ligament reconstruction: a systematic review. *Br J Sports Med*. 2014;48:356-364. https://doi.org/10.1136/bjsports-2012-091786
15. Dingenen B, Gokeler A. Optimization of the return-to-sport paradigm after anterior cruciate ligament reconstruction: a critical step back to move forward. *Sports Med*. 2017;47:1487-1500. https://doi.org/10.1007/s40279-017-0674-6
16. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). *J Med Internet Res*. 2004;6:e34. https://doi.org/10.2196/jmir.6.3.e34
17. Falconiero RP, DiStefano VJ, Cook TM. Revascularization and ligamentization of autogenous anterior cruciate ligament grafts in humans. *Arthroscopy*. 1998;14:197-205. https://doi.org/10.1016/S0749-8063(98)70041-6
18. Garrison JC, Bothwell JM, Wolf G, Aryal S, Thigpen CA. Y Balance Test™ anterior reach symmetry at three months is related to single leg functional performance at time of return to sports following anterior cruciate ligament reconstruction. *Int J Sports Phys Ther*. 2015;10:602-611.
19. Grindem H, Snyder-Mackler L, Moksnes H, Engebretsen L, Risberg MA. Simple decision rules can reduce reinjury risk by 84% after ACL reconstruction: the Delaware-Oslo ACL cohort study. *Br J Sports Med*. 2016;50:804-808. https://doi.org/10.1136/bjsports-2016-096031
20. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42:377-381. https://doi.org/10.1016/j.jbi.2008.08.010
21. Hewett TE, Di Stasi SL, Myer GD. Current concepts for injury prevention in athletes after anterior cruciate ligament reconstruction. *Am J Sports Med*. 2013;41:216-224. https://doi.org/10.1177/0363546512459638
22. Hewett TE, Ford KR, Hoogenboom BJ, Myer GD. Understanding and preventing ACL injuries: current biomechanical and epidemiologic considerations - update 2010. *N Am J Sports Phys Ther*. 2010;5:234-251.
23. Hui C, Salmon LJ, Kok A, Maeno S, Linklater J, Pinczewski LA. Fifteen-year outcome of endoscopic anterior cruciate ligament reconstruction with patellar tendon autograft for "isolated" anterior cruciate ligament tear. *Am J Sports Med*. 2011;39:89-98. https://doi.org/10.1177/0363546510379975
24. Joreitz R, Lynch A, Rabuck S, Lynch B, Davin S, Irrgang J. Patient-specific and surgery-specific factors that affect return to sport after ACL reconstruction. *Int J Sports Phys Ther*. 2016;11:264-278.
25. Kaplan Y. Identifying individuals with an anterior cruciate ligament-deficient knee as copers and noncopers: a narrative literature review. *J Orthop Sports Phys Ther*. 2011;41:758-766. https://doi.org/10.2519/jospt.2011.3384
26. Kvist J. Rehabilitation following anterior cruciate ligament injury: current recommendations for sports participation. *Sports Med*. 2004;34:269-280. https://doi.org/10.2165/00007256-200434040-00006
27. Kyritsis P, Bahr R, Landreau P, Miladi R, Witvrouw E. Likelihood of ACL graft rupture: not meeting six clinical discharge criteria before return to sport is associated with a four times greater risk of rupture. *Br J Sports Med*. 2016;50:946-951. https://doi.org/10.1136/bjsports-2015-095908
28. Lentz TA, Zeppieri G, Jr., George SZ, et al. Comparison of physical impairment, functional, and psychosocial measures based on fear of reinjury/lack of confidence and return-to-sport status after ACL reconstruction. *Am J Sports Med*. 2015;43:345-353. https://doi.org/10.1177/0363546514559707
29. Logerstedt D, Grindem H, Lynch A, et al. Single-legged hop tests as predictors of self-reported knee function after anterior cruciate ligament reconstruction: the Delaware-Oslo ACL cohort study. *Am J Sports Med*. 2012;40:2348-2356. https://doi.org/10.1177/0363546512457551
30. Logerstedt DS, Scalzitti D, Risberg MA, et al. Knee stability and movement coordination impairments: knee ligament sprain revision 2017. *J Orthop Sports Phys Ther*. 2017;47:A1-A47. https://doi.org/10.2519/jospt.2017.0303
31. Logerstedt DS, Snyder-Mackler L, Ritter RC, Axe MJ, Godges JJ. Knee stability and movement coordination impairments: knee ligament sprain. *J Orthop Sports Phys Ther*. 2010;40:A1-A37. https://doi.org/10.2519/jospt.2010.0303
32. Makhni EC, Crump EK, Steinhaus ME, et al. Quality and variability of online available physical therapy protocols from academic orthopaedic surgery programs for anterior cruciate ligament reconstruction. *Arthroscopy*. 2016;32:1612-1621. https://doi.org/10.1016/j.arthro.2016.01.033
33. Marx RG, Jones EC, Angel M, Wickiewicz TL, Warren RF. Beliefs and attitudes of members of the American Academy of Orthopaedic Surgeons regarding the treatment of anterior cruciate ligament injury. *Arthroscopy*. 2003;19:762-770. https://doi.org/10.1016/S0749-8063(03)00398-0
34. McRae SM, Chahal J, Leiter JR, Marx RG, MacDonald PB. Survey study of members of the Canadian Orthopaedic Association on the natural history and treatment of anterior cruciate ligament injury. *Clin J Sport Med*. 2011;21:249-258. https://doi.org/10.1097/JSM.0b013e318219a649
35. Morris ZS, Wooding S, Grant J. The answer is 17 years, what is the question: understanding time lags in translational research. *J R Soc Med*. 2011;104:510-520. https://doi.org/10.1258/jrsm.2011.110180
36. Noyes FR, Barber SD, Mangine RE. Abnormal lower limb symmetry determined by function hop tests after anterior cruciate ligament rupture. *Am J Sports Med*. 1991;19:513-518. https://doi.org/10.1177/036354659101900518
37. Pryor RR, Casa DJ, Vandermark LW, et al. Athletic training services in public secondary schools: a benchmark study. *J Athl Train*. 2015;50:156-162. https://doi.org/10.4085/1062-6050-50.2.03
38. Sugimoto D, Myer GD, McKeon JM, Hewett TE. Evaluation of the effectiveness of neuromuscular training to reduce anterior cruciate ligament injury in female athletes: a critical review of relative risk reduction and numbers-needed-to-treat analyses. *Br J Sports Med*. 2012;46:979-988. https://doi.org/10.1136/bjsports-2011-090895
39. Thomeé R, Kaplan Y, Kvist J, et al. Muscle strength and hop performance criteria prior to return to sports after ACL reconstruction. *Knee Surg Sports Traumatol Arthrosc*. 2011;19:1798-1805. https://doi.org/10.1007/s00167-011-1669-8
40. Toole AR, Ithurburn MP, Rauh MJ, Hewett TE, Paterno MV, Schmitt LC. Young athletes cleared for sports participation after anterior cruciate ligament reconstruction: how many actually meet recommended return-to-sport criterion cutoffs? *J Orthop Sports Phys Ther*. 2017;47:825-833. https://doi.org/10.2519/jospt.2017.7227

41. Undheim MB, Cosgrave C, King E, et al. Isokinetic muscle strength and readiness to return to sport following anterior cruciate ligament reconstruction: is there an association? A systematic review and a protocol recommendation. *Br J Sports Med*. 2015;49:1305-1310. https://doi.org/10.1136/bjsports-2014-093962

42. van Grinsven S, van Cingel RE, Holla CJ, van Loon CJ. Evidence-based rehabilitation following anterior cruciate ligament reconstruction. *Knee Surg Sports Traumatol Arthrosc*. 2010;18:1128-1144. https://doi.org/10.1007/s00167-009-1027-2

43. van Melick N, van Cingel RE, Brooijmans F, et al. Evidence-based clinical practice update: practice guidelines for anterior cruciate ligament rehabilitation based on a systematic review and multidisciplinary consensus. *Br J Sports Med*. 2016;50:1506-1515. https://doi.org/10.1136/bjsports-2015-095898

44. Wilk KE, Arrigo CA. Rehabilitation principles of the anterior cruciate ligament reconstructed knee: twelve steps for successful progression and return to play. *Clin Sports Med*. 2017;36:189-232. https://doi.org/10.1016/j.csm.2016.08.012

EARN CEUs With *JOSPT*'s Read for Credit Program

JOSPT's **Read for Credit (RFC)** program invites readers to study and analyze selected *JOSPT* articles and successfully complete online exams about them for continuing education credit. To participate in the program:

1. Go to **www.jospt.org** and click on **Read for Credit** in the top blue navigation bar that runs throughout the site.
2. Log in to read and study an article and to pay for the exam by credit card.
3. When ready, click **Take Exam** to answer the exam questions for that article.
4. Evaluate the RFC experience and receive a personalized certificate of continuing education credits.

The RFC program offers you 2 opportunities to pass the exam. You may review all of your answers—including your answers to the questions you missed. You receive **0.2 CEUs**, or 2 contact hours, for each exam passed.

JOSPT's website maintains a history of the exams you have taken and the credits and certificates you have been awarded in **My CEUs** and **Your Exam Activity**, located in the right rail of the Read for Credit page listing available exams.

APPENDIX

PHYSICAL THERAPIST SURVEY

ACLR PT/Ortho Survey

Dear Colleagues,

Anterior Cruciate Ligament (ACL) reconstruction is a common surgical procedure and patients often require an intensive and progressive course of rehabilitation. Although there has been extensive research on ACL reconstruction, clinical practice patterns detailing rehabilitation are currently unclear. In an attempt to gain insight into this factor, The Sports Medicine and Performance Center at The Children's Hospital of Philadelphia invites you to participate in this survey.

In order to participate you need to be a physical therapist or orthopaedic surgeon who currently works with patients following ACL reconstruction. The survey takes approximately 5-7 minutes to complete. Your participation is completely voluntary and your responses are anonymous. The survey includes a few demographics questions followed by a series of questions regarding your rehabilitation practices in the management of athletes after ACL reconstruction surgery.

Your responses will be kept completely confidential and analyzed anonymously. Please feel free to contact Dr. Elliot Greenberg (greenberge@email.chop.edu) or Dr. Theodore Ganley (ganley@email.chop.edu) with any questions, concerns or technical problems. If you have questions about your rights as a research subject, you can contact the Orthopedics research office at Children's Hospital of Philadelphia (267) 426-7607.

If you are interested in participating in this study, please select this option in the consent question below.
Thank you in advance for your participation.

Dr. Elliot M. Greenberg, PT, PhD, OCS
Dr. Theodore J. Ganley, MD

1. Consent

◯ I consent to participate in the study
◯ I do not consent to participate in the study

2. Are you an orthopaedic surgeon or a physical therapist?

◯ Orthopaedic surgeon
◯ Physical therapist
◯ Neither

APPENDIX

PRACTICE PATTERNS

3. How many anterior cruciate ligament (ACL) reconstructions do you treat per year?
- ◯ None
- ◯ 1-5
- ◯ 6-10
- ◯ 10+

4. What is your primary practice setting?
- ◯ Private practice
- ◯ Hospital-based outpatient facility
- ◯ Corporate-owned outpatient practice
- ◯ Academic/collegiate rehabilitation facility
- ◯ Other

4a. Other: _____ (Please specify)

5. How many years have you been practicing?
- ◯ 0-4 years
- ◯ 5-10 years
- ◯ 11-15 years
- ◯ 16 or more

6. Are you currently an APTA member?
- ◯ Yes
- ◯ No

7. Are you an APTA Orthopedic (OCS) or Sports (SCS) Certified Specialist?
- ◯ Yes
- ◯ No

APPENDIX

8. What state do you practice in?

- ◯ Alabama
- ◯ Alaska
- ◯ Arizona
- ◯ Arkansas
- ◯ California
- ◯ Colorado
- ◯ Connecticut
- ◯ Delaware
- ◯ Florida
- ◯ Georgia
- ◯ Hawaii
- ◯ Idaho
- ◯ Illinois
- ◯ Indiana
- ◯ Iowa
- ◯ Kansas
- ◯ Kentucky
- ◯ Louisiana
- ◯ Maine
- ◯ Maryland
- ◯ Massachusetts
- ◯ Michigan
- ◯ Minnesota
- ◯ Mississippi
- ◯ Missouri
- ◯ Montana
- ◯ Nebraska
- ◯ Nevada
- ◯ New Hampshire
- ◯ New Jersey
- ◯ New Mexico
- ◯ New York
- ◯ North Carolina
- ◯ North Dakota
- ◯ Ohio
- ◯ Oklahoma
- ◯ Oregon
- ◯ Pennsylvania
- ◯ Rhode Island
- ◯ South Carolina
- ◯ South Dakota
- ◯ Tennessee
- ◯ Texas
- ◯ Utah Vermont
- ◯ Virginia
- ◯ Washington
- ◯ West Virginia
- ◯ Wisconsin
- ◯ Wyoming

9. Where did you hear about this survey?

- ◯ Orthopedic section email
- ◯ Sports section email
- ◯ Other email
- ◯ Social media (Facebook, Twitter, etc.)

APPENDIX

For the remainder of this survey, please answer the questions based upon your typical treatment of the patient described below.

"Your patient is a 17-year-old female soccer player who underwent ACL reconstruction using a hamstring autograft. There were no concomitant injuries and she is having an uncomplicated post-operative recovery. Her goal is to return to soccer competition at the collegiate level upon full recovery."

10. For the patient described above, how long would your typical course of rehabilitation be? (i.e. how long would you treat them within an office setting?)

- ◯ 1-3 months
- ◯ 4-5 months
- ◯ 6-8 months
- ◯ 9-12 months
- ◯ >12 months

11. Who is responsible for determining this athlete's readiness to begin to run, initiate plyometric and agility training and unrestricted return to sports?

- ◯ Orthopaedic surgeon
- ◯ Rehabilitation specialist (PT, ATC)
- ◯ Both the orthopeadic and rehabilitation specialist
- ◯ Other

11a. Other: _____

PROGRESSION TO JOGGING AFTER ACL RECONSTRUCTION

12. I would typically allow the athlete in this example to begin jogging at _____ months post-surgery. (fill in the blank from the choices available)

- ◯ 2-3 months
- ◯ 3-4 months
- ◯ 4-5 months
- ◯ 6+ months

13. Are there specific physical tests, examination findings, or criteria that you utilize in order to assist in the decision to progress to jogging? (Check all that apply)

- ☐ Knee range of motion
- ☐ Strength assessment (manual muscle testing)
- ☐ Strength assessment (handheld dynamometry)
- ☐ Strength assessment (isokinetic testing)
- ☐ Knee effusion
- ☐ Lower extremity functional testing or balance assessment
- ☐ Patient-reported outcome measures
- ☐ ACL laxity test (e.g. Lachmans, Anterior Drawer, etc.)
- ☐ None
- ☐ Other

(Check all that apply)

13a. Other: _____

If "Knee strength" is selected:

14. Strength Assessment: What tests do you use? (Check all that apply)

- ☐ Manual Muscle Testing (MMT)
- ☐ Isometric Testing (i.e. Hand Held Dynamometry (HHD))
- ☐ Isokinetic testing
- ☐ Repetition maximum (RM) testing
- ☐ Other

(Check all that apply)

14a. Other: _____

APPENDIX

If "Isometric testing or isokinetic testing" is selected:

15. What QUADRICEPS strength criteria is required for progression to jogging?

- ◯ Side-to-side deficit of less than 30%
- ◯ Side-to-side deficit of less than 25%
- ◯ Side-to-side deficit of less than 20%
- ◯ Side-to-side deficit of less than 15%
- ◯ Side-to-side deficit of less than 10%
- ◯ Other

15a. Other: _____

16. What HAMSTRINGS strength criteria is required for progression to jogging?

- ◯ Side-to-side deficit of less than 30%
- ◯ Side-to-side deficit of less than 25%
- ◯ Side-to-side deficit of less than 20%
- ◯ Side-to-side deficit of less than 15%
- ◯ Side-to-side deficit of less than 10%
- ◯ Other

16a. Other: _____

If "Isokinetic testing" is selected:

17. What speed(s) of testing do you utilize for isokinetic testing? (Check all that apply)

- ☐ 60 degrees per second
- ☐ 120 degrees per second
- ☐ 180 degrees per second
- ☐ 240 degrees per second
- ☐ 300 degrees per second
- ☐ Other

(Check all that apply)

17a. Other: _____

If "Repetition maximum testing" is selected:

18. Repetition maximum (RM) Testing: What means of RM testing do you utilize? (Check all that apply)

- ☐ Leg Press
- ☐ Knee Extension
- ☐ Knee Flexion
- ☐ Other

18a. Other: _____

If "Leg press" is selected:

19. Leg Press: What criteria is required for progression to jogging?

- ◯ Side-to-side deficit of less than 30%
- ◯ Side-to-side deficit of less than 25%
- ◯ Side-to-side deficit of less than 20%
- ◯ Side-to-side deficit of less than 15%
- ◯ Side-to-side deficit of less than 10%
- ◯ Other

19a. Other: _____

If "Knee extension" is selected:

20. Knee Extension: What criteria is required for progression to jogging?

- ◯ Side-to-side deficit of less than 30%
- ◯ Side-to-side deficit of less than 25%
- ◯ Side-to-side deficit of less than 20%
- ◯ Side-to-side deficit of less than 15%
- ◯ Side-to-side deficit of less than 10%
- ◯ Other

20a. Other: _____

APPENDIX

If "Knee flexion" is selected:

21. Knee Flexion: What criteria is required for progression to jogging?

- ◯ Side-to-side deficit of less than 30%
- ◯ Side-to-side deficit of less than 25%
- ◯ Side-to-side deficit of less than 20%
- ◯ Side-to-side deficit of less than 15%
- ◯ Side-to-side deficit of less than 10%
- ◯ Other

21a. Other: _____

If "Lower extremity functional testing…" is selected:

22. Functional performance: What tests do you use to assist with the decision to progress to jogging? (Check all that apply)

- ☐ Straight leg raise
- ☐ Functional movement screen (FMS)
- ☐ Y or Star balance testing
- ☐ Lateral step down test
- ☐ Balance assessment tool (e.g. Balance Error Scoring System BESS)
- ☐ Other

22a. Other: _____

If "Functional movement screen" is selected:

23. What is your criteria for advancement on functional movement screen? (Check all that apply)

- ☐ Composite FMS score
- ☐ Performance on isolated movements
- ☐ Side-side symmetry for unilateral movements
- ☐ Other

23a. Other: _____

If "Y-balance test" is selected:

24. What is your criteria for advancement on Y-Balance Test? (Check all that apply)

- ☐ Anterior reach difference of less than 4cm
- ☐ Composite reach with less than 10% side-to-side asymmetry
- ☐ Other

24a. Other: _____

If "Patient-reported outcome measures" is selected:

25. Patient-Reported Outcome Measures(PROM): What PROM do you use to assist with the decision to progress to jogging? (Check all that apply)

- ☐ Lower Extremity Functional Scale (LEFS)
- ☐ International Knee Disability Committee (IKDC)
- ☐ Knee Outcome Survey (KOS)
- ☐ Short-Form 36 (SF-36)
- ☐ PROMIS Quality of Life Measures
- ☐ Pedi-Fabs Scale
- ☐ Fear or Self-Efficacy Based Survey (e.g. Tampa Scale of Kinesiophobia)
- ☐ Other

(Check all that apply)

25a. Other: _____

APPENDIX

PROGRESSION TO MODIFIED SPORTS ACTIVITY AFTER ACL RECONSTRUCTION

26. I would typically allow the athlete in this example to begin modified sports-specific activities (agilities, sports-specific drills/skills, etc) at _____ months post-surgery. (fill in the blank from the choices available)

- ◯ ≤ 3 months
- ◯ 4-5 months
- ◯ 6-7 months
- ◯ 8-9 months
- ◯ 9-12 months
- ◯ 12 or more months
- ◯ I typically do not see patient's during this phase of rehabilitation

27. Are there specific physical tests, examination findings, or criteria that you utilize in order to assist in the decision to progress to sports-specific training? (Check all that apply)

- ☐ Knee range of motion
- ☐ Knee strength
- ☐ Knee effusion
- ☐ Lower extremity functional testing or balance assessment
- ☐ Patient-reported outcome measures
- ☐ ACL laxity test (e.g. Lachmans, Anterior Drawer, etc.)
- ☐ None
- ☐ Other

(Check all that apply)

27a. Other: _____

If "Knee strength" is selected:

28. Strength Assessment: What tests do you use? (Check all that apply)

- ☐ Manual Muscle Testing (MMT)
- ☐ Isometric Testing (i.e. Hand Held Dynamometry (HHD))
- ☐ Isokinetic testing
- ☐ Repetition maximum (RM) testing
- ☐ Other

(Check all that apply)

28a. Other: _____

APPENDIX

If "Isometric testing or isokinetic testing" is selected:

29. What QUADRICEPS strength criteria is required for progression to sport-specific activities?
- ◯ Side-to-side deficit of less than 25%
- ◯ Side-to-side deficit of less than 20%
- ◯ Side-to-side deficit of less than 15%
- ◯ Side-to-side deficit of less than 10%
- ◯ Side-to-side deficit of less than 5%
- ◯ Other

29a. Other: _____

30. What HAMSTRINGS strength criteria is required for progression to sport-specific activities?
- ◯ Side-to-side deficit of less than 25%
- ◯ Side-to-side deficit of less than 20%
- ◯ Side-to-side deficit of less than 15%
- ◯ Side-to-side deficit of less than 10%
- ◯ Side-to-side deficit of less than 5%
- ◯ Other

30a. Other: _____

If "Isokinetic testing" is selected:

31. What speed(s) of testing do you utilize for isokinetic testing? (Check all that apply)
- ☐ 60 degrees per second
- ☐ 120 degrees per second
- ☐ 180 degrees per second
- ☐ 240 degrees per second
- ☐ 300 degrees per second
- ☐ Other

(Check all that apply)

31a. Other: _____

If "Repetition maximum testing" is selected:

32. Repetition maximum (RM) Testing: What means of RM testing do you utilize? (Check all that apply)
- ☐ Leg Press
- ☐ Knee Extension
- ☐ Knee Flexion
- ☐ Other

32a. Other: _____

If "Leg press" is selected:

33. Leg Press: What criteria is required for progression sport-specific activities?
- ◯ Side-to-side deficit of less than 25% to
- ◯ Side-to-side deficit of less than 20%
- ◯ Side-to-side deficit of less than 15%
- ◯ Side-to-side deficit of less than 10%
- ◯ Side-to-side deficit of less than 5%
- ◯ Other

33a. Other: _____

If "Knee extension" is selected:

34. Knee Extension: What criteria is required for progression to sport-specific activities?
- ◯ Side-to-side deficit of less than 25%
- ◯ Side-to-side deficit of less than 20%
- ◯ Side-to-side deficit of less than 15%
- ◯ Side-to-side deficit of less than 10%
- ◯ Side-to-side deficit of less than 5%
- ◯ Other

34a. Other: _____

APPENDIX

If "Knee flexion" is selected:

35. Knee Flexion: What criteria is required for progression to sport-specific activities?

- ◯ Side-to-side deficit of less than 25%
- ◯ Side-to-side deficit of less than 20%
- ◯ Side-to-side deficit of less than 15%
- ◯ Side-to-side deficit of less than 10%
- ◯ Side-to-side deficit of less than 5%
- ◯ Other

35a. Other: _____

If "Lower extremity functional testing…" is selected:

36. Functional Performance: what tests do you use to assist with the decision to progress to specific activities? (Check all that apply)

- ☐ Functional movement screen (FMS)
- ☐ Y balance test sport-
- ☐ Single leg hop test
- ☐ Drop vertical jump
- ☐ Balance assessment tool (e.g. Balance Error Scoring System BESS)
- ☐ Patient-reported outcome measure Landing
- ☐ Error Scoring System (LESS) test Vail sport
- ☐ test
- ☐ Other

(Check all that apply)

36a. Other: _____

If "Functional movement screen" is selected:

37. What is your criteria for advancement on functional movement screen? (Check all that apply)

- ☐ Composite FMS score
- ☐ Performance on isolated movements
- ☐ Side-side symmetry for unilateral movements
- ☐ Other

37a. Other: _____

If "Y-balance test" is selected:

38. What is your criteria for advancement on Y-Balance Test? (Check all that apply)

- ☐ Anterior reach difference of less than 4cm
- ☐ Composite reach with less than 10% side-to-side asymmetry
- ☐ Other

38a. Other: _____

If "Single leg hop test" is selected:

39. What hops do you use on hop testing? (Check all that apply)

- ☐ Single hop
- ☐ Triple hop
- ☐ Cross-over triple hop
- ☐ 6M timed hop
- ☐ Other

39a. Other: _____

40. What is your criteria for advancement on single leg hop testing?

- ◯ Side-to-side deficit of less than 25%
- ◯ Side-to-side deficit of less than 20%
- ◯ Side-to-side deficit of less than 15%
- ◯ Side-to-side deficit of less than 10%
- ◯ Side-to-side deficit of less than 5%
- ◯ Other

40a. Other: _____

APPENDIX

If "Patient-reported outcome measures" is selected:

41. Patient-Reported Outcome Measures(PROM): What PROM do you use to assist with the decision to progress to sport-specific activities? (Check all that apply)

- ☐ Lower Extremity Functional Scale (LEFS)
- ☐ International Knee Disability Committee (IKDC)
- ☐ Knee Outcome Survey (KOS)
- ☐ Short-Form 36 (SF-36)
- ☐ PROMIS Quality of Life Measures
- ☐ Pedi-Fabs Scale
- ☐ Fear or Self-Efficacy Based Survey (e.g. Tampa Scale of Kinesiophobia)
- ☐ Other

(Check all that apply)

41a. Other: _____

PROGRESSION TO FULL UNRESTRICTED SPORTS ACTIVITY

42. I would typically allow the athlete in this example to begin unrestricted sports at _____ months post-surgery. (fill in the blank from the choices available)

- ◯ ≤ 3 months
- ◯ 4-5 months
- ◯ 6-7 months
- ◯ 7-8 months
- ◯ 9-12 months
- ◯ 12 or more months
- ◯ I typically do not see patients during this phase of rehabilitation

43. Are there any additional tests, measures, or criteria, beyond those needed to initiate sports-related activities, that you require before allowing an athlete to participate in UNRESTRICTED sports activity?

- ◯ Yes
- ◯ No

43a. What additional tests or measures do you require? _____

RECOMMENDATION OF ONGOING INJURY PREVENTION PROGRAM

44. Would you recommend an ACL injury prevention program this patient at discharge?

- ◯ Yes for
- ◯ No

If ACL prevention program recommended:

45. What ACL prevention program do you recommend?

- ◯ Sportsmetrics
- ◯ Prevent Injury, Enhance Performance (PEP) Program
- ◯ FIFA 11+
- ◯ Other

45a. Other: _____

USE OF FUNCTIONAL BRACING AT TIME OF RETURN TO SPORTS

46. Would you typically recommend the use of a knee brace during sports activities for this patient?

- ◯ Yes
- ◯ No

If brace recommended:

47. What type of brace do you typically recommend?

- ◯ Functional ACL brace
- ◯ Neoprene knee sleeve
- ◯ Other

47a. Other: _____

ALAN C. LEE, DPT, PhD[1] • TODD E. DAVENPORT, DPT, MPH[2] • KEN RANDALL, PT, PhD[3]

Telehealth Physical Therapy in Musculoskeletal Practice

J Orthop Sports Phys Ther 2018;48(10):736-739. doi:10.2519/jospt.2018.0613

The World Health Organization defines *telehealth* as the delivery of health care services, where distance is a critical factor, by all health care professionals using information and communication technologies for the exchange of valid information for diagnosis, treatment, and prevention of disease and injuries; for research and evaluation; and for the continuing education of health care providers, all in the interests of advancing the health of individuals and their communities.[18] The American Physical Therapy Association (APTA) defines telehealth as the use of secure electronic communications to provide and deliver a host of health-related information and health care services, including but not limited to physical therapy–related information and services for patients and clients.[1]

Internationally, physical therapists utilize *telerehabilitation* as the common term for telehealth applications. For example, the Australian Physiotherapy Association's position statement[2] describes the provision of rehabilitation across the spectrum of acute, subacute, and community settings at a distance, using telecommunication technology to deliver real-time audio and video conferencing between providers and patients as synchronous telehealth. Other telehealth applications include secure electronic transmission of clinical information and medical data, described as asynchronous or store-and-forward telehealth.

Operationally, a health care professional at a distant site may interact with a patient who is at the originating site via synchronous and asynchronous telehealth. In addition, remote patient monitoring has gained support alongside the advent of emerging biotechnology, virtual reality, and wearable technology. Rehabilitation professionals utilize telerehabilitation to deliver physical, occupational, and speech therapy, while physicians, nurse practitioners, and other health care professionals utilize telemedicine in practice. As the digital health field grows with new technologies, an absolute definition of telehealth, telerehabilitation, and telemedicine remains elusive.[18]

Telehealth may enhance patient satisfaction, overcome barriers to access to physical therapy services, and reduce the costs of musculoskeletal care in society. Russell and colleagues[16] reported high patient satisfaction with a 6-week telerehabilitation intervention compared to usual care in outpatients after total knee arthroplasty in Australia. Palsbo[12] reported that telerehabilitation helped Medicaid programs in the United States deliver specialized physical therapy care to locations with provider shortages in rural communities. Tousignant and colleagues[19] found a cost differential in favor of the telerehabilitation group compared to the usual-care group after a patient following total knee arthroplasty had to travel more than 30 km (round trip) from home to a physical therapy clinic in Canada.

Overall, telehealth physical therapy has the potential to transform many critical areas of care in musculoskeletal practice. However, the amount of hype around telehealth needs to be carefully examined, because widespread implementation has been stalled by payment and regulatory barriers in physical therapy.[8,9] Therefore, the purposes of this Viewpoint are to highlight (1) the current level of implementation, (2) telehealth musculoskeletal evidence, and (3) future opportunities in the digital age.

Telehealth Implementation in Musculoskeletal Practice

Various health systems around the world have utilized telehealth to improve access to care in musculoskeletal practice. Numerous innovative musculoskeletal practices have emerged with the advancement

[1]Mount Saint Mary's University, Los Angeles, CA. [2]University of the Pacific, Stockton, CA. [3]University of Oklahoma Schusterman Center, Tulsa, OK. Drs Lee and Davenport serve as members of the American Physical Therapy Association's Frontiers in Rehabilitation, Science, and Technology Council and members of the advisory board at Bluejay Mobile Health, Inc. The authors certify that they have no other affiliations with or financial involvement in any organization or entity with a direct financial interest in the subject matter or materials discussed in the article. Address correspondence to Dr Alan C. Lee, 10 Chester Place, Los Angeles, CA 90007. E-mail: Allee@msmu.edu

of smart technologies, such as apps, mobile devices, and wearable technologies. More importantly, use of digital technology and devices has become ubiquitous in developed countries.

In the United States, the Department of Veterans Affairs and the Indian Health Service utilize telehealth to facilitate the need for musculoskeletal physical therapy expertise where provider shortages exist.[8] Kaiser Permanente, a US-based health maintenance organization, utilizes telehealth physical therapy for musculoskeletal triage and consultations.[8] In the United Kingdom, PhysioDirect telephone consultation services for patients with musculoskeletal problems are advancing toward a telehealth delivery model.[20] Online physical therapy musculoskeletal practice has emerged in Australia, while Alberta, Canada has allowed physical therapy across provincial borders with the courtesy register process for telerehabilitation.[13] Other international collaborations between 4 European Union member states (Italy, the Netherlands, Spain, and Poland) are establishing telerehabilitation delivery standards to achieve accreditation from their respective governmental agencies.[6]

As telehealth implementation in physical therapy ramps up, it is helpful to address real-world barriers, such as unpredictable weather hampering internet connectivity or the inability to manually assist or assess a patient during a telehealth encounter. Recently, the US Federation of State Boards of Physical Therapy developed its model regulation of telehealth to assist regulatory physical therapy stakeholders in addressing administrative, ethical, and technical standards in practice.[4] For example, patient-informed consent and verification of identity during a telehealth encounter, as well as privacy, patient safety, and emergency procedures, must be in place prior to telehealth practice. In addition, principles and guidelines developed by the American Telemedicine Association indicate a practitioner's ethical responsibility to discuss with a patient his or her right to refuse or discontinue a telehealth service.[15] The APTA, in partnership with the Private Practice Section and the Academy of Orthopaedic Physical Therapy, has developed educational webinars, monographs, and frequently-asked-questions documents for practitioners to address barriers and support implementation of telehealth.[1] Last, the Frontiers in Rehabilitation, Science, and Technology Council was established to advance science and technology in physical therapy, in concert with APTA leadership and approved by the APTA Board of Directors.

Hence, current guidelines and resources in telehealth should be reviewed by musculoskeletal practitioners (**TABLE**). Most recently, the World Confederation for Physical Therapy (WCPT) and the International Network of Physiotherapy Regulatory Authorities announced a formal collaboration to develop a model regulation of providing telehealth physical therapy rehabilitation services.[20] Therefore, musculoskeletal physical therapy practitioners have an opportunity to collaborate with international organizations, including the WCPT and International Network of Physiotherapy Regulatory Authorities, to improve access to quality care in physical therapy with telerehabilitation in the digital age. Overall, global engagement of telerehabilitation physical therapy professionals in musculoskeletal practice could address the WCPT goals of knowledge sharing and recognition of value for the physical therapy profession in the world.

Telehealth Musculoskeletal Evidence

As reported in the literature, technology advancement in the global marketplace has generated telerehabilitation research in musculoskeletal physical therapy.[5] Several academic research centers, including the University of Queensland (Australia) and the University of Sherbrooke (Canada), have led research evidence in musculoskeletal telerehabilitation. Early investigations in telehealth focused on the reliability and validity of musculoskeletal examinations of upper and lower extremities, as well as musculoskeletal dysfunction.[17] More recently, telehealth feasibility in total joint replacement was examined.[11] A systematic review of telehealth videoconferencing for physical therapy in people with musculoskeletal conditions noted a moderate quality of intervention and a positive impact on health outcomes and satisfaction.[5] However, this review and past systematic reviews[8] indicate lack of cost analysis as a weakness to remedy before promotion of delivery of telehealth musculoskeletal practice. The main justification for telerehabilitation was local shortages of practitioners, especially in rural areas.[12] Therefore, future telehealth research addressing the physical therapist workforce and cost analysis is urgently needed.

TABLE Key Topics Covered in Telehealth Resources[1,2,4,15]

Topics in Telehealth Resources	APA[2] (Under Review)	APTA[1]	ATA[15]	FSBPT[4]
Administrative			x	x
Clinical	x	x	x	x
Education		x		
Ethical	x	x	x	x
Regulation	x	x		x
Research		x		
Technical		x	x	x

Abbreviations: APA, Australian Physiotherapy Association; APTA, American Physical Therapy Association; ATA, American Telemedicine Association; FSBPT, Federation of State Boards of Physical Therapy.

In academia, a US-based study involving 139 nurse practitioner, occupational therapy, and physical therapist students (who comprised half of the sample) assessed student perceptions of telehealth in treating patients as interprofessional teams in both simulated and actual outpatient musculoskeletal clinics.[14] Findings revealed that students, although remaining positive about telehealth, perceived that it introduced barriers to the treatment process, particularly in establishing rapport with patients and other members of the health care team. The study, which corroborates similar findings from the literature,[10] recommends that physical therapy education about telehealth include content that specifically addresses potential barriers and ethical delivery, while optimizing the advantages that the technology can provide in the continuum of care.

According to one musculoskeletal researcher (Trevor G. Russell) in a March 2018 e-mail, a state-of-the-art telerehabilitation clinic is used at the University of Queensland (Australia), with physical and occupational therapy, speech pathology, and audiology students under the supervision of clinical educators, to develop unique skill sets in telehealth. Students are introduced to the theory and practical aspects of telehealth service delivery through online learning modules, including clinical scenarios, and also through a hands-on practicum in which students and clinical educators can work through clinical cases via telehealth. Overall, further academic research is necessary for telehealth, as a means for physical therapist musculoskeletal practice, to become a reality.

Future Telehealth Opportunities

Regulations allowing cross-state practice represent telehealth implementation opportunities in the future.[3,13] As of March 2018, 21 states in the United States have enacted legislation to allow cross-state physical therapy practice, known as the *interstate compact*.[3] The purpose of the compact is to increase consumer access to physical therapy services by reducing regulatory barriers to interstate practice. In Canada, a memorandum of understanding for cross-border physical therapy has been established as well.[13] Similar telemedicine regulatory actions removed barriers for nursing and medical professionals to address provider shortages and access to quality health care.[8] Given the potential importance of telehealth to reduce health care disparities associated with limited access to medical services, funding support for health services research with community collaborations for ethical delivery should be established within the physical therapy profession. These initiatives necessarily require collaboration between physical therapy professionals and their professional societies to advance telehealth in musculoskeletal practice.

Diffusion of telehealth practice will require clear solutions to key topics identified in the literature (**TABLE**). Experienced telehealth practitioners should lead in educating novice users and students on practical applications and lessons learned from the real world. For example, interpersonal aspects of physical therapist–patient communication using telehealth require unique verbal and nonverbal behaviors for information giving and seeking, social conversation, and partnership building.[10] By offering both in-person clinic visits and follow-up telehealth encounters with patients and caregivers, the opportunity for personalized care can be further established with secure technology. Hence, we invite musculoskeletal professionals in health services, engineering, and academic research to collaborate with the Frontiers in Rehabilitation, Science, and Technology Council and the WCPT to advance science and technology that will favorably influence the future of musculoskeletal telehealth practice.

As the aging population and health care expenditures grow in the future, telehealth in musculoskeletal practice may assist in solving the looming health care crisis. Currently, physical therapists are not listed as Medicare-eligible telehealth providers in the United States.[8] This barrier alone limits opportunities for millions of older adults to access high-quality physical therapy services via telehealth, as traveling to a physical therapy clinic can be a challenge for some older adults. In the literature, Medicaid programs have supported physical therapy telehealth services,[12] and cost benefits have been demonstrated by in-home telerehabilitation in Canada.[19] Therefore, we believe it is time to advocate for Medicare eligibility for telehealth physical therapy by funding musculoskeletal telehealth research, supporting payment policy, and educating providers in practice. Ultimately, telehealth is a tool to provide high-quality personalized physical therapy as part of musculoskeletal practice in society.

Conclusion

The use of telehealth in musculoskeletal practice has been stalled by implementation, research, and policy barriers in physical therapy.[7-9] Systematic reviews in musculoskeletal telehealth literature demonstrate comparable outcomes and patient satisfaction.[5] Barriers to consumer access to physical therapy services are being lessened by physical therapy stakeholders.[3,13] In the future, telehealth in musculoskeletal research and practice must aim to establish equitability in access, cost-effectiveness, clinical outcomes, and ethical delivery in the digital age. ◉

REFERENCES

1. American Physical Therapy Association. Telehealth. Available at: http://www.apta.org/Telehealth/. Accessed January 11, 2018.
2. Australian Physiotherapy Association. Telerehabilitation and physiotherapy: position statement. Available at: http://www.physiotherapy.asn.au/DocumentsFolder/Advocacy_Position_Telerehabilitation_2009.pdf. Accessed February 6, 2018.
3. Federation of State Boards of Physical Therapy. Physical therapy licensure compact. Available at: http://www.fsbpt.org/FreeResources/PhysicalTherapyLicensurecompact.aspx. Accessed August 3, 2018.

4. Federation of State Boards of Physical Therapy. Telehealth in physical therapy: policy recommendations for appropriate regulation. Available at: https://www.fsbpt.org/FreeResources/RegulatoryResources/TelehealthinPhysicalTherapy.aspx. Accessed January 11, 2018.
5. Grona SL, Bath B, Busch A, Rotter T, Trask C, Harrison E. Use of videoconferencing for physical therapy in people with musculoskeletal conditions: a systematic review. *J Telemed Telecare*. 2018;24:341-355. https://doi.org/10.1177/1357633X17700781
6. Habilis® Platform. Available at: https://www.signomotus.it/en/prodprot/habilisplatform_en.php. Accessed July 31, 2018.
7. Kairy D, Lehoux P, Vincent C, Visintin M. A systematic review of clinical outcomes, clinical process, healthcare utilization and costs associated with telerehabilitation. *Disabil Rehabil*. 2009;31:427-447. https://doi.org/10.1080/09638280802062553
8. Lee AC, Harada N. Telehealth as a means of health care delivery for physical therapist practice. *Phys Ther*. 2012;92:463-468. https://doi.org/10.2522/ptj.20110100
9. McAllister J. Is it time for mainstream 'telephysiotherapy' in Canada? Available at: https://physiotherapy.ca/blog/it-time-mainstream-telephysiotherapy-canada. Accessed February 6, 2018.
10. Miller EA. The technical and interpersonal aspects of telemedicine: effects on doctor-patient communication. *J Telemed Telecare*. 2003;9:1-7. https://doi.org/10.1258/135763303321159611
11. Nelson MJ, Crossley KM, Bourke MG, Russell TG. Telerehabilitation feasibility in total joint replacement. *Int J Telerehabil*. 2017;9:31-38. https://doi.org/10.5195/ijt.2017.6235
12. Palsbo SE. Medicaid payment for telerehabilitation. *Arch Phys Med Rehabil*. 2004;85:1188-1191. https://doi.org/10.1016/j.apmr.2003.09.008
13. Physiotherapy Alberta. Cross-border physiotherapy. Available at: https://www.physiotherapyalberta.ca/applicants/cross_border_physiotherapy#forms_and_instructions. Accessed February 6, 2018.
14. Randall K, Steinheider B, Isaacson M, et al. Measuring knowledge, acceptance, and perceptions of telehealth in an interprofessional curriculum for student nurse practitioners, occupational therapists, and physical therapists. *J Interact Learn Res*. 2016;27:339-353.
15. Richmond T, Peterson C, Cason J, et al. American Telemedicine Association's principles for delivering telerehabilitation services. *Int J Telerehabil*. 2017;9:63-68. https://doi.org/10.5195/ijt.2017.6232
16. Russell TG, Buttrum P, Wootton R, Jull GA. Internet-based outpatient telerehabilitation for patients following total knee arthroplasty: a randomized controlled trial. *J Bone Joint Surg Am*. 2011;93:113-120. https://doi.org/10.2106/JBJS.I.01375
17. Russell TG, Jull GA, Wootton R. Can the Internet be used as a medium to evaluate knee angle? *Man Ther*. 2003;8:242-246. https://doi.org/10.1016/S1356-689X(03)00016-X
18. Telehealth Alliance of Oregon. Telemedicine or Telehealth - Definitions. Available at: http://www.ortelehealth.org/content/telemedicine-or-telehealth-definitions. Accessed March 19, 2018.
19. Tousignant M, Moffet H, Nadeau S, et al. Cost analysis of in-home telerehabilitation for post-knee arthroplasty. *J Med Internet Res*. 2015;17:e83. https://doi.org/10.2196/jmir.3844
20. World Confederation for Physical Therapy. WCPT and INPTRA announce a formal collaboration. Available at: http://www.wcpt.org/sites/wcpt.org/files/files/resources/press/WCPT_PressRelease_20Feb17.pdf. Accessed February 9, 2018.

THIAGO JAMBO ALVES LOPES, PT, MSc[1,2] • **DEISI FERRARI,** PT, PhD[3,4] • **JOSHUA IOANNIDIS,** PT[1,5]
MILENA SIMIC, PT, PhD[1] • **FÁBIO MÍCOLIS DE AZEVEDO,** PT, PhD[6] • **EVANGELOS PAPPAS,** PT, PhD, OCS[1]

Reliability and Validity of Frontal Plane Kinematics of the Trunk and Lower Extremity Measured With 2-Dimensional Cameras During Athletic Tasks: A Systematic Review With Meta-analysis

BACKGROUND: Two-dimensional (2-D) analysis is commonly used to quantify frontal plane kinematics of the trunk and lower extremity. However, there are conflicting results regarding the reliability and validity of these measurements.

OBJECTIVE: To synthesize the current literature to determine whether 2-D analysis is a reliable and valid method of measuring frontal plane kinematics of the trunk and lower extremity during squatting, landing, and cutting tasks.

METHODS: For this systematic review with meta-analysis, MEDLINE, CINAHL, Embase, Scopus, and SPORTDiscus databases were searched from inception until March 2017. The authors included 16 studies that evaluated the reliability and/or validity of 2-D measurements of frontal plane trunk and/or lower extremity kinematics when compared to 3-D measurements during any of the following tasks: squatting, landing, or cutting.

RESULTS: Intrarater reliability (intraday and interday) and interrater reliability of the 2-D video measurements varied from moderate to excellent. In terms of validity, there was poor agreement between the 2-D and 3-D methods, with no correlation between 2-D knee frontal plane projection angle and 3-D knee frontal plane angles ($r = 0.127$, $P = .094$) for the single-leg squat, but a moderate to good relationship ($r = 0.619$, $P<.001$) for the landing task.

CONCLUSION: Two-dimensional video analysis of frontal plane trunk and lower extremity kinematics is reliable, but this appears to be dependent on the task and the type of reliability evaluated. The current evidence does not support the use of 2-D video analysis for measuring trunk and lower extremity frontal plane kinematics when accurate measures are required.

LEVEL OF EVIDENCE: Diagnosis, level 3.
J Orthop Sports Phys Ther 2018;48(10):812-822. Epub 12 Jun 2018. doi:10.2519/jospt.2018.8006

KEY WORDS: *2-D motion, 3-D motion, biomechanics, kinematics, video analysis*

Three-dimensional motion capture has been used extensively to measure movement in a variety of orthopaedic settings, from research[6] to clinical practice.[27] Previous studies have shown that 3-D motion-capture systems measure motion accurately and with small measurement error.[6,19] Because 3-D motion capture is a noninvasive method that does not involve radiation, it has become the gold standard for assessing frontal plane motions of the lower extremity and trunk.[9,27]

For example, in a prospective study of female athletes, 3-D motion capture was used to identify risk factors for anterior cruciate ligament injury, such as increased knee valgus angle and moment during landing from a jump task.[9] Similarly, uses of 3-D motion capture in persons with patellofemoral pain have ranged from investigating potential risk factors related to athletic tasks[26] to real-time gait retraining to improve pain levels and function in this population.[30]

Although 3-D motion capture is used to assess kinematics across different fields, its clinical applicability is limited due to the expense of the equipment and the requirement that system operators undergo extensive training.[20,22] As a result, it is unrealistic to expect this form of analysis to be carried out on a large scale; in fact, most biomechanical studies have limited sample sizes.

[1]The University of Sydney, Discipline of Physiotherapy, Faculty of Health Sciences, Sydney, Australia. [2]Research Laboratory of Exercise Science, Centro de Educação Física Almirante Adalberto Nunes, Brazilian Navy, Rio de Janeiro, Brazil. [3]Postgraduation Program-Interunits Bioengineering, School of Engineering of São Carlos/Ribeirão Preto Medical School/São Carlos Institute of Chemistry, University of São Paulo, São Carlos, Brazil. [4]Educational Union of Cascavel - University Center, Brazil. [5]Macarthur Physiotherapy, Sydney, Australia; [6]Physical Therapy Department, School of Science and Technology, São Paulo State University, Presidente Prudente, Brazil. The review protocol was registered with the International Prospective Register of Systematic Reviews (CRD42017060988). Mr Lopes is supported by the Brazilian National Council for Scientific and Technological Development program Science Without Borders and by the University of Sydney International Scholarships. The authors certify that they have no affiliations with or financial involvement in any organization or entity with a direct financial interest in the subject matter or materials discussed in the article. Address correspondence to Thiago Jambo Alves Lopes, The University of Sydney, Room S227, S Block, Faculty of Health Sciences, Cumberland Campus, 75 East Street, Lidcombe, NSW 2141 Australia. E-mail: tjam7956@uni.sydney.edu.au or thiagojambo@hotmail.com

However, recent advances in technology and affordability of high-speed, high-resolution cameras have made it possible to use 2-dimensional (2-D) video analysis to measure joint motion in athletic and clinical environments. For example, previous studies have shown an association between frontal plane projection angle (FPPA) and lower extremity kinematics during a single-leg squat task in persons with patellofemoral pain.[39] Recent evidence has shown that during a unilateral drop vertical jump task, a combination of increased knee valgus and ipsilateral trunk motion measured by 2-D video may identify female athletes at risk for noncontact knee injury.[3]

Before recommending its widespread use, though, it is important to establish the reliability and validity of 2-D video analysis for assessing frontal plane kinematics of the trunk and lower extremity during different athletic tasks, such as squatting, landing, and cutting.[8,9] Therefore, this systematic review sought to review and synthesize the existing literature regarding the reliability and validity of 2-D video to measure frontal plane kinematics of the trunk and lower extremity during various athletic tasks.

METHODS

Protocol and Registration

This systematic review followed the recommendations of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist.[24] The review protocol was registered with the International Prospective Register of Systematic Reviews (CRD42017060988).

Data Sources and Search Strategy

The authors conducted a systematic search of the databases MEDLINE, CINAHL, Embase, Scopus, and SPORTDiscus from inception to March 2017. The researchers used the following terms: *reliability/validity*, *athletic tasks (cutting, landing, squatting)*, and *2-D/3-D motion-capture systems* (**FIGURE 1 , APPENDIX D**). No restrictions were placed on the study type, population or language, or participant age and sex.

Study Selection

Titles and abstracts were independently screened by 2 blinded investigators. If at least 1 of the investigators included a paper during the abstract and title screening phase, then that study was included for the full-text screening. For full-text screening, the authors followed a previously described procedure, and any disagreement

A

Content 1		Content 2		Content 3		Content 4		Content 5
2D video		validity		inter-rater reliability		athletic tasks		limb alignment
OR		OR		OR		OR		OR
two dimensional video		validation		intra-rater reliability		cutting		kinematics
OR		OR		OR		OR		OR
3D video		concurrent validity		reliability		landing		lower extremity alignment
OR		OR		OR		OR		OR
three dimensional video		reproducibility of results		interrater agreement		squat		dynamic knee valgus
OR		OR		OR		OR		OR
markerless motion capture		validation studies		test-retest reliability		single leg squat		frontal plane knee motion
OR				OR		OR		OR
motion capture system				intra-rater agreement		functional tests		FPPA
OR				OR		OR		OR
real-time assessment				intraclass correlation coefficient		drop vertical jump		frontal plane projection angle
OR				OR		OR		OR
observational screening				intrasession reliability		single leg step down		knee abduction angle
OR				OR		OR		OR
motion analysis system				ICC		biomechanics		lower extremity kinematics
OR	AND		AND		AND	OR	AND	OR
movement analysis						drop-jump landings		joint angles
OR						OR		OR
screening tests						performance tests		knee control
OR						OR		OR
injury screenings						squatting		frontal plane angle
OR								OR
movement								knee angles
								OR
								hip angles
								OR
								trunk angles
								OR
								hip motion
								OR
								knee motion
								OR
								trunk motion

B

(TITLE-ABS-KEY ("2D video*" OR "Two dimensional video*" OR "3D video*" OR "three dimensional video*" OR "Markerless motion capture" OR "motion capture system" OR "real-time assessment" OR "Observational Screening" OR "motion analysis system") OR TITLE-ABS-KEY ("movement analysis" OR "screening tests" OR "injury screenings") AND TITLE-ABS-KEY (validity OR validation OR "concurrent validity" OR "Validation studies") AND TITLE-ABS- KEY ("athletic tasks*" OR cutting OR landing OR squat OR squatting OR "Single Leg Squat" OR "functional tests" OR "Drop Vertical Jump" OR "single leg step down" OR biomechanics OR "Drop-Jump Landings" OR movement OR "Performance tests") AND TITLE-ABS-KEY ("limb alignment" OR kinematics OR "Lower Extremity Alignment" OR "dynamic knee valgu" OR "frontal plane knee motion" OR fppa OR "frontal plane projection angle" OR "knee abduction angle" OR "lower extremity kinematics" OR "joint angles" OR "knee control") OR TITLE-ABS-KEY ("Frontal plane angle" OR "Knee angles" OR "Hip angles" OR "Trunk angles" OR "Hip motion" OR "knee motion" OR "Trunk motion"))

FIGURE 1. (A) Search strategy (content table), and (B) Scopus database search strategy. Abbreviations: FPPA, frontal plane projection angle; ICC, intraclass correlation coefficient.

between investigators over the eligibility of a study was discussed and resolved.[15] Finally, the investigators independently reviewed references of the selected full-text articles to determine whether additional studies should be assessed for inclusion.

Studies were included if they evaluated the reliability or validity of 2-D video measurements of trunk or lower extremity kinematics during any of the following tasks: cutting, landing, and squatting. The authors excluded studies that evaluated running or gait, used only 3-D motion-capture systems (including those that calculated 2-D angles from the 3-D motion-capture system), employed methods that were not based on kinematic angles (ie, Landing Error Scoring System), or did not report frontal plane kinematics. The investigators also excluded studies that were not full-text manuscripts (ie, abstracts, theses, dissertations), studies that did not assess 2-D and 3-D motion simultaneously, and studies that did not assess living human beings (ie, animals or human cadavers).

Risk-of-Bias Evaluation

Two pairs of independent investigators assessed the quality of the included studies, using the Quality Appraisal of Diagnostic Reliability (QAREL)[16,17] checklist for reliability studies and the Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2)[18,38] tool for validity studies. If there was any disagreement, a consensus was reached through discussion.

The QAREL tool, created in 2010, has been used to assess the methodological quality of reliability studies. It contains 11 items divided into the following domains: items 1 and 2, review of sampling bias and how representative participants and raters are; items 3 to 7, rater blinding; item 8, whether the test sequence varied; item 9, whether the time interval between measures was sufficient; item 10, whether the test was conducted and interpreted correctly; and item 11, the appropriateness of the statistical analyses. Furthermore, each item can be answered "yes," "no," "unclear," or "not applicable" (items 3, 4, 5, 6, and 8), where a "yes" response suggests a good-quality feature of the study and a "no" response suggests a poor-quality feature.

The QUADAS-2 tool is frequently used in systematic reviews to evaluate the accuracy of diagnostic tests.[38] This tool contains 4 domains (patient selection, index test, reference standard, and flow and timing) with signaling questions that can be answered as "yes," "no," or "unclear." The risk of bias is assessed for each domain, and the first 3 domains are also assessed for applicability concerns. In accordance with the QUADAS-2 instructions,[38] if all signaling questions within a specific domain are answered "yes," then the risk of bias should be graded as "low," but if any question is graded "no," then potential risk of bias exists. Finally, an "unclear" answer should be avoided and used only when insufficient data are provided.

Data Extraction

One investigator extracted the following data independently: study population, participant demographics and characteristics, details of the intervention (reliability or validity study), number of raters, tested leg, task performed, and relevant biomechanical variables. Two additional investigators double-checked these data for discrepancies and quality control.

For the meta-analysis, when the study reported results from either the left and right legs or both the dominant and nondominant legs, the investigators extracted data from the right or the dominant leg, as these were more commonly reported. In addition, when more than 1 type of landing task was reported by the same study (ie, cutting and jumping tasks), the researchers extracted the data most similar to those derived from a jumping task, as the latter was more commonly reported. For reliability studies, when the authors reported results of more than 1 rater, data from the first rater or the most experienced rater were used. Furthermore, when more than 2 test sessions were reported for reliability, the results comparing the first 2 sessions were used for analysis. Finally, the investigators extracted and presented data from the reliability studies based on the definitions provided by Portney and Watkins[31] as well as Koo and Li[13] for different types of reliability studies.

As for the studies evaluating 2-D video FPPA measurements, the researchers separated the reliability studies into intrarater reliability when the 2-D FPPA was measured by 1 rater across 2 or more different days, interrater reliability when 2 or more raters measured the same group of participants on the same trials, and intraday intrarater reliability when 1 rater assessed the same participant within the same test session or same day.

Data Analysis

Comprehensive Meta-Analysis Version 3.3.070 (Biostat, Inc, Englewood, NJ) software was used for the meta-analysis. For reliability, the investigators entered the mean intraclass correlation coefficient (ICC), the ICC lower limit (95% confidence interval [CI]), and sample size to calculate the pooled ICC.[5] The Pearson correlation coefficient strength (r) and sample size were used for validity. A random-effects model[33] was used and a heterogeneity value (I^2) of 50% or less was accepted, as higher values represent substantial heterogeneity, which might bias the study's conclusion.[10,33] Therefore, as per Schroll et al,[33] sensitivity analysis was performed in the presence of high heterogeneity (I^2 greater than 50%). The level of significance was set a priori at $\alpha = .05$.

The strength of the correlation coefficients (ICC and Pearson) was interpreted according to Koo and Li[13] and Portney and Watkins,[31] respectively. For the ICC interpretation, the following categories were used, considering the 95% CI as previously suggested[13]: less than 0.50 indicated poor reliability, 0.50 to 0.75 indicated moderate reliability, 0.75 to 0.90 indicated good reliability, and 0.90 or greater indicated excellent reliability.[13] A Pearson correlation coefficient strength

(r) of 0.00 to 0.25 was interpreted as little or no correlation, 0.25 to 0.50 as a fair relationship, 0.50 to 0.75 as a moderate to good relationship, and 0.75 or greater as a good to excellent relationship.[31]

Sensitivity Analysis

Reliability Studies For the single-leg squat task, 1 study[25] increased the heterogeneity (I^2 greater than 50%) for intrarater and intraday intrarater reliability; however, as the overall result did not change after removing this study, the authors retained it in the meta-analysis and reported both results (with and without the study).[33]

Validity Studies Of all studies that used a landing task, only 1 performed a unilateral landing task,[20] so the investigators performed a sensitivity analysis that included only bilateral tasks and then combined bilateral and unilateral landing tasks. As the overall result did not change, the authors combined both types of tasks for the meta-analysis.[33]

Two studies that used a landing task were assessed as having a high risk of bias, as they only included participants with specific characteristics[29] and excluded some participants from the statistical analysis.[2,29] Therefore, the investigators decided to run a sensitivity analysis excluding these studies to evaluate whether the results would change. As the overall result did not change, the authors retained both studies for the meta-analysis.[33]

RESULTS

Study Selection

A total of 16 studies[2,4,7,11,12,20-22,25,28,29,32,34-36,39] met the inclusion criteria and 13 were included in the quantitative analysis. Three studies[12,28,35] did not provide enough data in the papers or after contacting their authors. Eight studies[4,7,12,21,25,28,35,36] reported data on reliability of the 2-D video analysis only, 4 only reported data on the concurrent validity between the 2-D video and 3-D motion-capture system,[20,29,34,39] and 4 studies reported data on the reliability and concurrent validity[2,11,22,32] (**FIGURE 2**). For the title/abstract and full-text screening, the agreement rates between both raters were 94% (54 disagreements in 904 studies) and 85% (5 disagreements in 34 studies), respectively.

Study Characteristics

Across all studies that reported data on reliability, 11 presented data on intrarater reliability,[4,7,11,21,22,25,28,29,32,35,36] 6 on interrater reliability,[4,11,12,22,32,36] and 4 on intraday test-retest reliability,[12,25,32,35] including a total of 308 participants. Eight studies[2,11,20,22,29,32,34,39] that reported data on concurrent validity included 278 participants in total.

Overall, participants in the studies evaluated were predominantly female (8 of 16 studies), collegiate or elite athletes (7 of 16 studies), and had a mean age ranging from 15 to 24 years. Most studies (10 of 16) used a landing task (bilateral or unilateral) and almost all studies assessed the knee joint, with only 3 studies assessing different joints (ie, trunk, hip, and ankle joints). Characteristics of included studies are shown in **TABLE 1**.

Risk of Bias

Reliability Studies The agreement between both raters for risk of bias was 91% (12 disagreements in 132 questions). Overall, studies included participants considered representative of the population for which the results were intended to be used, respected the time interval between repeated measures, used and interpreted the test correctly, and performed the correct statistical analysis. The primary sources of bias were found in studies in which raters had not been blinded from prior analyses and other raters' assessments, and many studies did not vary the assessments' order (**APPENDIX A**, available at www.jospt.org).

Validity Studies The agreement between both raters for risk of bias was 91% (5 disagreements in 56 questions). Overall,

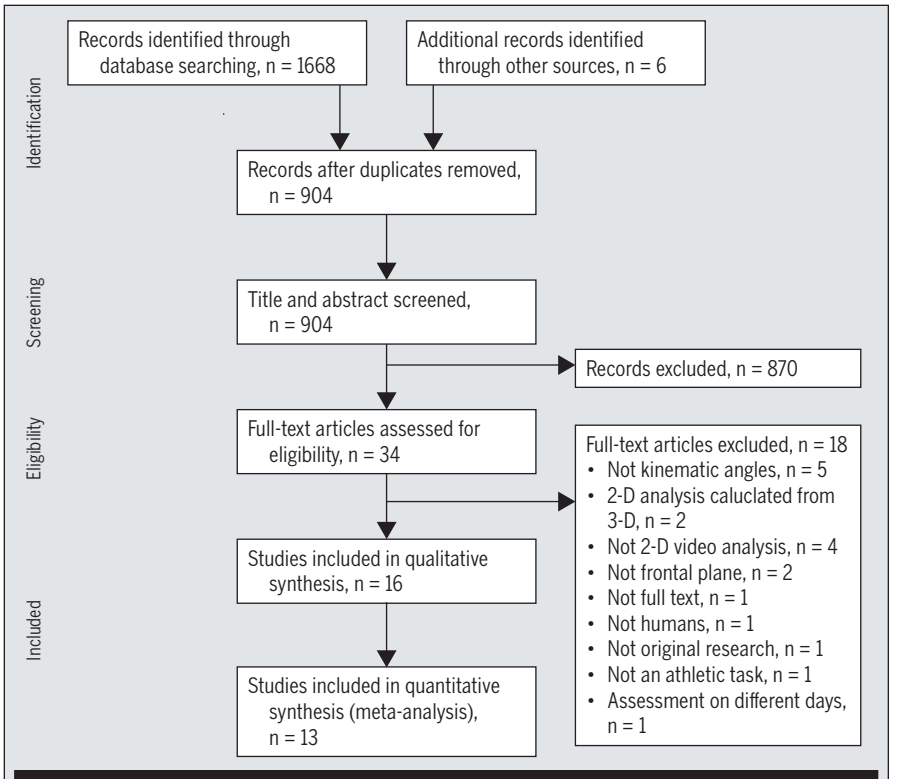


FIGURE 2. PRISMA flow diagram of study selection.

TABLE 1 Study Characteristics

Study	Participant Characteristics	Intrarater/ Interrater Reliability	Intraday Test-Retest Reliability	Raters, n	Tested Leg	Validity Study	Task/Body Region	2-D and 3-D Kinematic Variables Used as Outcome Measures
McLean et al[20]	Collegiate athletes, female (age, 21.2 ± 3.0 y; height, 176 ± 11.1 cm; weight, 76.1 ± 12.4 kg) and male (20.2 ± 1.9 y; 184.7 ± 8.0 cm; 81.9 ± 9.8 kg)	NA NA	NA	Unclear	Right	Yes (n = 20)	Side jump Knee	FPPA (2-D): peak frontal plane knee angle Dynamic knee valgus angle (3-D): combination of tibia and femur static alignment and dynamic motion Both measures were time normalized to 100% of stance phase
Nagano et al[29]	Female collegiate athletes (age, 21 ± 1 y; height, 1.66 ± 0.8 m; weight, 58.8 ± 7.7 kg)	NA NA	NA	Unclear	Dominant	Yes (n = 20)	Continuous jump test Knee	Knee valgus (2-D): frontal plane angle at maximum knee flexion minus knee frontal plane angle at standing position Knee abduction/adduction angle (3-D): knee angular displacements were defined as a variation from the position in the static standing trial
Willson and Davis[39]	Young women with and without PFP (age, 23.7 ± 3.6 y; height, 1.70 ± 0.06 m; weight, 61.1 ± 5.4 kg)	NA NA	NA	1 (not defined)	Random	Yes (n = 40)	SL squat Knee	FPPA (2-D): frontal plane angle at maximum knee flexion minus knee frontal plane angle at standing position Knee abduction angle (3-D): calculation not clearly defined
Stensrud et al[35]	Female handball players (age, 22 ± 4 y; height, 173 ± 7 cm; weight, 70 ± 8 kg)	Yes (n = 20) NA	Yes (n = 18)	1 (not defined)	Right and left	NA	SL squat, SL DVJ, and DVJ Knee	FPPA (2-D): knee frontal plane angle at maximum flexion subtracted from 180°
Miller and Callister[21]	Personal training students, female and male (age, 23.7 y; range, 21.2-26.3 y)	Yes (n = 48) NA	NA	1 inexperienced tester	Right and left	NA	SL DVJ Knee	Knee frontal plane (2-D): ankle, knee, and quadriceps angle, formed by the distal anterior tibial margin and the mid-patella and mid-thigh markers at maximum knee flexion
Munro et al[25]	Recreationally active adults, female (age, 21.5 ± 2.3 y; height, 170.1 ± 6.1 cm; weight, 66.2 ± 10.2 kg) and male (22.6 ± 3.1 y; 177.9 ± 6.0 cm; 75.8 ± 7.9 kg)	Yes (n = 10); yes (n = 10)	Yes (n = 10); yes (n = 10)	1 (not defined)	Average of left and right	NA	SL squat, DVJ, and SL landing Knee	FPPA (2-D): the angle formed between the line from the markers on the proximal thigh to the knee joint and the line from the knee joint to the ankle at the point of maximum knee flexion
Mizner et al[22]	Female collegiate athletes (age, 19.6 ± 1.2 y; BMI, 25 ± 2 kg/m^2)	Yes (n = 36) Yes (n = 36)	NA	2 experienced raters	Dominant	Yes (n = 36)	DVJ Knee	FPPA (2-D): knee frontal plane angle at the point of peak knee flexion during the initial landing phase (prior to starting the jump phase) Knee abduction angle (3-D): knee abduction angle at peak knee flexion during landing
Harris-Hayes et al[7]	Undergraduate and graduate students, female and male (age, 19.3 ± 4.5 y; BMI, 23.8 ± 3.6 kg/m^2)	Yes (n = 15) NA	NA	1 research assistant	Unclear	NA	SL squat Knee	FPPA (2-D): knee frontal plane angle at maximum knee flexion minus frontal plane angle at the standing position

Table continues on page 817.

TABLE 1 Study Characteristics (continued)

Study	Participant Characteristics	Intrarater/ Interrater Reliability	Intraday Test-Retest Reliability	Raters, n	Tested Leg	Validity Study	Task/Body Region	2-D and 3-D Kinematic Variables Used as Outcome Measures
Jones et al[11]	Female collegiate athletes (age, 19.4 ± 2.4 y; height, 170.3 ± 9.2 cm; weight, 65.8 ± 8.9 kg)	Yes (n = 35) Yes (n = 35)	NA	2 experienced physical therapists	Random	Yes (n = 81)	Step-down squat task Knee	FPPA (2-D): knee frontal plane angle at maximum flexion subtracted from 180° Knee abduction angle (3-D): knee abduction angle at peak knee flexion
King and Belyea[12]	College students, female and male (age, 21 ± 1 y; height, 1.73 ± 0.11 m; weight, 72.3 ± 13.8 kg)	NA Yes (n = 23)	Yes (n = 23)	2 novice raters and 2 expert raters	Right	NA	DVJ Knee	FPPA (2-D): calculated by subtracting initial contact values from maximum knee flexion values
Myer et al[28]	Female high school volleyball athletes (age, 15.27 ± 1.0 y; height, 1.69 ± 0.42 m; weight, 61.08 ± 7.9 kg)	Yes (n = 19) NA	NA	3 (not defined)	Right and left	NA	DVJ Knee	FPPA (2-D): maximum frontal plane knee angle was calculated from hip, knee, and ankle joint centers at maximum medial knee displacement
Dingenen et al[4]	Elite female athletes (age, 21.1 ± 3.4 y; height, 170.0 ± 8.3 cm; weight, 65.2 ± 8.0 kg)	Yes (n = 15) Yes (n = 15)	NA	2 (not defined)	Dominant and nondominant	NA	SL squat and SL DVJ Trunk	LTM (2-D): the angle between a vertical line starting at the ipsilateral ASIS and the line between the ipsilateral ASIS and the manubrium sterni. Smaller LTM angles represent more LTM in the direction of the supporting leg. The LTM angle was negative when the manubrium sterni was more lateral than the ipsilateral ASIS
Belyea et al[2]	Healthy subjects, female and male (age, 21 ± 1.4 y; height, 1.73 ± 0.12 m; weight, 71 ± 13 kg)	Yes (n = 21) NA	NA	1 student researcher	Random	Yes (n = 19)	DVJ Knee	FPPA (2-D): calculated by subtracting initial contact values from maximum knee flexion values Knee abduction ROM (3-D): knee frontal plane angle at peak knee flexion minus frontal plane angle at initial contact
Tate et al[36]	College students (sex not specified)	Yes (n = 20) Yes (n = 20)	NA	1 novice rater and 1 expert rater	Dominant	NA	SL squat Knee	FPPA (2-D): measured as the angle formed between a line drawn from the middle of the ankle joint to the middle of the knee joint and a line from the middle of the knee joint extended to bisect the thigh at maximum knee flexion
Schurr et al[34]	Physically active adults, female and male (age, 22.26 ± 2.99 y; height, 1.70 ± 0.12 m; weight, 67.43 ± 12.24 kg)	NA NA	NA	1 clinician	Average of left and right	Yes (n = 26)	SL squat Ankle, knee, hip, and trunk	FPPA (2-D): joint displacement at the trunk, hip, knee, and ankle in the frontal plane. Frontal plane angle calculated at maximum knee flexion minus frontal plane angle at the standing position Knee abduction ROM (3-D): joint frontal plane angle at peak knee flexion minus frontal plane angle at the standing position

Table continues on page 818.

TABLE 1 Study Characteristics (continued)

Study	Participant Characteristics	Intrarater/ Interrater Reliability	Intraday Test-Retest Reliability	Raters, n	Tested Leg	Validity Study	Task/Body Region	2-D and 3-D Kinematic Variables Used as Outcome Measures
Scholtes and Salsich[32]	Young women with and without PFP (age, 22.4 ± 4.3 y; BMI, 22.4 ± 3.2 kg/m^2)	Yes (n = 15) Yes (n = 10)	Yes (n = 36)	2 experienced physical therapists	Symptomatic leg (PFP participants) and random for healthy participants	Yes (n = 36)	SL squat Knee and hip	FPPA (2-D): hip and knee angles were obtained at peak knee flexion. The knee FPPA was calculated as 180° minus the angle between the thigh segment and the shank segment Knee abduction angle (3-D): knee frontal plane angle at peak knee flexion

Abbreviations: ASIS, anterior superior iliac spine; BMI, body mass index; DVJ, drop vertical jump; FPPA, frontal plane projection angle; LTM, lateral trunk motion; NA, not available; PFP, patellofemoral pain; ROM, range of motion; SL, single leg.

TABLE 2 Results of the QUADAS-2 Quality Assessment

	Risk of Bias				Applicability Concerns		
Study	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
McLean et al[20]	Low	Low	Low	Low	Low	Low	Low
Nagano et al[29]	High	Low	Low	High	Low	Low	Low
Willson and Davis[39]	Low	Low	Low	Low	Low	Low	Low
Mizner et al[22]	Low	Low	Low	Low	Low	Low	Low
Jones et al[11]	Low	Low	Low	Low	Low	Low	Low
Belyea et al[2]	Low	Low	Low	High	Low	Low	Low
Schurr et al[34]	Low	Low	Low	Low	Low	Low	Low
Scholtes and Salsich[32]	Low	Low	Low	Low	Low	Low	Low

Abbreviation: QUADAS, Quality Assessment of Diagnostic Accuracy Studies.

most studies presented no risk of bias for the patient selection domain. One study[29] excluded 8 participants who presented with knee varus displacement at baseline, as the study's purpose was to assess participants with knee valgus displacement only. On the other hand, there was no risk of bias for the index test and reference standard domains, as both devices (index and reference standard) were tested simultaneously. This prevented raters from having access to the reference standard's results. However, for the flow and timing domain, 2 studies[2,29] were graded as high risk of bias, as they excluded participants from the statistical analysis. Finally, for the applicability concerns domain, all studies were graded as low risk of bias (**TABLE 2**).

2-D Video Reliability

Single-Leg Squat Intrarater Reliability Regarding knee FPPA reliability, meta-analysis revealed that there was good to excellent intrarater reliability (**FIGURE 3**), and excellent reliability (ICC = 0.99; 95% CI: 0.97, 1.00) when the Munro et al[25] study was excluded after sensitivity analysis (I^2 = 0%). Studies that were not included in the meta-analysis due to lack of data reported mean ICC values ranging from 0.92[35] to 0.99.[11]

Scholtes and Salsich[32] reported excellent reliability for hip FPPA (ICC = 0.99; 95% CI: 0.97, 1.00), while Dingenen et al[4] reported excellent reliability (ICC = 0.99; 95% CI: 0.95, 1.00) for lateral trunk motion (LTM).

Single-Leg Squat Interrater Reliability Meta-analysis revealed that there was excellent interrater reliability for the knee FPPA (ICC = 0.97; 95% CI: 0.92, 1.00) (**APPENDIX B**, available at www.jospt.org). One study that was not included in the meta-analysis due to lack of data reported similar results (mean ICC = 0.94).[11]

For hip FPPA, Scholtes and Salsich[32] reported good to excellent reliability (ICC = 0.97; 95% CI: 0.88, 0.99), while Dingenen et al[4] reported good to excellent reliability for LTM (ICC = 0.98; 95% CI: 0.86, 1.00).

Single-Leg Squat Intraday Intrarater Reliability For knee FPPA, meta-analysis revealed that there was moderate to excellent intraday intrarater reliability

(ICC = 0.74; 95% CI: 0.57, 0.90; I^2 = 73%) (**APPENDIX C**, available at www.jospt.org). When male participants were excluded from Munro et al[25] based on the sensitivity analysis, moderate to good reliability was observed (ICC = 0.66; 95% CI: 0.53, 0.79; I^2 = 0%). One study that was not included in the meta-analysis due to lack of data reported moderate intraday intrarater reliability (mean ICC = 0.57).[35] Last, Scholtes and Salsich[32] reported moderate to excellent reliability (ICC = 0.83; 95% CI: 0.73, 0.91) for hip FPPA.

Landing Task Intrarater Reliability For knee FPPA, meta-analysis revealed that there was moderate to good reliability for unilateral landing tasks and good to excellent reliability for bilateral landing tasks (**FIGURE 3**). Studies that were not included in the meta-analysis because of lack of data reported a knee FPPA mean ICC of 0.89[35] (unilateral landing task) and 0.95[22,35] (bilateral landing task). Dingenen et al[4] reported excellent reliability (ICC = 0.99; 95% CI: 0.98, 1.00) for LTM.

Landing Task Interrater Reliability Because of a lack of data, it was not possible to run a meta-analysis for the bilateral landing task; however, 2 studies reported good and excellent knee FPPA reliability (mean ICC = 0.89[22]-0.92[12]). For LTM, Dingenen et al[4] reported excellent reliability (ICC = 0.98; 95% CI: 0.94, 0.99).

Landing Task Intraday Intrarater Reliability Knee FPPA reliability for unilateral landing tasks was moderate to good (ICC = 0.77; 95% CI: 0.68, 0.87), and for bilateral tasks was good to excellent (ICC = 0.86; 95% CI: 0.81, 0.92) (**APPENDIX C**). Studies that were not included in the meta-analysis due to lack of data reported moderate reliability for the unilateral landing task (mean ICC = 0.58)[35] and good to excellent reliability (mean ICC = 0.89[12,35] to 0.91[28]) for the bilateral landing task.

2-D Video Validity

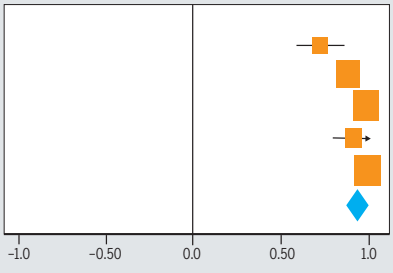
Single-Leg Squat Validity There was no correlation between 2-D knee FPPA and 3-D knee frontal plane angle (**FIGURE 4**). Importantly, the Bland-Altman plots provided by Schurr et al[34] revealed poor agreement between the 2-D video and 3-D motion-capture system, as evident by the wider limits of agreement (LoA) for all frontal plane movements, such as the trunk (7.92°; 95% LoA: −6.65°, 22.50°), hip (−8.72°; 95% LoA: −21.90°, 4.45°), knee (−6.62°; 95% LoA: −29.83°, 16.59°), and ankle (3.03°; 95% LoA: −7.96°, 14.02°).

Schurr et al[34] also reported no correlation between both methods for the trunk, knee, and ankle 2-D frontal plane angles. Conflicting results were found for 2-D frontal plane hip angles, as Scholtes and Salsich[32] reported a good to excellent relationship between 2-D and 3-D systems (r = 0.825, P<.001), while Schurr et al[34] reported no relationship (r = 0.150, P = .469). Due to high heterogeneity (I^2 = 93%) and lack of studies, it was not possible to run a meta-analysis for the 2-D hip frontal plane angle.

Landing Task Validity The meta-analysis revealed that there was a moderate to good relationship between the 2-D

Intrarater Reliability: Single-Leg Squat

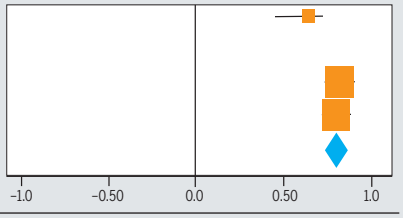
Study	n	ICC (95% CI)	*P* Value	ICC (95% CI)
Munro et al[25]				
Female	10	0.720 (0.581, 0.859)	<.001	
Male	10	0.880 (0.828, 0.932)	<.001	
Harris-Hayes et al[7]	15	0.980 (0.953, 1.000)	<.001	
Tate et al[36]	20	0.910 (0.788, 1.000)	<.001	
Scholtes and Salsich[32]	15	0.990 (0.974, 1.000)	<.001	
Total*	70	0.926 (0.873, 0.980)	<.001	



Abbreviations: CI, confidence interval; ICC, intraclass correlation coefficient.
**Heterogeneity: τ^2 = 0.003, Q = 30.185, df (Q) = 4 (P<.001), I^2 = 87%. Test for overall effect: z = 34.024 (P<.001).*

Intrarater Reliability: Unilateral Landing Task

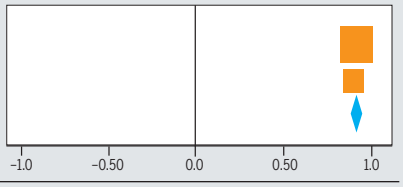
Study	n	ICC (95% CI)	*P* Value	ICC (95% CI)
Miller and Callister[21]	48	0.642 (0.448, 0.836)	<.001	
Munro et al[25]				
Female	10	0.820 (0.733, 0.907)	<.001	
Male	10	0.800 (0.713, 0.887)	<.001	
Total*	68	0.795 (0.736, 0.853)	<.001	



Abbreviations: CI, confidence interval; ICC, intraclass correlation coefficient.
**Heterogeneity: τ^2 = 0.001, Q = 2.725, df (Q) = 2 (P = .256), I^2 = 27%. Test for overall effect: z = 21.525 (P<.001).*

Intrarater Reliability: Bilateral Landing Task

Study	n	ICC (95% CI)	*P* Value	ICC (95% CI)
Munro et al[25]				
Female	10	0.910 (0.875, 0.945)	<.001	
Male	10	0.890 (0.838, 0.942)	<.001	
Total*	20	0.904 (0.875, 0.933)	<.001	



Abbreviations: CI, confidence interval; ICC, intraclass correlation coefficient.
**Heterogeneity: τ^2 = 0.000, Q = 0.394, df (Q) = 1 (P = .530), I^2 = 0%. Test for overall effect: z = 61.434 (P<.001).*

FIGURE 3. Forest plots describing the knee frontal plane projection angle intrarater reliability.

knee FPPA and 3-D knee frontal plane angle (**FIGURE 4**). As previously reported, a moderate to good relationship also was found when the only study that used a unilateral task[20] was excluded (r = 0.53, P = .001, I^2 = 14%) and when both studies that presented high risk of bias were removed from the analysis (r = 0.62, P = .04, I^2 = 82%). On the other hand, the Bland-Altman plots provided by Belyea et al[2] revealed poor agreement between the 2-D and 3-D methods for the knee FPPA measurements (approximately 6.0°; approximate 95% LoA: −29°, 17°).

DISCUSSION

The main findings of the current meta-analysis were as follows: (1) the reliability of 2-D video analysis varies from moderate to excellent, depending on the reliability type and task performed; and (2) measurement of frontal plane angles using 2-D video is not comparable to 3-D measures.

2-D Video Reliability

For a device or measurement to be considered valid, it must first be reliable.[31] Although there is no consensus on what constitutes good reliability, as this is context dependent, it is suggested that an ICC higher than 0.90 is desirable when a diagnosis or a clinical decision is the main purpose.[31] In the current study, for almost all tasks and joints, intrarater and interrater reliability for 2-D frontal plane angles ranged from good to excellent, with most results presenting a mean ICC higher than 0.90.

Considering that the 2-D kinematic assessment is rater dependent, these results are important and provide a certain degree of confidence that, regardless of rater or knee, hip, and trunk frontal plane angle, reliability will be within acceptable values. Although the knee FPPA intrarater reliability (ICC = 0.79; 95% CI: 0.72, 0.86) and intraday intrarater reliability (mean ICC = 0.74-0.86) for the unilateral task ranged from moderate to good, these results are still within the acceptable values for clinical practice.[37]

Despite these acceptable ICC values, however, it is important to highlight the possibility of bias and overestimation related to these results, as almost none of the studies reported whether raters were blinded from their own or other raters' scores, and it is unclear whether assessment order varied in most studies. Therefore, the authors recommend that future studies on this topic follow the QAREL[16,17] checklist and use the Guidelines for Reporting Reliability and Agreement Studies[14] for reporting purposes.

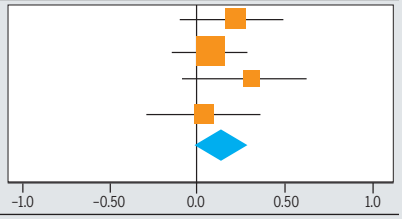
2-D Video Validity

Of the 8 studies that compared 2-D video with the 3-D motion-capture system, only 2 studies[2,34] reported the level of agreement in Bland-Altman plots. These studies reported, regardless of body region, poor agreement between the 2 methods and a clinically unacceptable LoA. Although there is no consensus regarding an acceptable LoA for lower extremity frontal plane angles, a previous study evaluating low back motion suggested that an LoA of less than 5° in either direction (−5° to +5°) would be clinically acceptable.[23] Therefore, based on the current evidence, the findings of this systematic review concur with previous suggestions that 2-D video may not be a valid assessment of lower extremity frontal plane motion.[1,2,20,39]

The poor agreement between 2-D and 3-D methods may be explained by the fact that the 2-D video system is not able to capture the rotational movements (transverse plane) that occur concomitantly with frontal plane movements during some athletic tasks, such as the single-leg squat and drop jump tests.[1,39] For instance, knee valgus is the combined result of hip adduction, hip internal rotation, and ankle eversion.[8] Previous research has shown that knee FPPA during a single-leg squat is greatly affected by hip rotation angle, maybe even to a greater extent than 3-D knee valgus/varus angle.[1] Recent studies have

Single-Leg Squat Task

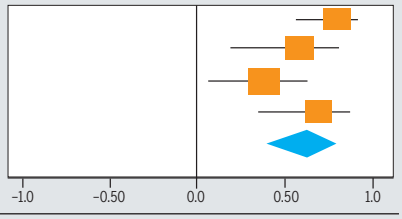
Study	n	Correlation	*P* Value	Correlation (95% CI)
Willson and Davis[39]	40	0.210	.195	
Jones et al[11]	81	0.070	.536	
Schurr et al[34]	26	0.310	.124	
Scholtes and Salsich[32]	36	0.036	.836	
Total*	183	0.127	.094	



Abbreviation: CI, confidence interval.
**Heterogeneity: τ^2 = 0.000, Q = 1.662, df (Q) = 3 (P = .645), I^2 = 0%. Test for overall effect: z = 1.676 (P = .094).*

Landing Task

Study	n	Correlation	*P* Value	Correlation (95% CI)
McLean et al[20]	20	0.800	<.001	
Nagano et al[29]	20	0.580	.006	
Mizner et al[22]	36	0.381	.021	
Belyea et al[2]	19	0.690	.001	
Total*	95	0.619	<.001	



Abbreviation: CI, confidence interval.
**Heterogeneity: τ^2 = 0.050, Q = 5.999, df (Q) = 3 (P = .112), I^2 = 50%. Test for overall effect: z = 4.554 (P<.001).*

FIGURE 4. Forest plots describing the correlation between the knee frontal plane projection angle and 3-D knee frontal plane angle.

tried to combine the 2-D hip and knee FPPA measures to compare to 3-D motion-capture measures[32]; however, more studies are needed to confirm that this approach provides meaningful results.

Although the 2-D video system does not appear to be a valid method to assess frontal plane kinematics for the trunk and lower extremity joints during athletic tasks, especially for the knee joint, this method may have the potential to correctly identify people with excessive knee valgus displacement.[20] In the current study, the meta-analysis revealed a moderate to good relationship (r = 0.619, P< .001) between the 2-D video (knee FPPA) and 3-D motion-capture system (knee frontal plane angle) when a landing task was assessed. However, there was no relationship between the 2 methods during the single-leg squat (r = 0.127, P = .094).

Based on these findings, the investigators speculate that as landing tasks become more challenging, greater knee displacement makes this motion more detectable. For the less demanding squatting task, knee displacement values may be more closely clustered and harder to detect on 2-D video, which may have implications for validity findings. It has been suggested that the accuracy of the 2-D video system to measure knee frontal plane motion is task and participant dependent.[20,29] For instance, 2-D knee FPPA may have better association with 3-D knee abduction angle for those athletes who land with excessive knee valgus (ie, greater than 15°).[29]

In the clinical or athletic setting, where the goal may be to identify athletes at high risk of injury (eg, with excessive knee valgus) and measurement accuracy is less important, 2-D video analysis may be useful.

Strengths and Limitations

This study has several strengths, in that the authors followed a systematic approach and a strict protocol controlling all steps of the systematic review process. Moreover, the investigators are confident that there is a low chance of missing studies reporting the reliability and validity of the 2-D FPPA for measuring trunk and lower extremity frontal plane angles.

The current study also has some limitations. For instance, most studies did not report the 95% CI for the ICC, and although the researchers contacted the respective authors of these studies twice, they only received a reply from 1 author. Therefore, the authors of the present study were not able to include all studies in the reliability meta-analysis. In addition, most studies that assessed reliability did not report important methodological considerations to minimize bias, such as blinding raters from their own results or varying the order of their assessments.

CONCLUSION

The current systematic review revealed that reliable measurements of frontal plane trunk and lower extremity kinematics can be obtained using 2-D video analysis; however, reliability varies depending on the task and type of reliability. Conversely, the current evidence does not support the use of 2-D video analysis for measuring trunk and lower extremity frontal plane angles when accurate (valid) measures are required. ◉

KEY POINTS

FINDINGS: Two-dimensional (2-D) measures of trunk and lower extremity frontal plane kinematics appear to be reliable, but their reliability varies depending on the task and reliability type. On the other hand, 2-D video is not valid for measuring knee FPPA during the single-leg squat, but may be useful for landing tasks.

IMPLICATIONS: Clinicians and researchers should be aware that 2-D video analysis is not recommended when accurate trunk and lower extremity frontal plane kinematics are required. Moreover, there is poor evidence to support the use of 2-D video to measure knee FPPA, especially during the single-leg squat task.

CAUTION: Most studies that assessed reliability did not blind raters from their own results or vary the assessment order as recommended. As such, these results should be viewed with caution.

REFERENCES

1. Ageberg E, Bennell KL, Hunt MA, Simic M, Roos EM, Creaby MW. Validity and inter-rater reliability of medio-lateral knee motion observed during a single-limb mini squat. *BMC Musculoskelet Disord*. 2010;11:265. https://doi.org/10.1186/1471-2474-11-265
2. Belyea BC, Lewis E, Gabor Z, Jackson J, King DL. Validity and intrarater reliability of 2-dimensional motion analysis using a handheld tablet compared to traditional 3-dimensional motion analysis. *J Sport Rehabil*. 2015;24:technical report 14. https://doi.org/10.1123/jsr.2014-0194
3. Dingenen B, Malfait B, Nijs S, et al. Can two-dimensional video analysis during single-leg drop vertical jumps help identify non-contact knee injury risk? A one-year prospective study. *Clin Biomech (Bristol, Avon)*. 2015;30:781-787. https://doi.org/10.1016/j.clinbiomech.2015.06.013
4. Dingenen B, Malfait B, Vanrenterghem J, Verschueren SM, Staes FF. The reliability and validity of the measurement of lateral trunk motion in two-dimensional video analysis during unipodal functional screening tests in elite female athletes. *Phys Ther Sport*. 2014;15:117-123. https://doi.org/10.1016/j.ptsp.2013.05.001
5. Duffy L, Gajree S, Langhorne P, Stott DJ, Quinn TJ. Reliability (inter-rater agreement) of the Barthel Index for assessment of stroke survivors: systematic review and meta-analysis. *Stroke*. 2013;44:462-468. https://doi.org/10.1161/STROKEAHA.112.678615
6. Ford KR, Myer GD, Hewett TE. Reliability of landing 3D motion analysis: implications for longitudinal analyses. *Med Sci Sports Exerc*. 2007;39:2021-2028. https://doi.org/10.1249/mss.0b013e318149332d
7. Harris-Hayes M, Steger-May K, Koh C, Royer NK, Graci V, Salsich GB. Classification of lower extremity movement patterns based on visual assessment: reliability and correlation with 2-dimensional video analysis. *J Athl Train*. 2014;49:304-310. https://doi.org/10.4085/1062-6050-49.2.21
8. Hewett TE, Ford KR, Hoogenboom BJ, Myer GD. Understanding and preventing ACL injuries: current biomechanical and epidemiologic considerations - update 2010. *N Am J Sports Phys Ther*. 2010;5:234-251.
9. Hewett TE, Myer GD, Ford KR, et al. Biomechanical measures of neuromuscular control and valgus loading of the knee predict anterior cruciate ligament injury risk in female athletes: a prospective study. *Am J Sports Med*. 2005;33:492-501. https://doi.org/10.1177/0363546504269591
10. Higgins JPT, Green S. *Cochrane Handbook for*

Systematic Reviews of Interventions. Chichester, UK: Wiley-Blackwell; 2014.
11. Jones D, Tillman SM, Tofte K, et al. Observational ratings of frontal plane knee position are related to the frontal plane projection angle but not the knee abduction angle during a step-down task. *J Orthop Sports Phys Ther*. 2014;44:973-978. https://doi.org/10.2519/jospt.2014.4960
12. King DL, Belyea BC. Reliability of using a handheld tablet and application to measure lower-extremity alignment angles. *J Sport Rehabil*. 2015;24:technical report 15. https://doi.org/10.1123/jsr.2014-0195
13. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15:155-163. https://doi.org/10.1016/j.jcm.2016.02.012
14. Kottner J, Audigé L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol*. 2011;64:96-106. https://doi.org/10.1016/j.jclinepi.2010.03.002
15. Lopes TJA, Simic M, Myer GD, Ford KR, Hewett TE, Pappas E. The effects of injury prevention programs on the biomechanics of landing tasks: a systematic review with meta-analysis. *Am J Sports Med*. 2018;46:1492-1499. https://doi.org/10.1177/0363546517716930
16. Lucas N, Macaskill P, Irwig L, et al. The reliability of a quality appraisal tool for studies of diagnostic reliability (QAREL). *BMC Med Res Methodol*. 2013;13:111. https://doi.org/10.1186/1471-2288-13-111
17. Lucas NP, Macaskill P, Irwig L, Bogduk N. The development of a quality appraisal tool for studies of diagnostic reliability (QAREL). *J Clin Epidemiol*. 2010;63:854-861. https://doi.org/10.1016/j.jclinepi.2009.10.002
18. Maclachlan L, White SG, Reid D. Observer rating versus three-dimensional motion analysis of lower extremity kinematics during functional screening tests: a systematic review. *Int J Sports Phys Ther*. 2015;10:482-492.
19. Maletsky LP, Sun J, Morton NA. Accuracy of an optical active-marker system to track the relative motion of rigid bodies. *J Biomech*. 2007;40:682-685. https://doi.org/10.1016/j.jbiomech.2006.01.017
20. McLean SG, Walker K, Ford KR, Myer GD, Hewett TE, van den Bogert AJ. Evaluation of a two dimensional analysis method as a screening and evaluation tool for anterior cruciate ligament injury. *Br J Sports Med*. 2005;39:355-362. https://doi.org/10.1136/bjsm.2005.018598
21. Miller A, Callister R. Reliable lower limb musculoskeletal profiling using easily operated, portable equipment. *Phys Ther Sport*. 2009;10:30-37. https://doi.org/10.1016/j.ptsp.2008.10.003
22. Mizner RL, Chmielewski TL, Toepke JJ, Tofte KB. Comparison of 2-dimensional measurement techniques for predicting knee angle and moment during a drop vertical jump. *Clin J Sport Med*. 2012;22:221-227. https://doi.org/10.1097/JSM.0b013e31823a46ce
23. Mjøsund HL, Boyle E, Kjaer P, Mieritz RM, Skallgård T, Kent P. Clinically acceptable agreement between the ViMove wireless motion sensor system and the Vicon motion capture system when measuring lumbar region inclination motion in the sagittal and coronal planes. *BMC Musculoskelet Disord*. 2017;18:124. https://doi.org/10.1186/s12891-017-1489-1
24. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA statement. *PLoS Med*. 2009;6:e1000097. https://doi.org/10.1371/journal.pmed.1000097
25. Munro A, Herrington L, Carolan M. Reliability of 2-dimensional video assessment of frontal-plane dynamic knee valgus during common athletic screening tasks. *J Sport Rehabil*. 2012;21:7-11. https://doi.org/10.1123/jsr.21.1.7
26. Myer GD, Ford KR, Barber Foss KD, et al. The incidence and potential pathomechanics of patellofemoral pain in female athletes. *Clin Biomech (Bristol, Avon)*. 2010;25:700-707. https://doi.org/10.1016/j.clinbiomech.2010.04.001
27. Myer GD, Ford KR, Khoury J, Succop P, Hewett TE. Clinical correlates to laboratory measures for use in non-contact anterior cruciate ligament injury risk prediction algorithm. *Clin Biomech (Bristol, Avon)*. 2010;25:693-699. https://doi.org/10.1016/j.clinbiomech.2010.04.016
28. Myer GD, Wordeman SC, Sugimoto D, et al. Consistency of clinical biomechanical measures between three different institutions: implications for multi-center biomechanical and epidemiological research. *Int J Sports Phys Ther*. 2014;9:289-301.
29. Nagano Y, Sakagami M, Ida H, Akai M, Fukubayashi T. Statistical modelling of knee valgus during a continuous jump test. *Sports Biomech*. 2008;7:342-350. https://doi.org/10.1080/14763140802233223
30. Noehren B, Scholz J, Davis I. The effect of real-time gait retraining on hip kinematics, pain and function in subjects with patellofemoral pain syndrome. *Br J Sports Med*. 2011;45:691-696. https://doi.org/10.1136/bjsm.2009.069112
31. Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice*. 3rd ed. Upper Saddle River, NJ: Pearson/Prentice Hall; 2008.
32. Scholtes SA, Salsich GB. A dynamic valgus index that combines hip and knee angles: assessment of utility in females with patellofemoral pain. *Int J Sports Phys Ther*. 2017;12:333-340.
33. Schroll JB, Moustgaard R, Gøtzsche PC. Dealing with substantial heterogeneity in Cochrane reviews. Cross-sectional study. *BMC Med Res Methodol*. 2011;11:22. https://doi.org/10.1186/1471-2288-11-22
34. Schurr SA, Marshall AN, Resch JE, Saliba SA. Two-dimensional video analysis is comparable to 3D motion capture in lower extremity movement assessment. *Int J Sports Phys Ther*. 2017;12:163-172.
35. Stensrud S, Myklebust G, Kristianslund E, Bahr R, Krosshaug T. Correlation between two-dimensional video analysis and subjective assessment in evaluating knee control among elite female team handball players. *Br J Sports Med*. 2011;45:589-595. https://doi.org/10.1136/bjsm.2010.078287
36. Tate J, True H, Dale B, Baker C. Expert versus novice interrater and intrarater reliability of the frontal plane projection angle during a single-leg squat. *Int J Athl Ther Train*. 2015;20:23-27. https://doi.org/10.1123/ijatt.2014-0116
37. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Stat Med*. 1998;17:101-110.
38. 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. https://doi.org/10.7326/0003-4819-155-8-201110180-00009
39. Willson JD, Davis IS. Utility of the frontal plane projection angle in females with patellofemoral pain. *J Orthop Sports Phys Ther*. 2008;38:606-615. https://doi.org/10.2519/jospt.2008.2706

APPENDIX A

QUALITY APPRAISAL OF DIAGNOSTIC RELIABILITY (QAREL CHECKLIST*)

	Item†										
Study	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Miller and Callister[21]	Y	N	NA	N	NA	NA	U	N	Y	Y	Y
Stensrud et al[35]	Y	U	NA	N	NA	NA	U	N	Y	Y	Y
Jones et al[11]	Y	Y	N	N	NA	NA	U	N	Y	Y	Y
Mizner et al[22]	Y	Y	N	N	NA	NA	U	N	Y	Y	Y
Munro et al[25]	Y	U	NA	N	NA	NA	U	N	Y	Y	Y
Myer et al[28]	N	U	NA	N	NA	NA	U	N	Y	Y	Y
Harris-Hayes et al[7]	Y	Y	NA	N	NA	NA	U	N	Y	Y	Y
Dingenen et al[4]	Y	U	Y	N	NA	NA	U	N	Y	Y	Y
Belyea et al[2]	Y	Y	NA	Y	NA	NA	U	N	U	Y	Y
King and Belyea[12]	Y	Y	N	N	NA	NA	U	N	Y	Y	Y
Tate et al[36]	Y	Y	N	N	NA	NA	U	Y	Y	Y	Y
Scholtes and Salsich[32]	N	Y	N	N	NA	NA	U	N	Y	Y	Y

Abbreviations: N, no; NA, not applicable; QAREL, Quality Appraisal of Diagnostic Reliability; U, unclear; Y, yes.

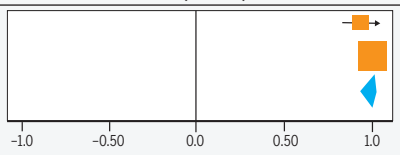
**Used with permission from Lucas et al.[17]*

†1, Was the test evaluated in a sample of subjects who were representative of those to whom the authors intended the results to be applied? 2, Was the test performed by raters who were representative of those to whom the authors intended the results to be applied? 3, Were raters blinded to the findings of other raters during the study? 4, Were raters blinded to their own prior findings of the test under evaluation? 5, Were raters blinded to the results of the accepted reference standard or disease status for the target disorder (or variable) being evaluated? 6, Were raters blinded to clinical information that was not intended to be provided as part of the testing procedure or study design? 7, Were raters blinded to additional cues that were not part of the test? 8, Was the order of examination varied? 9, Was the stability (or theoretical stability) of the variable being measured considered when determining the suitability of the time interval between repeated measures? 10, Was the test applied correctly and interpreted appropriately? 11, Were appropriate statistical measures of agreement used?

APPENDIX B

INTERRATER RELIABILITY: SINGLE-LEG SQUAT

Study	n	ICC (95% CI)	*P* Value	ICC (95% CI)
Tate et al[36]	20	0.920 (0.817, 1.000)	<.001	
Scholtes and Salsich[32]	10	0.990 (0.963, 1.000)	<.001	
Total*	30	0.973 (0.915, 1.000)	<.001	

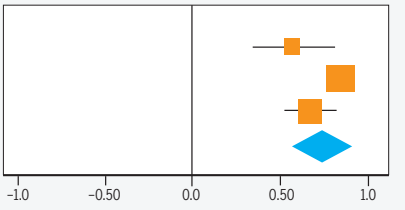


Abbreviations: CI, confidence interval; ICC, intraclass correlation coefficient.
**Heterogeneity: τ^2 = 0.001, Q = 1.661, df (Q) = 1 (P = .197), I^2 = 40%. Test for overall effect: z = 32.685 (P<.001).*

APPENDIX C

INTRADAY INTRARATER RELIABILITY: SINGLE-LEG SQUAT

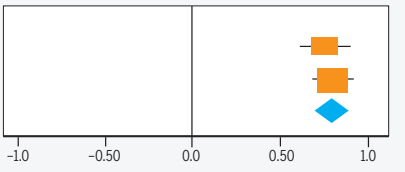
Study	n	ICC (95% CI)	*P* Value	ICC (95% CI)
Munro et al[25]				
Female	10	0.590 (0.347, 0.833)	<.001	
Male	10	0.860 (0.782, 0.938)	<.001	
Scholtes and Salsich[32]	36	0.682 (0.528, 0.836)	<.001	
Total*	56	0.737 (0.571, 0.902)	<.001	



Abbreviations: CI, confidence interval; ICC, intraclass correlation coefficient.
**Heterogeneity: τ^2 = 0.015, Q = 7.363, df (Q) = 2 (P = .025), I^2 = 73%. Test for overall effect: z = 8.733 (P<.001).*

INTRADAY INTRARATER RELIABILITY: UNILATERAL LANDING TASK

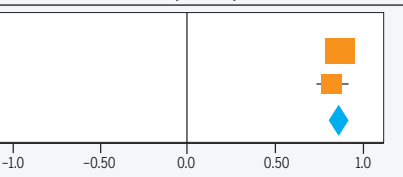
Study	n	ICC (95% CI)	*P* Value	ICC (95% CI)
Munro et al[25]				
Female	10	0.750 (0.603, 0.897)	<.001	
Male	10	0.790 (0.669, 0.911)	<.001	
Total*	20	0.774 (0.680, 0.867)	<.001	



Abbreviations: CI, confidence interval; ICC, intraclass correlation coefficient.
**Heterogeneity: τ^2 = 0.000, Q = 0.169, df (Q) = 1 (P = .681), I^2 = 0%. Test for overall effect: z = 16.198 (P<.001).*

INTRADAY INTRARATER RELIABILITY: BILATERAL LANDING TASK

Study	n	ICC (95% CI)	*P* Value	ICC (95% CI)
Munro et al[25]				
Female	10	0.880 (0.811, 0.949)	<.001	
Male	10	0.830 (0.735, 0.925)	<.001	
Total*	20	0.863 (0.807, 0.919)	<.001	



Abbreviations: CI, confidence interval; ICC, intraclass correlation coefficient.
**Heterogeneity: τ^2 = 0.000, Q = 0.692, df (Q) = 1 (P = .406), I^2 = 0%. Test for overall effect: z = 30.164 (P<.001).*

APPENDIX D

SEARCH STRINGS

Date of Search: March 29, 2017

MEDLINE

1. 2D video*.mp.
2. Two dimensional video*.mp.
3. 3D video*.mp.
4. three dimensional video*.mp.
5. "Reproducibility of Results"/
6. Markerless motion capture*.mp.
7. motion capture system*.mp.
8. real-time assessment*.mp.
9. Observational Screening*.mp.
10. Movement/
11. motion analysis system*.mp.
12. movement analysis*.mp.
13. screening tests*.mp.
14. injury screenings*.mp.
15. movement*.mp.
16. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15
17. Validity*.mp.
18. Validation Studies/
19. Validation*.mp.
20. "Reproducibility of Results"/
21. concurrent validity*.mp.
22. 17 or 18 or 19 or 20 or 21
23. inter?rater reliability*.mp.
24. intra?rater reliability*.mp.
25. inter?rater agreement*.mp.
26. reliability*.mp.
27. intra?rater agreement*.mp.
28. interrater agreement*.mp.
29. intrarater agreement*.mp.
30. Intraclass Correlation Coefficient*.mp.
31. Intrasession Reliability*.mp.
32. ICC*.mp.
33. 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32
34. athletic tasks*.mp.
35. cutting*.mp.
36. landing*.mp.
37. Squat*.mp.
38. Single Leg Squat*.mp.
39. functional tests*.mp.
40. Drop Vertical Jump*.mp.
41. single leg step down*.mp.
42. Biomechanics*.mp.
43. Drop-Jump Landings*.mp.
44. Performance tests*.mp.
45. Squating*.mp.
46. 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44 or 45
47. limb alignment*.mp.

APPENDIX D

48. kinematics*.mp.
49. Lower Extremity Alignment*.mp.
50. dynamic knee valgus*.mp.
51. frontal plane knee motion*.mp.
52. FPPA*.mp.
53. frontal plane projection angle*.mp.
54. knee abduction angle*.mp.
55. lower extremity kinematics*.mp.
56. joint angles*.mp.
57. knee control*.mp.
58. Frontal plane angle*.mp.
59. Knee angles*.mp.
60. Hip Angles*.mp.
61. Trunk angles*.mp.
62. hip motion*.mp.
63. Knee motion*.mp.
64. trunk motion*.mp.
65. 47 or 48 or 49 or 50 or 51 or 52 or 53 or 54 or 55 or 56 or 57 or 58 or 59 or 60 or 61 or 62 or 63 or 64
66. 16 and 22 and 46 and 65
67. 16 and 33 and 46 and 65
68. 16 and 22 and 33 and 46 and 65

Scopus

(TITLE-ABS-KEY ("2D video*" OR "Two dimensional video*" OR "3D video*" OR "three dimensional video*" OR "Markerless motion capture" OR "motion capture system" OR "real-time assessment" OR "Observational Screening" OR "motion analysis system") OR TITLE-ABS-KEY ("movement analysis" OR "screening tests" OR "injury screenings") AND TITLE-ABS-KEY (validity OR validation OR "concurrent validity" OR "Validation studies") AND TITLE-ABS- KEY ("athletic tasks*" OR cutting OR landing OR squat OR squatting OR "Single Leg Squat" OR "functional tests" OR "Drop Vertical Jump" OR "single leg step down" OR biomechanics OR "Drop-Jump Landings" OR movement OR "Performance tests") AND TITLE-ABS-KEY ("limb alignment" OR kinematics OR "Lower Extremity Alignment" OR "dynamic knee valgu" OR "frontal plane knee motion" OR fppa OR "frontal plane projection angle" OR "knee abduction angle" OR "lower extremity kinematics" OR "joint angles" OR "knee control") OR TITLE-ABS-KEY ("Frontal plane angle" OR "Knee angles" OR "Hip angles" OR "Trunk angles" OR "Hip motion" OR "knee motion" OR "Trunk motion"))

SPORTDiscus

S66. S13 AND S18 AND S31 AND S44 AND S63
S65. S13 AND S31 AND S44 AND S63
S64. S13 AND S18 AND S44 AND S63
S63. S45 OR S46 OR S47 OR S48 OR S49 OR S50 OR S51 OR S52 OR S53 OR S54 OR S55 OR S56 OR S57 OR S58 OR S59 OR S60 OR S61 OR S62
S62. Trunk motion*
S61. knee motion*
S60. Hip motion*
S59. Trunk angles*
S58. Hip Angles*
S57. Knee angles*
S56. Frontal plane angle*
S55. knee control*
S54. joint angles*
S53. lower extremity kinematics*
S52. knee abduction angle*
S51. frontal plane projection angle*
S50. FPPA*
S49. frontal plane knee motion*
S48. dynamic knee valgus*
S47. Lower Extremity Alignment*

APPENDIX D

S46. kinematics*
S45. limb alignment*
S44. S32 OR S33 OR S34 OR S35 OR S36 OR S37 OR S38 OR S39 OR S40 OR S41 OR S42 OR S43
S43. Performance tests*
S42. Movement*
S41. Drop-Jump Landings*
S40. Biomechanics*
S39. single leg step down*
S38. Drop Vertical Jump*
S37. functional tests*
S36. Single Leg Squat*
S35. Squat*
S34. Landing*
S33. cutting*
S32. athletic tasks*
S31. S19 OR S20 OR S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28 OR S29 OR S30
S30. intrarater reliability*
S29. intrarater agreement*
S28. ICC*
S27. Intraclass Correlation Coefficient*
S26. Intrasession Reliability*
S25. interrater reliability*
S24. Intra-session Reliability*
S23. Test-retest reliability*
S22. interrater agreement*
S21. reliability*
S20. intra-rater reliability*
S19. inter-rater reliability*
S18. S14 OR S15 OR S16 OR S17
S17. Validation studies*
S16. concurrent validity*
S15. Validation*
S14. Validity*
S13. S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12
S12. injury screenings*
S11. screening tests*
S10. movement analysis*
S9. motion analysis system*
S8. Observational Screening*
S7. real-time assessment*
S6. motion capture system*
S5. Markerless motion capture*
S4. three dimensional*
S3. 3D video*
S2. Two dimensional video*
S1. 2D video*

Embase

1. 2D video*.mp.
2. Two dimensional video*.mp.
3. 3D video*.mp.
4. three dimensional video*.mp.

APPENDIX D

5. Markerless motion capture*.mp.
6. motion capture system*.mp.
7. real-time assessment*.mp.
8. Observational Screening*.mp.
9. motion analysis system*.mp.
10. movement analysis*.mp.
11. screening tests*.mp.
12. injury screenings*.mp.
13. movement*.mp.
14. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13
15. concurrent validity/
16. validity/
17. Validity*.mp.
18. Validation*.mp.
19. concurrent validity*.mp.
20. accuracy/
21. Accuracy*.mp.
22. 15 or 16 or 17 or 18 or 19 or 20 or 21
23. inter?rater reliability*.mp.
24. intra?rater reliability*.mp.
25. inter?rater agreement*.mp.
26. reliability*.mp.
27. intra?rater agreement*.mp.
28. interrater agreement*.mp.
29. intrarater agreement*.mp.
30. Intraclass Correlation Coefficient*.mp.
31. Intrasession Reliability*.mp.
32. ICC*.mp.
33. 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32
34. athletic tasks*.mp.
35. cutting*.mp.
36. landing*.mp.
37. Squat*.mp.
38. Single Leg Squat*.mp.
39. functional tests*.mp.
40. Drop Vertical Jump*.mp.
41. single leg step down*.mp.
42. Drop-Jump Landings*.mp.
43. Performance tests*.mp.
44. Squating*.mp.
45. 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 or 42 or 43 or 44
46. limb alignment*.mp.
47. kinematics*.mp.
48. Lower Extremity Alignment*.mp.
49. dynamic knee valgus*.mp.
50. frontal plane knee motion*.mp.
51. FPPA*.mp.
52. frontal plane projection angle*.mp.
53. knee abduction angle*.mp.
54. lower extremity kinematics*.mp.
55. joint angles*.mp.
56. knee control*.mp.

APPENDIX D

57. Frontal plane angle*.mp.
58. Knee angles*.mp.
59. Hip Angles*.mp.
60. Trunk angles*.mp.
61. hip motion*.mp.
62. Knee motion*.mp.
63. trunk motion*.mp.
64. Biomechanics*.mp.
65. 46 or 47 or 48 or 49 or 50 or 51 or 52 or 53 or 54 or 55 or 56 or 57 or 58 or 59 or 60 or 61 or 62 or 63 or 64
66. 14 and 22 and 45 and 65
67. 14 and 33 and 45 and 65
68. 14 and 22 and 33 and 45 and 65

CINAHL

S69. S14 AND S20 AND S32 AND S47 AND S66
S68. S14 AND S32 AND S47 AND S66
S67. S14 AND S20 AND S47 AND S66
S66. S48 OR S49 OR S50 OR S51 OR S52 OR S53 OR S54 OR S55 OR S56 OR S57 OR S58 OR S59 OR S60 OR S61 OR S62 OR S63 OR S64 OR S65
S65. "Trunk motion*"
S64. "knee motion*"
S63. "Hip motion*"
S62. "Trunk angles*"
S61. "Hip Angles*"
S60. "Knee angles*"
S59. "Frontal plane angle*"
S58. "knee control*"
S57. "joint angles*"
S56. "lower extremity kinematics*"
S55. "knee abduction angle*"
S54. "frontal plane projection angle*"
S53. "FPPA*"
S52. "frontal plane knee motion*"
S51. "dynamic knee valgus*"
S50. "Lower Extremity Alignment*"
S49. "kinematics*"
S48. "limb alignment*"
S47. S33 OR S34 OR S35 OR S36 OR S37 OR S38 OR S39 OR S40 OR S41 OR S42 OR S43 OR S44 OR S45 OR S46
S46. (MH "Athletic Performance")
S45. "Performance tests*"
S44. "Movement*"
S43. "Drop-Jump Landings*"
S42. "Biomechanics*"
S41. "single leg step down*"
S40. "Drop Vertical Jump*"
S39. "functional tests*"
S38. "Single Leg Squat*"
S37. "Squat*"
S36. (MH "Squatting")
S35. "Landing*"
S34. "cutting*"
S33. "athletic tasks*"
S32. S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28 OR S29 OR S30 OR S31

APPENDIX D

S31. "Intrasession reliability*"
S30. "Intra-session Reliability*"
S29. "Test-Retest Reliability*"
S28. (MH "Test-Retest Reliability")
S27. "interrater agreement*"
S26. "Interrater Reliability*"
S25. (MH "Intraclass Correlation Coefficient")
S24. (MH "Interrater Reliability")
S23. "reliability*"
S22. "intra-rater reliability*"
S21. "inter-rater reliability*"
S20. S15 OR S16 OR S17 OR S18 OR S19
S19. "Validation studies*"
S18. (MH "Validation Studies")
S17. "concurrent validity*"
S16. "Validation*"
S15. "Validity*"
S14. S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13
S13. "injury screenings*"
S12. "screening tests*"
S11. "movement analysis*"
S10. "motion analysis system*"
S9. "real-time assessment*"
S8. "Observational Screening*"
S7. "motion capture system*"
S6. (MH "Motion Analysis Systems")
S5. "Markerless motion capture*"
S4. "three dimensional video*"
S3. "3D video*"
S2. "Two dimensional video*"
S1. "2D video*"

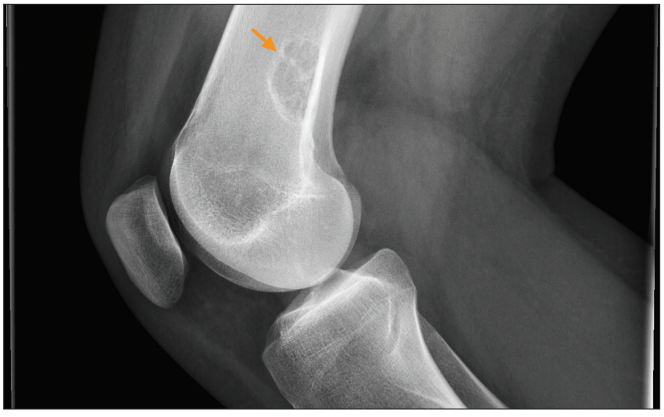


FIGURE 1. Lateral radiograph view of the right knee showing a nonaggressive-appearing, eccentrically located, cortically based mixed lytic and sclerotic bone lesion (arrow) in the posteromedial distal femoral metaphysis, with a sclerotic rim and narrow zone of transition most consistent with fibroxanthoma. There is no evidence of periosteal reaction or fracture of the medial tibial plateau.

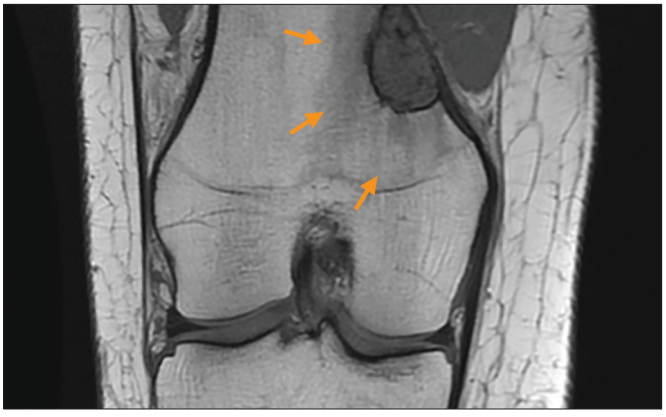


FIGURE 2. Coronal, T1-weighted magnetic resonance image with mostly isointense internal T1 signal to muscle, and mild surrounding marrow and soft tissue edema with corresponding decreased T1 marrow signal surrounding the lesion (indicated by arrows), indicative of bone stress injury.[2]

Differential Diagnosis of Complicated Bone Stress Injury in a Female Collegiate Swimmer

RONALD P. MILLER, PT, DPT, OCS, SCS, FAAOMPT, *Tactical Sports and Orthopedic Manual Physical Therapy Fellowship, US Air Force Academy, CO.*
R. JUDD ROBINS, MD, *Orthopedic Sports Medicine Service, US Air Force Academy, CO.*
SHANE D. DIECKMAN, DO, *10th Medical Group, US Air Force Academy, CO.*

An 18-year-old woman presented to a direct-access military physical therapy clinic after a fall directly onto the knee, with subsequent medial knee pain. Pain at rest was aggravated with higher-impact activities. Localized edema was present at the medial femoral condyle and tibial plateau, without loss of range of motion. A National Collegiate Athletic Association Division I swimming recruit, she recently incorporated land-based training 6 weeks prior to initiating service academy training. She was screened for female athlete triad upon intake, without positive findings, though decreased bone mineral density was probable.

The patient's history, rapid activity progression, absence of clinical signs of ligamentous or meniscal involvement, and pain with palpation to the medial femoral condyle and tibial plateau led the physical therapist to order radiographs to evaluate for bone stress injury. Radiographs revealed a pre-existing metaphyseal fibrous defect, common in 30% to 40% of adolescents[1] (**FIGURE 1**). On the same day, an orthopaedic consultation recommended magnetic resonance imaging (**FIGURE 2**) and computed tomography (**FIGURES 3** and **4**, available at www.jospt.org), which showed a progression from a benign cortical defect to disruption of the posterior femur and surrounding bone marrow edema consistent with bone stress injury.

The physical therapist recommended partial weight bearing until the patient's pain resolved, which occurred in approximately 3 weeks. The patient returned to competitive swimming pain free and remained asymptomatic, reducing her impact activity to preinjury levels.

This case exemplifies the occurrence of bone stress injury in young adult bone when subjected to a rapid increase in land-based training. Under similar conditions, benign lesions may progress to pathologic fractures. Finally, in direct-access environments, thorough subjective screening and utilization of imaging can assist in differential diagnoses to optimize care. ◉ *J Orthop Sports Phys Ther 2018;48(10):823. doi:10.2519/jospt.2018.7731*

References
1. Betsy M, Kupersmith LM, Springfield DS. Metaphyseal fibrous defects. *J Am Acad Orthop Surg*. 2004;12:89-95.
2. Nattiv A, Kennedy G, Barrack MT, et al. Correlation of MRI grading of bone stress injuries with clinical risk factors and return to play: a 5-year prospective study in collegiate track and field athletes. *Am J Sports Med*. 2013;41:1930-1941. https://doi.org/10.1177/0363546513490645

OLUWATOYOSI B.A. OWOEYE, PT, PhD[1,2] • **J. PRESTON WILEY,** MPE, MD[1,3] • **RICHARD E.A. WALKER,** MD[4,5]
LUZ PALACIOS-DERFLINGHER, PhD[1,6] • **CAROLYN A. EMERY,** PT, PhD[1,2,5-7]

Diagnostic Accuracy of a Self-report Measure of Patellar Tendinopathy in Youth Basketball

BACKGROUND: To engage clinicians in diagnosing patellar tendinopathy in large surveillance studies is often impracticable. The Oslo Sports Trauma Research Centre-patellar tendinopathy (OSTRC-P) questionnaire, a self-report measure adapted from the OSTRC questionnaire, may provide a viable alternative.

OBJECTIVES: To evaluate the diagnostic accuracy of the OSTRC-P questionnaire in detecting patellar tendinopathy in youth basketball players when compared to clinical evaluation.

METHODS: Following the STAndards for Reporting of Diagnostic accuracy studies guidelines, the researchers recruited 208 youth basketball players (13-18 years of age) for this prospective diagnostic accuracy validation study. Participants completed the OSTRC-P questionnaire (index test) prior to a clinical evaluation (reference standard) by a physical therapist blinded to OSTRC-P questionnaire results. Sensitivity, specificity, predictive value, likelihood ratio, and posttest probability were calculated. Linear regression was used to examine the association between the OSTRC-P questionnaire severity score and the patellar tendinopathy severity rating during a single-leg decline squat.

RESULTS: The final analysis included 169 players. The OSTRC-P questionnaire had a sensitivity of 79% (95% confidence interval [CI]: 65%, 90%), specificity of 98% (95% CI: 94%, 100%), positive predictive value of 95% (95% CI: 83%, 99%), negative predictive value of 92% (95% CI: 86%, 96%), positive likelihood ratio of 48 (95% CI: 12, 191), and negative likelihood ratio of 0.21 (95% CI: 0.12, 0.37). The posttest probabilities were 95% and 8%, given positive and negative results, respectively. A positive association was found between OSTRC-P questionnaire severity score and single-leg decline squat rating (β = 0.08; 95% CI: 0.03, 0.12; P = .001).

CONCLUSION: The OSTRC-P questionnaire is an acceptable alternative to clinical evaluation for self-reporting patellar tendinopathy and grading its severity in settings involving youth basketball players.

LEVEL OF EVIDENCE: Diagnosis, level 1b. *J Orthop Sports Phys Ther 2018;48(10):758-766. Epub 27 Apr 2018. doi:10.2519/jospt.2018.8088*

KEY WORDS: *epidemiology, jumper's knee, OSTRC questionnaire, overuse injury*

The knee is one of the most vulnerable joints for overuse injuries in jumping sports such as basketball, and patellar tendinopathy is the most common overuse injury among jumping athletes.[8,24] Nonetheless, the prevalence and etiology of patellar tendinopathy has yet to be fully described due to the daunting task of diagnosing patellar tendinopathy, particularly in large surveillance studies. The diagnosis of patellar tendinopathy remains a challenge, because there is currently no gold standard diagnostic technique.[13,23]

Clinical evaluation involving patient history of pain during or after physical activity and physical examination of local tenderness at the inferior pole of the patella is currently regarded as the reference standard.[23] However, some experts have cautioned against relying solely on palpation at the inferior pole of the patella for diagnosing patellar tendinopathy in youth athletes[7] and, instead, have advocated the use of functional loading tests, such as the single-leg decline squat.[15,21] In studies involving the diagnosis of patellar tendinopathy by a clinician, researchers have used clinical evaluation with or without the single-leg decline squat stress test.[3,7,12,14,21,22]

Although a remarkable number of studies have been published on the prevalence of patellar tendinopathy,[7,8,12,25] these studies are often cross-sectional designs, and their methodologies require further validation to expose the true burden of patellar tendinopathy. Further, most studies are of adult elite players, and literature on youth sport remains

[1]Sport Injury Prevention Research Centre, Faculty of Kinesiology, University of Calgary, Calgary, Canada. [2]Alberta Children's Hospital Research Institute, University of Calgary, Calgary, Canada. [3]Sport Medicine Centre, University of Calgary, Calgary, Canada. [4]Department of Radiology, Cumming School of Medicine, University of Calgary, Calgary, Canada. [5]McCaig Institute for Bone and Joint Health, University of Calgary, Calgary, Canada. [6]Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, Canada. [7]Department of Pediatrics, Cumming School of Medicine, University of Calgary, Calgary, Canada. The Conjoint Health Research Ethics Board of the University of Calgary, Calgary, Canada approved the study protocol (REB16-0864). This study was funded by the National Basketball Association (NBA) and General Electric (GE) Healthcare through the NBA/GE Orthopedics and Sports Medicine Collaboration Grant on Tendinopathy. The Sport Injury Prevention Research Centre is supported by an International Olympic Committee Research Centre Award. Dr Owoeye has a postdoctoral fellowship award from Alberta Children's Hospital Research Institute. Dr Emery holds a Chair in Pediatric Rehabilitation funded by Alberta Children's Hospital Foundation. The authors certify that they have no affiliations with or financial involvement in any organization or entity with a direct financial interest in the subject matter or materials discussed in the article. Address correspondence to Dr Oluwatoyosi Owoeye, Sport Injury Prevention Research Centre, Faculty of Kinesiology, University of Calgary, 2500 University Drive NW, Calgary, AB T2N 1N4 Canada. E-mail: oluwatoyosi.owoeye@ucalgary.ca

sparse. Recent evidence suggests that regular and consistent monitoring of tendinopathy is imperative for a robust understanding of the true burden and etiology of tendinopathy and other overuse injuries.[2,4,10,19] However, field diagnosis of patellar tendinopathy by a physician or physical therapist is expensive and impracticable for large surveillance studies involving weekly or biweekly monitoring of tendinopathy in athletes.

van der Worp et al[25] and de Vries et al[8] have published the most comprehensive studies on the etiology of patellar tendinopathy to date. These studies used a self-administered pain map to detect patellar tendinopathy in adult basketball and volleyball athletes 18 to 35 years of age. The participants were classified as having patellar tendinopathy when they indicated having pain at the inferior pole of the patella. However, the self-report tool was not previously validated. There is currently no criterion-validated questionnaire for reporting tendinopathy in surveillance studies.

In order to accurately quantify the problem of overuse injuries in sports, the Oslo Sports Trauma Research Center (OSTRC) developed and validated an Overuse Injury Questionnaire.[5] This questionnaire has been used to describe the prevalence and severity of overuse injuries in numerous sports.[1,4,10,19] However, this questionnaire does not provide information about individual overuse injuries and, therefore, is limited in its scope of application to specific injury types and diagnoses, such as patellar tendinopathy.

Given that overuse injuries can be captured through self-report and that the clinical diagnosis of patellar tendinopathy hinges on an initial self-report of anterior knee pain localized to the inferior pole of the patella,[13,23] the authors of the current study posited that it would be possible to identify ongoing patellar tendinopathy (ie, irritable tendon pain) using an adapted version of the OSTRC Overuse Injury Questionnaire specifically designed for self-reporting patellar tendinopathy: the Oslo Sports Trauma Research Center-patellar tendinopathy (OSTRC-P) questionnaire (**FIGURE 1**). The OSTRC-P comprises the original OSTRC Overuse Injury Questionnaire and features additional questions about knee injury onset and symptoms for identifying patellar tendinopathy. If reasonable diagnostic accuracy could be determined, then this new questionnaire could provide sport injury prevention researchers, experts in tendinopathy research, and clinicians with a simple and low-cost instrument for reporting and monitoring patellar tendinopathy in athletes. This would be significant for advancing the prevention and management of patellar tendinopathy in athletes who participate in sports, such as basketball, in which overuse injuries are common.

The authors hypothesized that there would be minimal discordance between the OSTRC-P questionnaire and a reference standard of clinical evaluation by a physical therapist for the presence or absence of patellar tendinopathy in competitive youth basketball players with or without other knee conditions. Therefore, the objective of this study was to evaluate the diagnostic accuracy of the OSTRC-P questionnaire in detecting patellar tendinopathy in this population, using clinical evaluation as a reference standard. Further, the researchers examined the criterion-based validity of the OSTRC-P questionnaire severity score against the patellar tendinopathy severity rating during clinical evaluation using the single-leg decline squat test.

METHODS

Design and Setting

A prospective diagnostic accuracy validation study was conducted within a larger cohort study aimed at examining risk factors for patellar and Achilles tendinopathy (PAT). The PAT Prevention Study was a prospective cohort study in 518 youth basketball players during the 2016-2017 high school and club basketball seasons in Calgary, Canada (Research Ethics Board approval REB16-0864). The objectives of the current substudy were formulated within the larger PAT Prevention Study a priori. A total of 208 players (34 teams) were invited from the PAT Prevention Study cohort in January and February (school basketball) and May and June (club basketball) 2017 to participate in the present study, based on feasibility. All players were asked to complete the weekly OSTRC-P questionnaire (index test) as part of the PAT Prevention Study. The index test was then followed up to a week later with an "on-court" clinical evaluation (reference standard) by the study physical therapist (O.O.) for all players present at scheduled team practices. The STAndards for Reporting of Diagnostic accuracy studies guidelines[6] were followed for this study.

The OSTRC-P questionnaire and clinical evaluation were used to assess for patellar tendinopathy on both knees in players. Further, specific questions on the OSTRC-P questionnaire and a functional stress test of the single-leg decline squat (using player-perceived patellar tendinopathy pain rating) during clinical examination were used to evaluate patellar tendinopathy severity.

Index Test

The OSTRC-P questionnaire (**FIGURE 1**) is a 10-item questionnaire designed to detect patellar tendinopathy in epidemiological studies in youth basketball. Six new questions were added to the 4 original questions in the OSTRC Overuse Injury Questionnaire, relating to self-reported problems of the knee in the previous week.[5] These new questions, appropriately designed for adolescent basketball players, probe any ongoing knee pain to "rule in" or "rule out" patellar tendon pain. Only players with potential patellar tendinopathy based on self-reported knee pain are prompted to answer all additional questions, which follow in sequence until the final question of whether a player's anterior knee pain is "on the bottom tip of the kneecap" (**FIGURE 1**).

The 6 additional questions were formulated by research team members and

face validated through a consortium of clinicians and 2 high school basketball teams after a round of reviews and multiple drafts. Specifically, these additional questions underwent a 3-step screening process: (1) initial draft and review by a large group of researchers, comprising all study coauthors and a team of expert collaborators in the fields of epidemiology, medicine, rehabilitation, strength and conditioning, coaching, and biomechanics, on the PAT Prevention Study; (2) pilot testing and feedback among 2 high school basketball teams (1 boys team and 1 girls team), each comprising 2 coaches and 12 youth basketball players; and (3) subsequent reviews and corrections by the same group of researchers and

Please answer all questions regardless of whether or not you have problems with your knees. Select **(tick or circle)** the option that is most appropriate for you, and in the case that you are unsure, try to give an answer as best you can anyway.

The term "knee problems" refers to pain, ache, stiffness, swelling, instability/giving way, locking or other complaints related to one or both knees. Please note that all questions in this questionnaire refer to the previous week.

Question 1 - Have you had any difficulties participating in normal practice and game due to knee problems this past week?

a) Full participation without knee problems
b) Full participation but with knee problems
c) Reduced participation due to knee problems
d) Cannot participate due to knee problem

Question 2 - To what extent have you reduced your practice volume due to knee problems this past week?

a) No reduction b) To a minor extent c) To a moderate extent
d) To a major extent e) Cannot participate at all

Question 3 - To what extent have knee problems affected your performance this past week?

a) No effect b) To a minor extent c) To a moderate extent
d) To a major extent e) Cannot participate at all

Question 4 - To what extent have you experienced knee pain related to playing basketball this past week?

a) No pain b) Mild pain c) Moderate pain d) Severe pain

** If you answered "a" to all 4 questions, questionnaire is completed for the week; if otherwise please answer the following questions:*

Question 5 – Do you still experience any knee pain, especially during and/or after basketball participation?

a) Yes b) No **If "yes" please proceed to Question 6, if otherwise questionnaire is completed*

Question 6 - Is the knee pain you are reporting?

a) The same knee pain as in previous week(s) b) A return of a knee pain that had gone away
c) A knee pain that is being experienced for the first time this past week

Question 7 - On which knee do you have pain?

a) Right knee b) Left knee c) Both knees (right and left)

Complete this section as applicable		
	Right knee	Left knee
Question 8 **Describing the onset of your knee pain, was it:**	**a)** Of a gradual or sudden onset that is **unidentifiable** with any event? **b)** Of a sudden onset that is **clearly identifiable** (e.g. impact or collision with another player)?	**a)** Of a gradual or sudden onset that is **unidentifiable** with any event? **b)** Of a sudden onset that is **clearly identifiable** (e.g. impact or collision with another player)?
If your answer to Question 8 is "a" please proceed to Question 9, if otherwise questionnaire is completed		
Question 9 **Describe the location of your knee pain (you can select multiple):**	**a)** Front of the knee **b)** Back of the knee **c)** Inside of the knee (medial) **d)** Outside of the knee (lateral)	**a)** Front of the knee **b)** Back of the knee **c)** Inside of the knee (medial) **d)** Outside of the knee (lateral)
If your answer to Question 9 is "a" please proceed to Question 10, if otherwise questionnaire is completed		
Question 10 **Is the pain in the front of your knee on the bottom tip of your kneecap?**	**a)** Yes **b)** No	**a)** Yes **b)** No

FIGURE 1. The OSTRC-patellar tendinopathy questionnaire: adapted OSTRC Overuse Injury Questionnaire for patellar tendinopathy. Abbreviation: OSTRC, Oslo Sports Trauma Research Center. Adapted from Clarsen et al,[5] with permission from BMJ Publishing Group Ltd. Copyright ©2012 BMJ Publishing Group Ltd.

collaborators to produce a final version of the questionnaire.

All participants were asked to complete the OSTRC-P questionnaire weekly, either online using a mobile phone or by hand using an identical hardcopy paper version. The response to the last query in the questionnaire, which as described above is dependent on the preceding questions, classified a player as either having patellar tendinopathy or not having it. The first 4 questions of the OSTRC-P questionnaire (as in the original OSTRC Overuse Injury Questionnaire) were used to calculate players' patellar tendinopathy severity score as instructed by Clarsen et al.[5] Each of these items has a score ranging from 0 (no problems) to 25 (the maximum level), with the sum of the scores on all 4 items ranging from 0 to 100.

Reference Standard

Currently, there is no gold standard for the clinical diagnosis of patellar tendinopathy. The reference standard chosen for comparison with the OSTRC-P questionnaire in the current study was clinical evaluation. All players were examined by the same physical therapist (O.O.), a clinician-scientist (ie, a physical therapist with extensive additional training in health research) with 13 years of experience in sports physical therapy practice and research. The physical therapist was blinded to the outcomes on the OSTRC-P questionnaire for all participants during the testing periods.

The criteria used for the diagnosis of patellar tendinopathy were standardized by 2 of the study's authors (O.O. and J.W., an experienced sports medicine physician) and were informed by current evidence for best practice.[11,13,21-23] This involved the following 3-step process: (1) the participant reported ongoing or history of pain (previous week) localized to the inferior pole of the patella during or after basketball practice or a game; (2) the examining physical therapist conducted a focused physical examination in which the knee was placed in extension and the patella pushed distally (participant in long sitting), and pain was reproduced with palpation at the inferior pole of the patella (localized tenderness); and (3) pain was reproduced during the single-leg decline squat test up to 90° of knee flexion or to the threshold of pain tolerance on a 25° decline board (Physiofoam.co.uk; Physiofoam Ltd, Knebworth, UK). Participants needed to have a positive result in step 1 to proceed to step 2, and in step 2 to proceed to step 3.

To assess ongoing patellar tendinopathy severity (pain intensity) during clinical evaluation, participants with patellar tendinopathy were asked to rate their pain on a numeric pain-rating scale, with 0 as no pain and 10 as the worst pain imaginable, while completing the single-leg decline squat test. Cases in which patellar tendinopathy was identified as coexisting with other knee conditions were regarded as mixed diagnosis. For a "yes patellar tendinopathy" to be legitimate, participants needed to have the minimum severity score on each of the test measures (ie, 6/100 on the OSTRC-P questionnaire and 1/10 on the single-leg decline squat test during clinical evaluation).

Data and Statistical Analysis

Player characteristics and other descriptive variables were presented using mean ± SD values, frequencies, and percentages where applicable. Players were classified as either "no patellar tendinopathy" or "yes patellar tendinopathy" on each knee for the index (OSTRC-P questionnaire) and reference (clinical evaluation) tests. Final diagnosis of patellar tendinopathy was decided by clinical evaluation. Because limbs were not independent of each other in individuals, players rather than knees were used as the unit of analysis to avoid the possibility of inflating the association between the OSTRC-P questionnaire and clinical evaluation.[16,17] However, because some players with bilateral patellar tendinopathy might have a combination of accurate and inaccurate results on the OSTRC-P questionnaire when compared with clinical evaluation, the authors decided that in such players, the more severe knee (based on severity rating during the clinical evaluation using the single-leg decline squat) would be considered for data analyses.

Instances of "no patellar tendinopathy" included completely injury- and symptom-free knees or knees with other conditions aside from patellar tendinopathy. Instances of "yes patellar tendinopathy" included all knees reported or diagnosed as patellar tendinopathy, with or without other acute or overuse knee injuries (referred to as mixed diagnosis). Sensitivity (true positives), specificity (true negatives), positive and negative predictive values, positive and negative likelihood ratios, and pretest and posttest probabilities with associated confidence intervals (CIs) were calculated for the outcome of yes or no patellar tendinopathy. Because there are currently no recommended thresholds for sensitivity, specificity, and predictive values, the researchers operationally determined an acceptable/important level of accuracy in this study to be 66.6% (the ability of the OSTRC-P questionnaire to detect at least 2 of 3 youth basketball players with patellar tendinopathy). For likelihood ratio estimates, positive likelihood ratios greater than 5 and negative likelihood ratios less than 0.2 were considered important based on the thresholds suggested by Portney and Watkins.[20]

Exploratory analysis investigating the effect of age groups—individuals younger than 16 years of age (younger players, aged 13 to 15 years) and those younger than 19 years of age (older players, aged 16 to 18 years)—on the sensitivity and specificity of the OSTRC-P questionnaire was also performed. Finally, to validate the OSTRC-P questionnaire severity score against ongoing patellar tendinopathy severity rating during clinical evaluation using the single-leg decline squat, simple linear regression analyses ($\alpha = .05$) were conducted.

RESULTS

Overall, 187 of 208 players (90%) aged 13 to 18 years (30 teams) completed the OSTRC-P questionnaire

between 1 and 4 days prior to clinical evaluation. Of these 208 players, 169 (81%) completed the clinical evaluation and were included in final analyses. The flow of participants through this study and reasons for exclusion from the final analysis are described in **FIGURE 2**, and participant characteristics are provided in **TABLE 1**. Six of 48 patellar tendinopathy cases diagnosed (12.5%, all males) by the study physical therapist were reported as mixed diagnosis, in which patellar tendinopathy coexisted with other conditions such as patellofemoral pain syndrome, Osgood-Schlatter disease, and quadriceps tendinitis. All players with positive results on either or both the OSTRC-P questionnaire and/or the clinical evaluation (step 2) reported severity scores of at least 6 and 1, respectively. There were no adverse events reported by any player from completing the OSTRC-P questionnaire or undergoing the clinical evaluation.

The OSTRC-P questionnaire detected 38 of 48 patellar tendinopathy cases diagnosed through clinical evaluation, estimating a sensitivity of 79% (95% CI: 65%, 90%). Further, the OSTRC-P questionnaire cleared 119 of the 121 cases as not having patellar tendinopathy, estimating a specificity of 98% (95% CI: 94%, 100%). Details of the positive predictive values, negative predictive values, and likelihood ratios are presented in **TABLE 2**. The pretest probability of patellar tendinopathy in this study was 28% (95% CI: 22%, 36%). Given a positive result, the posttest probability of the OSTRC-P questionnaire was 95% (95% CI: 83%, 99%), and given a negative result, the posttest probability was 8% (95% CI: 5%, 13%).

Exploratory analysis investigating the effect of age groups on the sensitivity and specificity of the OSTRC-P questionnaire is presented in **TABLE 3**. A sensitivity of 83% (95% CI: 69%, 93%) was reported in older players, and a sensitivity of 50% (95% CI: 12%, 88%) was found in younger players.

A significant positive association was found between OSTRC-P questionnaire score and clinical evaluation single-leg decline squat rating of patellar tendinopathy severity, as demonstrated in linear regression analysis ($\beta = 0.08$; 95% CI: 0.03, 0.12; $P = .001$; $R^2 = 0.29$).

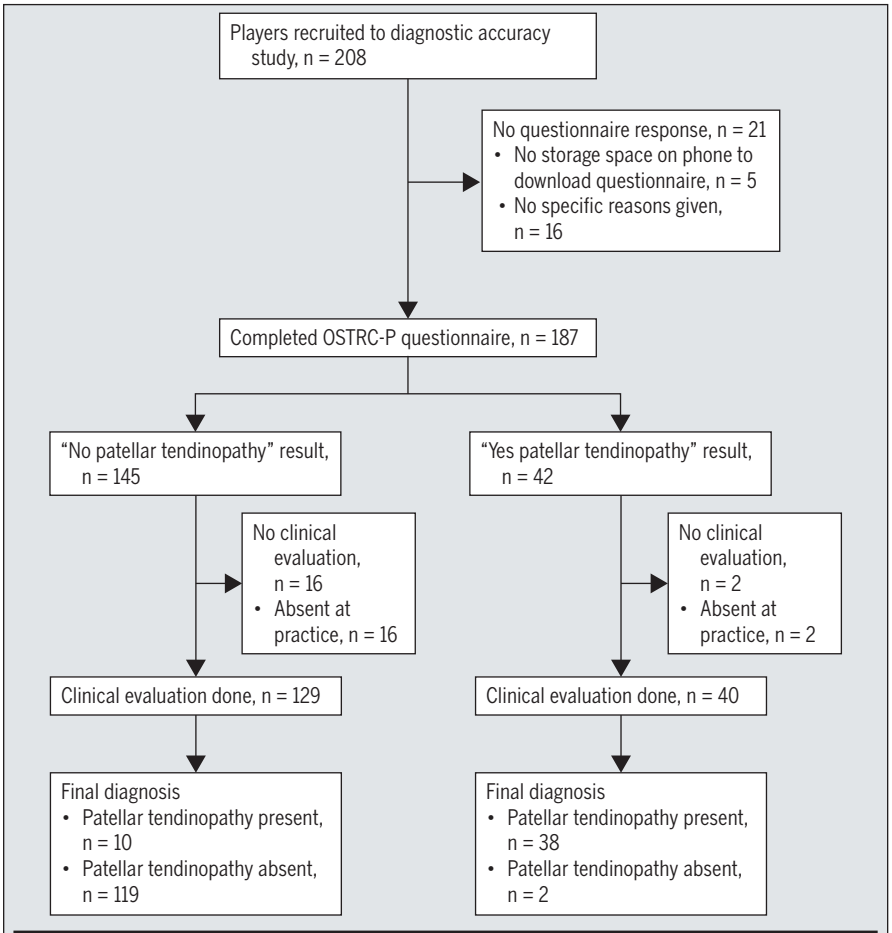


FIGURE 2. Flow of players through the study. Abbreviation: OSTRC-P, Oslo Sports Trauma Research Center-patellar tendinopathy.

DISCUSSION

The objective of this study was to evaluate the diagnostic accuracy of an adapted version of the OSTRC Overuse Injury Questionnaire, referred to as the OSTRC-P questionnaire, in comparison with a clinical evaluation. Beyond the basic measures of accuracy for a new test or tool, which are sensitivity and specificity,[6] the researchers evaluated other relevant measures of diagnostic test accuracy such as predictive value, likelihood ratio, and pretest/posttest probability of patellar tendinopathy in order to thoroughly examine the diagnostic utility of the OSTRC-P questionnaire in youth basketball players.

Based on the thresholds the authors considered as clinically important, findings from this study demonstrate that overall, the OSTRC-P questionnaire, with a sensitivity of 79% and specificity of 98%, offers an acceptable level of accuracy in detecting patellar tendinopathy among youth basketball players. The OSTRC-P questionnaire showed excellent predictive values and likelihood ratios. For example, among those who had a positive test result, the probability of patellar tendinopathy truly occurring was 95%, and among those who had a negative result, the probability of not having patellar tendinopathy was 92%. Also, the

positive likelihood ratio of 48 found in this study is far greater than 5 (a threshold for an important effect), indicating that patellar tendinopathy is likely to be present with a positive test.[20] Similarly, the negative likelihood ratio of 0.21 from this study approximates 0.2, a threshold that indicates that the disorder has a likelihood of being absent with a negative test.[20] Additionally, the high posttest probability of 95% for a positive test result on the OSTRC-P questionnaire provides assurance of the presence of patellar tendinopathy, and the low probability of 8% for a negative test result provides reasonable assurance of the absence of the condition. These findings will potentially improve clinicians' and researchers' confidence about the diagnostic certainty of the OSTRC-P questionnaire.

However, exploratory evaluation of the effects of age on the accuracy of the OSTRC-P questionnaire in stratified data analysis suggests that the questionnaire may be less sensitive in younger adolescents (sensitivity of 50% in players 13 to 15 years of age versus 83% in players 16 to 18 years of age). Although the original OSTRC Overuse Injury Questionnaire has been used in players younger than 15 years of age,[19] it was originally validated in athletes older than 15 years of age, and its use is primarily with older athletes.[1,5,10,19] This explains, in part, the reason why the OSTRC-P questionnaire recorded a low sensitivity in younger players. Nonetheless, it is possible that the small sample size and very few cases of patellar tendinopathy in the stratified analysis impacted the sensitivity estimate reported. The diagnostic utility of the OSTRC-P questionnaire in athletes younger than 16 years of age demands further investigation.

The secondary objective of this study was to investigate the association between the OSTRC-P questionnaire severity score and the patellar tendinopathy severity rating based on clinical examination. Although a linear regression analysis was used to explore this relationship, the researchers' intention was not to predict patellar tendinopathy severity rating from the OSTRC-P questionnaire, but to examine whether these variables would be practically associated with each other.

Accordingly, the previously validated severity measure in the original OSTRC

TABLE 1 Player Characteristics*

Variable	Overall (n = 169)
Age, y	16.3 ± 1.3
Weight, kg	65.5 ± 15.7
Height, m	1.75 ± 0.12
Age group, n (%)	
<16 y	41 (24.3)
<19 y	128 (75.7)
Sex, n (%)	
Male	117 (69.2)
Female	52 (30.8)
Play position, n (%)	
Post	44 (26.0)
Guard	53 (31.4)
Combo	45 (26.6)
Missing	27 (16.0)
Dominant leg, n (%)	
Right	111 (65.6)
Left	29 (17.2)
Missing	29 (17.2)
Setting, n (%)	
School basketball	91 (53.8)
Club basketball	78 (46.2)

**Values are mean ± SD unless otherwise indicated.*

TABLE 2 Diagnostic Accuracy of the OSTRC-P Questionnaire in Comparison With Clinical Evaluation as Criterion Standard in Youth Basketball Players*

	Clinical Examination: Condition Present, n							
OSTRC-P Questionnaire	Yes	No	Sensitivity	Specificity	+PV	–PV	+LR	–LR
Condition present, n			79 (65, 90)	98 (94, 100)	95 (83, 99)	92 (86, 96)	48 (12, 191)	0.21 (0.12, 0.37)
Yes	38	2						
No	10	119						

Abbreviations: –LR, negative likelihood ratio; +LR, positive likelihood ratio; OSTRC-P, Oslo Sports Trauma Research Center-patellar tendinopathy; –PV, negative predictive value; +PV, positive predictive value.
**Values are percent (95% confidence interval) unless otherwise indicated.*

Overuse Injury Questionnaire,[5] which the current study's authors adapted in the OSTRC-P questionnaire, may be specifically useful in estimating patellar tendinopathy severity in surveillance studies and in clinical settings. The OSTRC-P questionnaire severity scores demonstrated a significant positive linear relationship with patellar tendinopathy severity ratings during clinical evaluation. The regression model suggests that on average, for each unit rise in OSTRC-P questionnaire severity score, the patellar tendinopathy severity rating based on clinical examination using the single-leg decline squat test increases by 0.08. Although the R^2 value suggests that 29% of the variation in clinical patellar tendinopathy severity score may be explained by the linear relationship with the OSTRC-P questionnaire severity score, the hypothesis test demonstrates that the overall relationship is statistically significant (P = .001). Thus, the initial 4 questions on the OSTRC-P questionnaire potentially provide added benefit when scoring severity in patellar tendinopathy cases identified in any given population of young basketball athletes using prespecified calculations.[5] Researchers, coaches, and clinicians are then able to monitor players' conditions. This is particularly useful for decisions on load modification in players with patellar tendinopathy and in prospective studies with panel designs of weekly or biweekly overuse injury surveillance, similar to what was done in previous studies for nonspecific overuse problems in athletes.[1,4,10,19]

The original questionnaire, the OSTRC Overuse Injury Questionnaire, was developed to monitor overuse problems in specific body regions in athletes. However, the 4 questions contained therein would only identify region-specific problems (eg, knee problems) in the past week. To specifically report whether the problem identified by the questionnaire is an acute or overuse knee injury, a clinician/researcher would need to follow up with the athlete via phone conversation at the least. This has been the practice in most of the studies that have used the OSTRC Overuse Injury Questionnaire.[1,4,5,10,19]

The authors adapted the OSTRC Overuse Injury Questionnaire to allow athletes to self-report all details relating to knee problems experienced in the previous week. This comprises reporting whether the knee problem includes ongoing pain; is acute, overuse, recurrent, or new; affects the right, left, or both knees; affects the anterior knee; and, ultimately, whether there is pain at the inferior pole of the patella. Apart from providing other specific and important information on the knee problems and/or pain reported by a player (eg, differentiating acute injuries from overuse), the OSTRC-P questionnaire helps identify the presence of patellar tendinopathy, negating the need for any initial or follow-up contact with a clinician and providing an opportunity for continued player monitoring.

The OSTRC-P questionnaire is a practical and economical tool that can be used to detect the prevalence and severity of patellar tendinopathy in a large sample of youth basketball players. The OSTRC-P questionnaire may also be useful for continued player monitoring during treatment/rehabilitation. The questionnaire may be administered either electronically (eg, via e-mail or software application on mobile phones) or in paper form. The authors are confident that this simple tool will be invaluable in future studies and has the potential to advance knowledge in the field of patellar tendinopathy research.

Study Limitations

In this study, the researchers estimated all relevant measures of diagnostic accuracy to enhance the potential use of the OSTRC-P questionnaire in other settings. However, application of this questionnaire in settings other than youth basketball should be made with caution. Further research is needed to evaluate the validity of the OSTRC-P questionnaire in other sports. Further validation of the OSTRC-P questionnaire is particularly warranted in sports with high patellar tendinopathy prevalence, such as volleyball and running.

The OSTRC-P questionnaire lacks the ability to rule out other knee conditions when they coexist with patellar tendinopathy. The extent to which cases of mixed diagnosis affect the accuracy of the OSTRC-P questionnaire is unknown, as this was not evaluated in the current study due to the small number of cases designated as mixed diagnosis. Further, this study lacked the power for stratified analyses to fully investigate the possible effects of age and sex on the diagnostic accuracy of the OSTRC-P. Thus, the authors only evaluated the sensitivity and specificity of the OSTRC-P questionnaire in comparison to clinical evaluation in an exploratory analysis to gain an insight

TABLE 3 Exploratory Analysis of the Sensitivity and Specificity of the OSTRC-P Questionnaire by Age Group

	Clinical Examination			
	<16 y: Condition Present, n		<19 y: Condition Present, n	
OSTRC-P Questionnaire	Yes	No	Yes	No
Condition present, n				
Yes	3	1	35	1
No	3	34	7	85
Sensitivity, %*	50 (12, 88)		83 (69, 93)	
Specificity, %*	97 (85, 100)		99 (94, 100)	

Abbreviation: OSTRC-P, Oslo Sports Trauma Research Center-patellar tendinopathy.
**Values in parentheses are 95% confidence interval.*

into potential differences in the accuracy of the questionnaire between older and younger adolescents. Future research with larger samples will aid a more robust understanding of the potential effects of age on the diagnostic accuracy of the OSTRC-P questionnaire.

It was impracticable for the study physical therapist to evaluate the entire cohort in the PAT Prevention Study (ie, both patellar tendinopathy cases and non–patellar tendinopathy cases) within days of completing the OSTRC-P questionnaire; thus, the sample in the present study was based on a convenience sampling of participants through enrollment of teams that were approached and willing to participate. The extent to which this sample is biased is unknown. Also, the small number of participant cases/noncases in some of the 2-by-2 cells resulted in wide CIs reported for some of the diagnostic accuracy estimates, especially in the positive likelihood ratio value, suggesting imprecision in the plausible range of values around the point estimate. However, the lower limit of the positive likelihood ratio (12) and the upper limit of the negative likelihood ratio (0.37) are within the thresholds considered to be important.

The OSTRC-P questionnaire was compared with a reference standard and not a gold standard test, which would have been ideal. Therefore, its accuracy depends on the validity of clinical evaluation as a reference standard in diagnosing patellar tendinopathy. The pretest probability (prevalence) of patellar tendinopathy reported in this study is probably due to convenience sampling; a patellar tendinopathy prevalence of 28.4% is high compared to the 7% reported by Cook et al[7] on clinical grounds 18 years ago. However, so much has changed in tendinopathy research, and given the level of competitiveness among youth participants since then, it is very likely that the prevalence of patellar tendinopathy has increased accordingly.

Nevertheless, it is possible that a few cases reported as patellar tendinopathy are actually other conditions with a similar pattern of pain in the anterior knee region. For example, it is clinically impossible to differentiate between Sinding-Larsen-Johansson syndrome and patellar tendinopathy in adolescents. Sinding-Larsen-Johansson syndrome may be clinically considered as part of a spectrum of tendon pathology in adolescents in this age bracket. Furthermore, the authors acknowledge that a report of 6 of 48 cases (12.5%) as mixed diagnosis may well be an underestimation. The clinical evaluation process described in this study was focused on diagnosing patellar tendinopathy only and was not a complete knee evaluation. It is possible that the on-court clinical evaluation process missed a few other conditions that would need further diagnostic processes. Although clinical evaluation is not without limitations, it remains the reference standard of choice for diagnosing patellar tendinopathy, as there is currently no gold standard test.[13,23]

Finally, the OSTRC-P questionnaire was administered both electronically and in paper form. It is unknown whether the means of delivering the questionnaire had any impact on its accuracy. The researchers did not do a subanalysis of data to evaluate this because that was not included in the present study's objectives. Nevertheless, it is unlikely that there are differences in the accuracy of these 2 methods, as there is extensive evidence from systematic reviews showing that electronic and paper administrations of patient-reported outcomes are comparable.[9,18]

CONCLUSION

This study demonstrates that the OSTRC-P questionnaire has an acceptable level of diagnostic accuracy for self-report of patellar tendinopathy and patellar tendinopathy severity. This questionnaire provides a low-cost alternative to clinical evaluation for detecting patellar tendinopathy and grading its severity in large-scale surveillance studies and athlete care. These findings have implications for future development of primary and secondary prevention strategies for tendinopathy in youth, school, and community basketball settings. ◉

KEY POINTS

FINDINGS: The authors adapted (and tested against a physical therapist's clinical evaluation) a questionnaire that was previously developed for any knee problem to specifically identify patellar tendinopathy in youth basketball players. This study found that the adapted questionnaire, the Oslo Sports Trauma Research Center-patellar tendinopathy (OSTRC-P) questionnaire, is a viable alternative to a physical therapist's clinical evaluation in detecting patellar tendinopathy.

IMPLICATIONS: The OSTRC-P questionnaire offers the opportunity for close monitoring of youth basketball players for patellar tendinopathy. This tool provides the opportunity for future development of prevention and treatment strategies for tendinopathy in youth basketball.

CAUTION: Use of the OSTRC-P questionnaire in younger adolescent athletes (younger than 16 years of age) and settings other than youth basketball warrants caution and further investigation.

REFERENCES

1. Andersen CA, Clarsen B, Johansen TV, Engebretsen L. High prevalence of overuse injury among iron-distance triathletes. *Br J Sports Med*. 2013;47:857-861. https://doi.org/10.1136/bjsports-2013-092397
2. Bahr R. No injuries, but plenty of pain? On the methodology for recording overuse symptoms in sports. *Br J Sports Med*. 2009;43:966-972. https://doi.org/10.1136/bjsm.2009.066936
3. Cassel M, Baur H, Hirschmüller A, Carlsohn A, Fröhlich K, Mayer F. Prevalence of Achilles and patellar tendinopathy and their association to intratendinous changes in adolescent athletes. *Scand J Med Sci Sports*. 2015;25:e310-e318. https://doi.org/10.1111/sms.12318
4. Clarsen B, Bahr R, Heymans MW, et al. The prevalence and impact of overuse injuries in five Norwegian sports: application of a new surveillance method. *Scand J Med Sci Sports*. 2015;25:323-330. https://doi.org/10.1111/sms.12223
5. Clarsen B, Myklebust G, Bahr R. Development

and validation of a new method for the registration of overuse injuries in sports injury epidemiology: the Oslo Sports Trauma Research Centre (OSTRC) overuse injury questionnaire. *Br J Sports Med*. 2013;47:495-502. https://doi.org/10.1136/bjsports-2012-091524

6. Cohen JF, Korevaar DA, Altman DG, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open*. 2016;6:e012799. https://doi.org/10.1136/bmjopen-2016-012799
7. Cook JL, Khan KM, Kiss ZS, Griffiths L. Patellar tendinopathy in junior basketball players: a controlled clinical and ultrasonographic study of 268 patellar tendons in players aged 14-18 years. *Scand J Med Sci Sports*. 2000;10:216-220. https://doi.org/10.1034/j.1600-0838.2000.010004216.x
8. de Vries AJ, van der Worp H, Diercks RL, van den Akker-Scheek I, Zwerver J. Risk factors for patellar tendinopathy in volleyball and basketball players: a survey-based prospective cohort study. *Scand J Med Sci Sports*. 2015;25:678-684. https://doi.org/10.1111/sms.12294
9. Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review. *Value Health*. 2008;11:322-333. https://doi.org/10.1111/j.1524-4733.2007.00231.x
10. Harøy J, Clarsen B, Thorborg K, Hölmich P, Bahr R, Andersen TE. Groin problems in male soccer players are more common than previously reported. *Am J Sports Med*. 2017;45:1304-1308. https://doi.org/10.1177/0363546516687539
11. Krabak BJ, Snitily B, Milani CJ. Running injuries during adolescence and childhood. *Phys Med Rehabil Clin N Am*. 2016;27:179-202. https://doi.org/10.1016/j.pmr.2015.08.010
12. Lian ØB, Engebretsen L, Bahr R. Prevalence of jumper's knee among elite athletes from different sports: a cross-sectional study. *Am J Sports Med*. 2005;33:561-567. https://doi.org/10.1177/0363546504270454
13. Malliaras P, Cook J, Purdam C, Rio E. Patellar tendinopathy: clinical diagnosis, load management, and advice for challenging case presentations. *J Orthop Sports Phys Ther*. 2015;45:887-898. https://doi.org/10.2519/jospt.2015.5987
14. Malliaras P, Cook JL, Kent P. Reduced ankle dorsiflexion range may increase the risk of patellar tendon injury among volleyball players. *J Sci Med Sport*. 2006;9:304-309. https://doi.org/10.1016/j.jsams.2006.03.015
15. Mendonça LM, Ocarino JM, Bittencourt NF, Fernandes LM, Verhagen E, Fonseca ST. The accuracy of the VISA-P questionnaire, single-leg decline squat, and tendon pain history to identify patellar tendon abnormalities in adult athletes. *J Orthop Sports Phys Ther*. 2016;46:673-680. https://doi.org/10.2519/jospt.2016.6192
16. Menz HB. Analysis of paired data in physical therapy research: time to stop double-dipping? *J Orthop Sports Phys Ther*. 2005;35:477-478. https://doi.org/10.2519/jospt.2005.0108
17. Menz HB. Two feet, or one person? Problems associated with statistical analysis of paired data in foot and ankle medicine. *Foot*. 2004;14:2-5. https://doi.org/10.1016/S0958-2592(03)00047-6
18. Muehlhausen W, Doll H, Quadri N, et al. Equivalence of electronic and paper administration of patient-reported outcome measures: a systematic review and meta-analysis of studies conducted between 2007 and 2013. *Health Qual Life Outcomes*. 2015;13:167. https://doi.org/10.1186/s12955-015-0362-x
19. Pluim BM, Loeffen FG, Clarsen B, Bahr R, Verhagen EA. A one-season prospective study of injuries and illness in elite junior tennis. *Scand J Med Sci Sports*. 2016;26:564-571. https://doi.org/10.1111/sms.12471
20. Portney LG, Watkins MP. *Foundations of Clinical Research: Applications to Practice*. 3rd ed. Upper Saddle River, NJ: Pearson/Prentice Hall; 2008.
21. Purdam CR, Cook JL, Hopper DM, Khan KM, VIS Tendon Study Group. Discriminative ability of functional loading tests for adolescent jumper's knee. *Phys Ther Sport*. 2003;4:3-9. https://doi.org/10.1016/S1466-853X(02)00069-X
22. Rio E, van Ark M, Docking S, et al. Isometric contractions are more analgesic than isotonic contractions for patellar tendon pain: an in-season randomized clinical trial. *Clin J Sport Med*. 2017;27:253-259. https://doi.org/10.1097/JSM.0000000000000364
23. Scott A, Docking S, Vicenzino B, et al. Sports and exercise-related tendinopathies: a review of selected topical issues by participants of the second International Scientific Tendinopathy Symposium (ISTS) Vancouver 2012. *Br J Sports Med*. 2013;47:536-544. https://doi.org/10.1136/bjsports-2013-092329
24. van der Worp H, van Ark M, Roerink S, Pepping GJ, van den Akker-Scheek I, Zwerver J. Risk factors for patellar tendinopathy: a systematic review of the literature. *Br J Sports Med*. 2011;45:446-452. https://doi.org/10.1136/bjsm.2011.084079
25. van der Worp H, van Ark M, Zwerver J, van den Akker-Scheek I. Risk factors for patellar tendinopathy in basketball and volleyball players: a cross-sectional study. *Scand J Med Sci Sports*. 2012;22:783-790. https://doi.org/10.1111/j.1600-0838.2011.01308.x

DANIEL RAMSKOV, PT, PhD[1,2] • **STEN RASMUSSEN,** MD, PhD[3] • **HENRIK SØRENSEN,** PhD[1]
ERIK THORLUND PARNER, PhD[4] • **MARTIN LIND,** MD, PhD[5] • **RASMUS NIELSEN,** PT, PhD[1]

Progression in Running Intensity or Running Volume and the Development of Specific Injuries in Recreational Runners: Run Clever, a Randomized Trial Using Competing Risks

BACKGROUND: It has been proposed that training intensity and training volume are associated with specific running-related injuries. If such an association exists, secondary preventive measures could be initiated by clinicians, based on symptoms of a specific injury diagnosis.

OBJECTIVES: To test the following hypotheses: (1) a running schedule focusing on running intensity (S-I) would increase the risk of sustaining Achilles tendinopathy, gastrocnemius injuries, and plantar fasciitis compared with hypothesized volume-related injuries; and (2) a running schedule focusing on running volume (S-V) would increase the risk of sustaining patellofemoral pain syndrome, iliotibial band syndrome, and patellar tendinopathy compared with hypothesized intensity-related injuries.

METHODS: In this randomized clinical trial and etiology study, healthy recreational runners were included in a 24-week follow-up, divided into 8-week preconditioning and 16-week specific-focus training periods. Participants were randomized to 1 of 2 running schedules: S-I or S-V. The S-I group progressed the amount of high-intensity running (88% maximal oxygen consumption [VO_2max] or greater) each week, and the S-V group progressed total weekly running volume. A global positioning system watch or smartphone collected data on running. Running-related injuries were diagnosed based on a clinical examination. Estimates were reported as risk difference and 95% confidence interval (CI).

RESULTS: Of 447 runners, a total of 80 sustained an injury (S-I, n = 36; S-V, n = 44). Risk differences (95% CIs) of intensity injuries in the S-I group were −0.8% (−5.0%, 3.4%) at 2 weeks, −0.8% (−6.7%, 5.1%) at 4 weeks, −2.0% (−9.2%, 5.2%) at 8 weeks, and −5.1% (−16.5%, 6.3%) at 16 weeks. Risk differences (95% CIs) of volume injuries in the S-V group were −0.9% (−5.0%, 3.2%) at 2 weeks, −2.0% (−7.5%, 3.5%) at 4 weeks, −3.2% (−9.1%, 2.7%) at 8 weeks, and −3.4% (−13.2%, 6.2%) at 16 weeks.

CONCLUSION: No difference in risk of hypothesized intensity- and volume-specific running-related injuries exists between the 2 running schedules focused on progression in either running intensity or volume.

LEVEL OF EVIDENCE: Etiology, level 1b.
J Orthop Sports Phys Ther 2018;48(10):740-748. Epub 12 Jun 2018. doi:10.2519/jospt.2018.8062

KEY WORDS: *injury diagnoses, running, running-related injury, training variables*

Running-related injury (RRI) is a problem among people to whom running is part of a physically active lifestyle.[42] Because health-related benefits of running are the primary motivation for many runners, it is concerning that 31% of males and 18% of females who discontinue running over a 10-year period report injury as the main reason.[18,23,33]

The predominant anatomical locations of injury include the foot, lower leg, and knee, with medial tibial stress syndrome, patellofemoral pain, Achilles tendinopathy, and plantar fasciitis being the common diagnoses reported by health care professionals.[24,32,44] However, the

[1]Section for Sports Science, Department of Public Health, Aarhus University, Aarhus, Denmark. [2]Department of Physiotherapy, University College Northern Denmark, Aalborg, Denmark. [3]Orthopaedic Surgery Research Unit, Aalborg University Hospital, Aalborg, Denmark. [4]Section for Biostatistics, Department of Public Health, Aarhus University, Aarhus, Denmark. [5]Department of Orthopaedics, Aarhus University Hospital, Aarhus, Denmark. The study protocol was approved by the Danish Data Protection Agency and the Northern Region Ethics Committee, Denmark (N-20140069). The trial is registered at www.clinicaltrials.gov (NCT02349373). This trial was supported by the Graduate School of Health at Aarhus University and the Danish Rheumatism Association. The authors certify that they have no affiliations with or financial involvement in any organization or entity with a direct financial interest in the subject matter or materials discussed in the article. Address correspondence to Dr Daniel Ramskov, Selma Lagerloefsvej 2, 9220 Aalborg Øst, Denmark. E-mail: dar@ucn.dk

mechanism of these injuries remains uncertain.[20] Importantly, a runner's susceptibility to injury, as described in dynamic models of etiology, is highly dependent on running participation.[25,46] Therefore, to effectively examine the etiology of specific RRIs, it is necessary to understand the role of specific training-related variables, such as running volume and intensity, in RRI.[28]

A theoretical association between specific injuries and the training variables of running volume and running intensity has been proposed, based on epidemiological and biomechanical findings.[30] More precisely, sudden increases in running volume were hypothesized to be associated with an increased risk of patellofemoral pain, iliotibial band friction syndrome, and patellar tendinopathy, while sudden increases in running intensity were assumed to be associated with Achilles tendinopathy, plantar fasciitis, and gastrocnemius injuries.[30] This theoretical assumption was explored in a data set from an observational cohort study, concluding that a progression of 30% or greater, compared with a progression of 10% or less, in running distance over a 2-week period was associated with a 59% greater volume-related injury rate.[31] Therefore, injuries such as patellofemoral pain, iliotibial band friction syndrome, and patellar tendinopathy were labeled as volume-related injuries.

Biomechanical findings from 2 studies by Petersen et al[35,36] also contribute positively to this suggested association. First, a greater increase in load of ankle joint plantar flexors, compared with knee joint extensors, with increased running speed was observed.[35] Second, the cumulative load at the knee joint increased more with slow-speed running, compared to faster running, at a given distance.[36] Consequently, Achilles tendinopathy, plantar fasciitis, and gastrocnemius injuries were labeled as intensity-related injuries.

If these assumptions are true, then such knowledge would make an important contribution to load management in injury prevention and could improve the efficiency of both primary prevention, using online tailored advice, and secondary prevention by modifying running based on symptoms of injury.[8,10,15] However, before runners and clinicians apply preventive measures based on this knowledge, additional research to support the hypotheses surrounding volume-related and intensity-related injuries is necessary. The Run Clever trial aimed to conduct a training schedule intervention study, investigating the following hypothesis: (1) a running schedule focusing on intensity would increase the risk of sustaining Achilles tendinopathy, gastrocnemius injuries, and plantar fasciitis (ie, intensity RRIs) compared with those categorized as volume RRIs; and (2) a running schedule focusing on running volume would increase the risk of sustaining patellofemoral pain syndrome, iliotibial band syndrome, and patellar tendinopathy (ie, volume RRIs) compared with those categorized as intensity RRIs.

METHODS

The Run Clever trial was approved by the Danish Data Protection Agency and the Northern Region Ethics Committee, Denmark (N-20140069). All included participants received written and verbal information about the aim of the Run Clever trial, and provided verbal and written consent to participate. Prior to recruitment, the trial was registered at www.clinicaltrials.gov (January 23, 2015; NCT02349373) and the trial protocol was published.[38] An in-depth description of methods, intervention content, and outcome assessment can be found in the published protocol. Participants were included and followed from April 2015 to March 2016. Reporting according to the 2010 Consolidated Standards of Reporting Trials was followed, with the exception that power or sample-size calculation (item 7a) and relative effect sizes (item 17b) are not reported.[26,41]

Design

Using a randomized parallel-group design, recreational runners were randomized at enrollment to one of two 24-week interventions. The initial 8 weeks (preconditioning period) of the intervention functioned as physical preparation, with all runners following a similar running schedule. During the final 16 weeks, runners followed 1 of 2 running schedules based on their randomization: (1) running schedule intensity (S-I) focused on increasing the weekly volume of running intensity (ie, greater percentage of training at hard intensity), or (2) running schedule volume (S-V) focused on increasing total weekly running volume. Specific injury diagnoses collected as secondary outcomes were used to test the hypotheses.

Participants

Eligible participants were healthy recreational runners between 18 and 65 years of age who owned an iOS- or Android-based smartphone. A recreational runner was operationally defined as one who runs 1 to 3 times per week for at least 6 consecutive months. Runners interested in participating answered an internet-based questionnaire, distributed through social media, magazines, and announcements of the trial in running stores and clubs. Researchers accessed the submitted questionnaires and assessed participation eligibility.

Runners were excluded from participation if they reported having had an injury within the past 6 months or being pregnant, or if vigorous physical activity was contraindicated.[3] Before inclusion, researchers contacted eligible participants by telephone and provided verbal information, instructions on the rigidity of the running schedules, the weekly administration of the Oslo Sports Trauma Research Center Overuse Injury Questionnaire,[6] and the global positioning system–based data collection. At baseline, all included participants provided self-reported information on sex, age, height, weight, previous RRIs, and

running experience. Baseline height and weight were used to calculate body mass index.

Data Collection

An internet-based, trial-specific system was used to collect data. Only researchers working on the Run Clever trial had access to the back-end of the system. Through the back-end system, researchers could administer the running schedule interventions to individual participants, follow up on performed running, administer questionnaires, and review submitted questionnaires. Participants were provided with access to a personal internet-based training diary and a smartphone application. The smartphone application synchronized with the global positioning system unit in the smartphone, and data on running performed were collected and uploaded to the runner's internet-based training diary. Further, the smartphone application provided runners with real-time feedback during running on adherence to the scheduled running intensity, running volume, and rest periods, when the scheduled running involved interval running.

Interventions

Both running schedules involved running 3 times per week and followed a 4-week periodization cycle that was repeated a total of 6 times. The first week in every cycle involved a 23% progression in running volume. The second and third weeks were adaptation weeks, with 0% progression in running. The fourth week had a regression in running volume of 10%.[5] Running frequency and structure of the 4-week periodization cycle were constant during the entire 24-week intervention (8-week preconditioning and 16-week specific-focus training).

The initial 8 weeks (preconditioning) of both groups followed a similar running schedule. The beginning weekly volume of running was 15 km at an easy intensity. After the first 4-week periodization cycle, weekly running volume progressed 23%, based on the scheduled running volume in the preceding regression week, resulting in 3 km of running at moderate intensity being introduced into the running schedule. The 4-week periodization cycle was repeated 1 more time during the preconditioning period.

During the subsequent 16-week, specific-focus training period, S-V progressed the total weekly running volume and S-I progressed the proportion of weekly running at a hard intensity. Specifically, progression/regression of running in S-V consisted of a percentage change in total weekly kilometers. Progression/regression of running in S-I consisted of a percentage change in weekly kilometers at an intensity equal to or above 88% maximal oxygen consumption (VO_2max) (hard intensity). The 4-week periodization cycle was repeated 4 times during the specific-focus training period.

Individual relative running intensity was estimated using field-based maximal running tests, incorporated into the participants' running schedules. Using an estimation of VDOT (measure of running performance, arbitrary unit), running intensities categorized as easy (80% VO_2max or less), moderate (81%-87% VO_2max), and hard (88% VO_2max or greater) were prescribed.[7] Running tests were performed at baseline and every 8 weeks during follow-up. Detailed information on the content of the running schedule is presented in the protocol article.[38]

Outcome

The Oslo Sports Trauma Research Center Overuse Injury Questionnaire was screened weekly to identify participants reporting symptoms and time loss. An RRI was defined as "an injury sustained on muscles, joints, tendons, and/or bones during or after running and attributed to running. The injury must have caused a training reduction (reduced distance, intensity, frequency, etc) for at least 7 days." If a participant sustained an RRI, that individual was referred to a member of the diagnostic team of physical therapists.[5,27]

All physical therapists had volunteered to be part of a diagnostic team working without payment. Before the trial began, all members of the diagnostic team received information on the trial's purpose and were introduced to a standardized examination procedure and the diagnostic criteria to be used.[32] If diagnosis of the injured participant was impossible and the physical therapist deemed diagnostic imaging a necessity, then the participant was referred to the Department of Orthopaedics at Aarhus University Hospital (Aarhus, Denmark), or to the Department of Orthopaedics at Aalborg University Hospital (Aalborg, Denmark) for further examination and diagnostic imaging.

Power

In the design of the Run Clever trial,[38] power was calculated based on the primary hypothesis and, accordingly, the prespecified between-group risk difference of the primary outcome, RRI. The results of this primary objective have been published elsewhere.[39] The hypotheses investigated in the present study concerned secondary categorical outcomes, including specific RRI categories (ie, intensity and volume). Therefore, no power calculation based on the hypotheses and outcome was performed.

Randomization

Randomization was performed at inclusion, with the actual difference in intervention presenting itself after 8 weeks of running. The random allocation of participants to the S-I and S-V groups was applied through the trial-specific back-end system. In blocks of 10, the back-end system randomly allocated participants to 1 of the 2 schedules. Group allocation was concealed from investigators until after inclusion. Members of the diagnostic team were blinded to group allocation, though they were informed of the aim of the Run Clever trial.[38]

Statistical Methods

The primary analytical approach was an instrumental variable analysis used

to control for noncompliance, with randomization as the instrument.[13] Secondary analyses were based on intention-to-treat and per-protocol principles. Proportional completion of 80% or more of the scheduled running sessions was the percentage cutoff and was used to dichotomize compliance in the instrumental variable to control for confounding due to noncompliance. All participants who started a running session in the specific-focus training period were included in the intention-to-treat analysis. The per-protocol analysis only included compliant participants, utilizing the same percentage cutoff of 80% or greater.

Using the pseudo-observation method, a competing-risk analysis applying the Aalen-Johansen estimator was used to test the hypotheses.[21,34,37] Five categories of end points were included as competing risks in the analysis: intensity-related injuries, volume-related injuries, Taunton RRI (injury diagnoses presented by Taunton et al,[44] but not included in the intensity category or volume category), other RRI (RRIs matching the time-loss definition but not 1 of the previous 3 categories), and non-RRI (non-RRIs causing a permanent cessation of running). All end-point categories were included in the analysis as possible outcomes, and the Aalen-Johansen estimator accounted for the occurrence of 1 end point, disqualifying the participant's possibility of experiencing a secondary end point.[29]

Data were analyzed using follow-up time in weeks and days as the time scales. Differences in risk were analyzed at 2, 4, 8, and 16 weeks (14, 28, 56, and 112 days, respectively) after the beginning of the specific-focus training period.

Censoring, withdrawal of a participant at a specific time point before injury occurred and including only time at risk until the time point of censoring, was utilized in the analysis.[2,29] Using different time scales allowed for censoring on a daily or weekly basis.[29] Participants were right-censored in case of pregnancy, disease, lack of motivation, unwillingness to attend the clinical examination in case of injury, and when more than 10% of all training sessions were uploaded manually or at end of follow-up, whichever came first.

Estimates were presented as cumulative risk difference (risks were multiplied over time to take censoring into account) with 95% confidence interval (CI), and differences were considered statistically significant at $P<.05$. All analyses were performed using Stata/SE Version 13 (StataCorp LLC, College Station, TX).

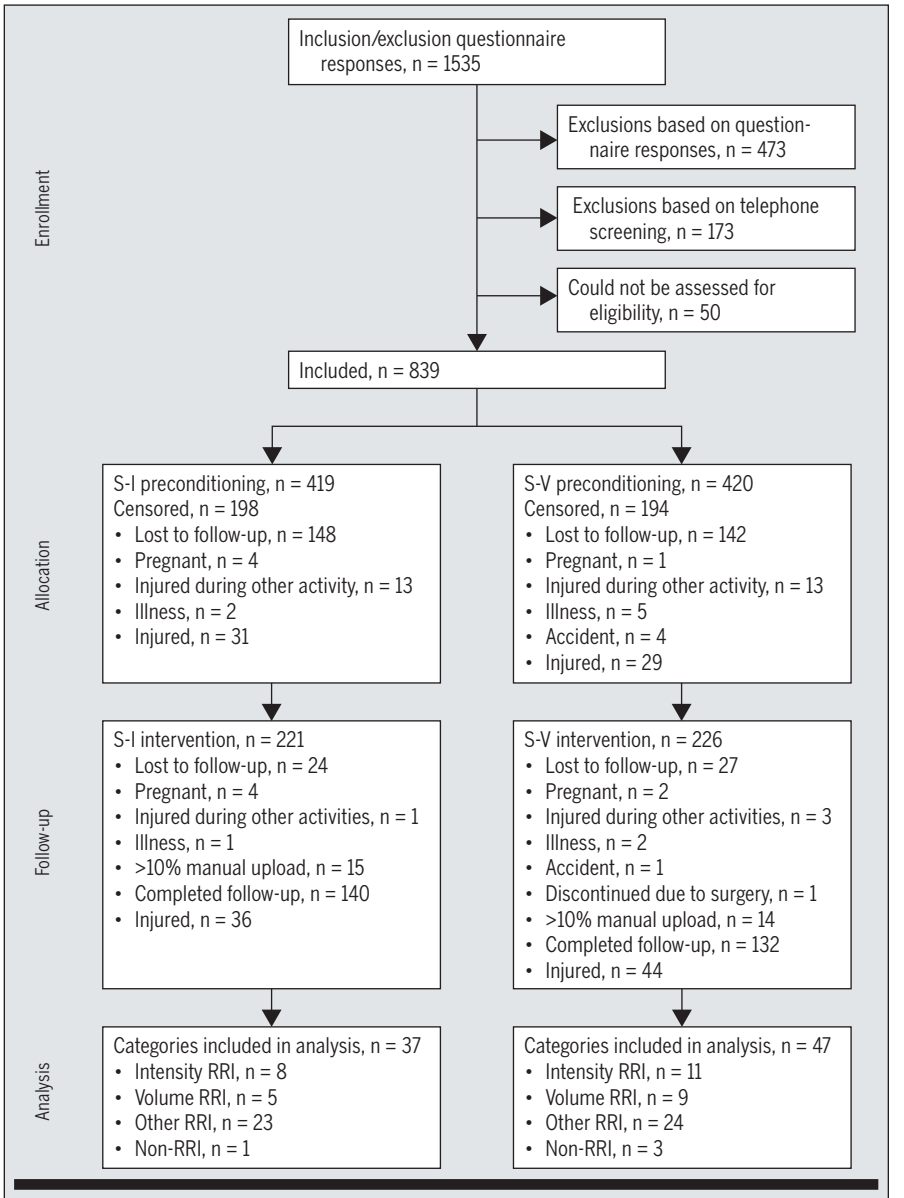


FIGURE 1. Flow chart of participants. Reasons for censoring during follow-up are listed. Injured are the number of events. The category "other RRI" is a combination of 2 categories: Taunton RRI (n = 37) and other RRI (n = 10). Abbreviations: RRI, running-related injury; S-I, schedule intensity; S-V, schedule volume.

RESULTS

Of the 1535 persons assessed for eligibility, 839 were included and allocated to S-I (n = 419) and S-V (n = 420) groups (**FIGURE 1**). During preconditioning, 198 participants from the S-I group and 194 participants from the S-V group were censored. Of the participants who completed the preconditioning period, 140 (63%) of 221 participants

TABLE 1 Baseline Characteristics*

	Preconditioning		Specific-Focus Training	
Participant Information	**S-I (n = 419)**	**S-V (n = 420)**	**S-I (n = 221)**	**S-V (n = 226)**
Sex, n (%)				
Female	261 (62)	260 (62)	136 (62)	138 (61)
Male	158 (38)	160 (38)	85 (38)	88 (39)
Age, y	39.1 ± 10.4	39.1 ± 10.1	39.6 ± 10.3	39.4 ± 9.6
BMI, kg/m^2	24.1 ± 2.9	24.2 ± 3.1	24.0 ± 2.9	24.3 ± 2.9
Previous injury, n (%)				
No	190 (45)	194 (46)	104 (47)	107 (47)
Yes	229 (55)	226 (54)	117 (53)	119 (53)
Running experience, y†	6 (3-10)	5 (3-10)	7 (3-15)	6 (3-12)
General activity level, min/wk†‡	2110 (1224-3480)	2302.5 (1327.5-3672)	2390 (1465-3520)	2372 (1450-3573)
Preconditioning compliance§	NA	NA	87% ± 16%	89% ± 19%

Abbreviations: BMI, body mass index; NA, not available; S-I, schedule intensity; S-V, schedule volume.
**Values are mean ± SD unless otherwise indicated.*
†Values are median (interquartile range).
‡Assessed using the Short QUestionnaire to ASsess Health-enhancing physical activity.[45]
§Completed sessions divided by scheduled sessions during the preconditioning period.

TABLE 2 Running-Related Injuries*

Diagnosis	Total (n = 80)	S-I (n = 36)	S-V (n = 44)	Category
MTSS	10 (12.5)	7 (19.4)	3 (6.8)	Taunton RRI
Achilles tendinopathy	8 (10.0)	3 (8.3)	5 (11.4)	Intensity RRI
Gastrocnemius injury	8 (10.0)	2 (5.6)	6 (13.6)	Intensity RRI
PFP	7 (8.7)	4 (11.1)	3 (6.8)	Volume RRI
ITBS	6 (7.5)	1 (2.8)	5 (11.4)	Volume RRI
Gluteus medius injury	6 (7.5)	2 (5.6)	4 (9.1)	Taunton RRI
Hamstring injury	5 (6.3)	3 (8.3)	2 (4.5)	Taunton RRI
Meniscal injury	4 (5.0)	1 (2.8)	3 (6.8)	Taunton RRI
Ankle inversion injury	4 (5.0)	3 (8.3)	1 (2.3)	Taunton RRI
Instability problem	3 (3.7)	1 (2.8)	2 (4.5)	Other RRI
Quadriceps injury	3 (3.7)	0 (0)	3 (6.8)	Other RRI
Plantar fasciitis	3 (3.7)	3 (8.3)	0 (0)	Intensity RRI
Greater trochanteric bursitis	2 (2.5)	1 (2.8)	1 (2.3)	Taunton RRI
Iliopsoas injury	2 (2.5)	2 (5.6)	0 (0)	Taunton RRI
Lumbar spine injury	2 (2.5)	0 (0)	2 (4.5)	Taunton RRI
Pes anserinus injury	2 (2.5)	1 (2.8)	1 (2.3)	Other RRI
Hallux valgus	1 (1.3)	0 (0)	1 (2.3)	Other RRI
Peroneal tendinopathy	1 (1.3)	1 (2.8)	0 (0)	Taunton RRI
Stress fracture, femur	1 (1.3)	1 (2.8)	0 (0)	Taunton RRI
Stress fracture, metatarsal	1 (1.3)	0 (0)	1 (2.3)	Other RRI
Patellar tendinopathy	1 (1.3)	0 (0)	1 (2.3)	Volume RRI

Abbreviations: ITBS, iliotibial band syndrome; MTSS, medial tibial stress syndrome; PFP, patellofemoral pain; RRI, running-related injury; S-I, schedule intensity; S-V, schedule volume.
**Diagnosed during the specific-focus training period. Values are n (%).*

in the S-I group and 132 (58%) of the 226 participants in the S-V group completed the 24-week follow-up. The number of noninjured participants who were censored during the 16-week, specific-focus training period for reasons other than end of follow-up was 45 in the S-I group and 50 in the S-V group. A total of 231 participants completed 80% or more of all scheduled running sessions (S-I, n = 110; S-V, n = 121). Baseline characteristics are presented in **TABLE 1**.

Running-Related Injuries

A total of 80 (S-I, n = 36; S-V, n = 44) participants sustained an RRI during the specific-focus training period. Of the 80 injured participants, 41.3% sustained an RRI categorized either as intensity or volume related. In the S-I group, there were 8 (22.2%) intensity RRIs and 5 (13.9%) volume RRIs. In the S-V group, there were 11 (25%) intensity RRIs and 9 (20.5%) volume RRIs. The frequency of RRI diagnoses is presented in **TABLE 2**.

Risk of Specific Injuries

Results from the competing-risk analysis are presented in **TABLES 3** and **4**. Estimated

risk differences indicated no difference in risk of intensity RRI and volume RRI between runners in the S-I and S-V groups. Results from the intention-to-treat and per-protocol analyses are presented in **APPENDICES A** and **B** (available at www.jospt.org). The Aalen-Johansen curves visualizing the cause-specific cumulative injury proportion as a function of follow-up time in weeks and days are presented in **FIGURE 2**.

DISCUSSION

This is the first prospective study to use a randomized design to compare the effect of running intensity progression and running volume progression on the risk of specific RRIs in recreational runners. The hypothesis that running schedules focused on progression in intensity or volume would result in more runners sustaining intensity- or volume-related injuries was not supported by estimates of risk differences over the study period.

These results contrast the proposed association, the exploratory prospective study findings, and mechanistic biomechanical findings.[30,31,35,36] However, possible reasons for this discrepancy could be (1) the periodization of the running schedules, (2) the scheduled running intensities, or (3) the categorizations of injuries.

The Periodization of Running Schedules

The importance of managing training load to avoid fatigue, illness, and injury is well recognized and was considered in the current trial when providing participants with running schedules.[9,43] Ethical considerations concerning appropriate training-load progression supported the choice of a 4-week progression cycle. To allow participants time to adapt to the increased training, a theoretical step-loading approach was chosen and implemented using a 4-week periodization cycle.[4] Based on findings by Buist et al,[5] a progression in running volume of 23% between 2 weeks was incorporated into the step-loading approach, followed by 2 adaptation weeks and a regression week.

Recently, the acute-chronic workload ratio has received much attention as a method of predicting injury likelihood.[11] Applying the acute-chronic workload ratio to the progression in the specific-focus training period of emphasized training variables in both the S-I and S-V groups resulted in an acute-chronic workload ratio between 0.8 and 1.2, the range in which injury risk is considered low. Therefore, the change in training load (progression in running intensity or running volume) might not have been rapid or sudden enough to apply a combination of stress and frequency that would result in the proposed association.[19,30] The definition of a rapid or sudden change in training load is probably not uniform across sports, and few studies within running have investigated how much of a progression is too much. At present, the limit for a biweekly progression is no greater than 30%.[5,11,31]

TABLE 3 Risk of Intensity RRI*

Analysis Time Point	Reference Risk†	RD‡	*P* Value
2 wk	1.4	−0.8 (−5.0, 3.4)	.70
14 d	1.4	−1.9 (−5.7, 1.9)	.32
4 wk	2.5	−0.8 (−6.7, 5.1)	.79
28 d	2.0	−1.9 (−6.7, 2.8)	.43
8 wk	3.7	−2.0 (−9.2, 5.1)	.57
56 d	3.1	−1.0 (−7.7, 5.9)	.80
16 wk	7.5	−5.1 (−16.5, 6.3)	.38
112 d	6.4	−3.0 (−14.0, 7.6)	.58

Abbreviations: RD, risk difference; RRI, running-related injury; S-I, schedule intensity; S-V, schedule volume.
**Values are absolute percentage points unless otherwise indicated. When testing difference in risk of injuries related to intensity (S-I), S-V was used as the reference group.*
†The risk of sustaining an injury among runners randomized to S-V.
‡Values in parentheses are 95% confidence interval. A negative RD is protective, whereas a positive value is more injurious.

TABLE 4 Risk of Volume RRI*

Analysis Time Point	Reference Risk†	RD‡	*P* Value
2 wk	1.4	−0.9 (−5.0, 3.2)	.65
14 d	1.4	−1.0 (−5.2, 3.0)	.61
4 wk	2.5	−2.0 (−7.5, 3.5)	.47
28 d	2.0	−2.1 (−6.7, 2.5)	.36
8 wk	3.1	−3.2 (−9.1, 2.7)	.29
56 d	3.1	−3.3 (−9.3, 2.6)	.27
16 wk	5.3	−3.4 (−13.2, 6.2)	.49
112 d	5.5	−6.5 (−15.0, 2.0)	.14

Abbreviations: RD, risk difference; RRI, running-related injury; S-I, schedule intensity; S-V, schedule volume.
**Values are absolute percentage points unless otherwise indicated. When testing difference in risk of injuries related to volume (S-V), S-I was used as the reference group.*
†The risk of sustaining an injury among runners randomized to S-I.
‡Values in parentheses are 95% confidence interval. A negative RD is protective, whereas a positive value is more injurious.

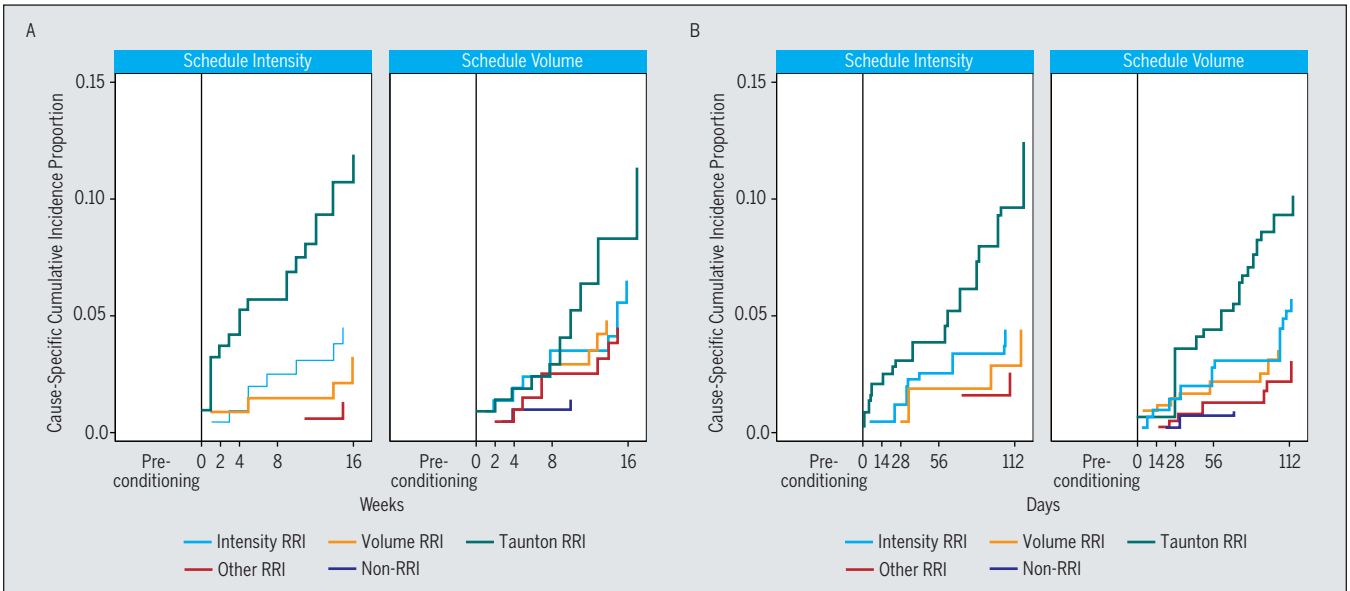


FIGURE 2. Aalen-Johansen curves visualizing the cause-specific cumulative incidence proportion in schedule intensity and schedule volume, as a function of the time scales of follow-up: (A) weeks as time scale (2, 4, 8, and 16 weeks), and (B) days as time scale (14, 28, 56, and 112 days). No differences in risk of intensity and volume RRIs exist between running schedules that focused on intensity or volume progression. Abbreviation: RRI, running-related injury.

The Intensity of Running

Absolute joint load and muscle force contributions increase with increased running speed. Further, with increased running speed, knee and ankle peak joint moments differ in increments at specific running speeds, possibly resulting in differences in structure-specific loads.[14,35] However, at running speeds below 12 km/h, differences in peak joint moments of the ankle and knee seem to be less pronounced.[1,14,35]

To accommodate individual differences in fitness levels, the scheduling of running intensity in the current trial was based on a measure of relative intensity.[12] Despite the interventional focus of S-I being a progression of a hard relative running intensity, only 8% of running sessions averaged an absolute intensity of 12 km/h or faster. Because all running sessions consisted of running at both easy and hard intensities, averaging absolute intensity might obscure the true amount of running performed at 12 km/h or faster. It is possible that only a small proportion of recreational runners achieve an average absolute running intensity of 12 km/h or faster.[16,22]

Surprisingly, 10% of the running sessions in the S-V group averaged an absolute running intensity of 12 km/h or faster. The reason for this result may be a compliance problem among participants in the S-I group, due to unfamiliarity with running at higher speeds and executing interval training or to beliefs related to excessive speed or pace; however, these explanations are speculative.[17,40]

Injury Categorization and RRI Diagnoses

Of the 80 RRIs, only 4 were caused by an isolated identifiable event related to running (eg, ankle distortion caused by a fall due to an icy path) and categorized as acute injuries. The remaining 76 RRIs were assumed to be related to overuse and were so categorized based on the diagnosis. The RRI diagnoses included in the intensity and volume categories have consistently been reported as some of the most frequent RRIs, emphasizing the importance of both primary and secondary preventive measures targeting these RRIs.[24,31,44] Only 1 study in the running-injury literature, Nielsen et al,[31] categorized volume and intensity injuries in a way similar to that applied in the current trial. However, an important difference may be the categories used and the diagnoses within the different categories in the present study. Nielsen et al[31] utilized 2 additional categories: traumatic injury and other overuse injury, whereas the current trial utilized Taunton and non-RRI categories, as stated a priori, with the addition of the other RRI category.[31,38]

Further, as stated a priori, the current trial only included Achilles tendinopathy, gastrocnemius injuries, plantar fasciitis, patellar tendinopathy, patellofemoral pain, and iliotibial band syndrome as outcomes of interest. The total proportion of volume and intensity injuries in the current trial, therefore, was 41.3%. Of these, intensity RRIs were more frequent (23.8%) than volume RRIs (17.5%). This contrasts the findings of Nielsen et al,[31] in which additional diagnoses were categorized as an intensity RRI (tibial stress fracture, hamstring injury, iliopsoas injury) and a volume RRI (medial tibial stress syndrome, gluteus medius injury, greater trochanteric bursitis, tensor fascia latae injury). This expanded diagnosis inclusion resulted in a total proportion of 66.3% of all RRIs being volume and

intensity injuries, with volume injuries (37.6%) being more frequent than intensity injuries (28.7%). Two sensitivity analyses have been performed to categorize injuries (1) as done by Nielsen et al[31] and (2) by anatomical region (**APPENDICES C** and **D**, available at www.jospt.org).

Limitations and Future Considerations

The large number of participants censored during preconditioning is a limitation. The purpose of preconditioning was to increase sample homogeneity related to the most recent running performed prior to the specific-focus training period. However, the possibility of unmeasured random confounding is increased, and external validity is affected by the many participants being censored due to lack of motivation (lost to follow-up). Further, the low number of hypothesized intensity and volume injuries included as outcomes of interest should also be considered, as it affects the power of the study and possibly introduces a type II error.

The diagnostic criteria employed by the diagnostic team might also have caused a misclassification of injuries, because the standardized examination and diagnostic criteria are not validated. However, members of the diagnostic team were experienced musculoskeletal physical therapists. Further, both the standardized examination and diagnostic criteria could be considered evidence-based best practice.

The appropriateness of the study design should also be considered. Allocating study participants to running schedules with large progressions in running volume or intensity, disregarding what is considered appropriate training-load progression, cannot be justified. Therefore, an observational prospective cohort might be a sounder study design in future research.

There is a need for future biomechanical studies investigating changes in structure-specific loads in sagittal, frontal, and transverse planes of movement. In addition, these studies should incorporate comparable isolated changes in running speed and duration.

CONCLUSION

The respective focus on intensity or volume of the running schedules did not result in more recreational runners sustaining intensity-related or volume-related RRIs. Therefore, the findings indicate that no difference in risk of intensity- and volume-specific RRIs exists between the 2 running schedules focused on intensity or volume progression. ◉

KEY POINTS

FINDINGS: The progressions of running intensity and running volume were not associated with a difference in the frequency of specific running-related injuries (RRIs) between groups.

IMPLICATIONS: When choosing to modify training load to reduce the risk of a specific RRI, clinicians should consider modifying both running intensity and running volume, as the currently available evidence is both basic and exploratory.

CAUTION: The low number of RRIs hypothesized to be associated with running intensity and running volume should be considered, as it affects the power of this study. The applicability of a randomized design should also be considered, due to the ethical aspect of prescribing running schedules with changes in training load that are currently considered unsafe.

ACKNOWLEDGMENTS: *The authors thank the team of physical therapists who helped examine injured runners, Martin Elkjær from Help2Run for his assistance, all members of RUNSAFE who spent time talking with participants, the Danish Rheumatism Association for its contribution, and all the recreational runners who participated in this study.*

REFERENCES

1. Arampatzis A, Brüggemann GP, Metzler V. The effect of speed on leg stiffness and joint kinetics in human running. *J Biomech*. 1999;32:1349-1353. https://doi.org/10.1016/S0021-9290(99)00133-5
2. Bahr R, Holme I. Risk factors for sports injuries—a methodological approach. *Br J Sports Med*. 2003;37:384-392. https://doi.org/10.1136/bjsm.37.5.384
3. Balady GJ, Chaitman B, Driscoll D, et al. Recommendations for cardiovascular screening, staffing, and emergency policies at health/fitness facilities. *Circulation*. 1998;97:2283-2293.
4. Bompa TO. *Periodization Training for Sports*. Champaign, IL: Human Kinetics; 1999.
5. Buist I, Bredeweg SW, van Mechelen W, Lemmink KA, Pepping GJ, Diercks RL. No effect of a graded training program on the number of running-related injuries in novice runners: a randomized controlled trial. *Am J Sports Med*. 2008;36:33-39. https://doi.org/10.1177/0363546507307505
6. Clarsen B, Myklebust G, Bahr R. Development and validation of a new method for the registration of overuse injuries in sports injury epidemiology: the Oslo Sports Trauma Research Centre (OSTRC) Overuse Injury Questionnaire. *Br J Sports Med*. 2013;47:495-502. https://doi.org/10.1136/bjsports-2012-091524
7. Daniels J. *Daniels' Running Formula*. 2nd ed. Champaign, IL: Human Kinetics; 2005.
8. Drew MK, Cook J, Finch CF. Sports-related workload and injury risk: simply knowing the risks will not prevent injuries. *Br J Sports Med*. 2016;50:1306-1308. https://doi.org/10.1136/bjsports-2015-095871
9. Drew MK, Finch CF. The relationship between training load and injury, illness and soreness: a systematic and literature review. *Sports Med*. 2016;46:861-883. https://doi.org/10.1007/s40279-015-0459-8
10. Esculier JF, Bouyer LJ, Dubois B, et al. Is combining gait retraining or an exercise programme with education better than education alone in treating runners with patellofemoral pain? A randomised clinical trial. *Br J Sports Med*. 2018;52:659-666. https://doi.org/10.1136/bjsports-2016-096988
11. Gabbett TJ. The training–injury prevention paradox: should athletes be training smarter *and* harder? *Br J Sports Med*. 2016;50:273-280. https://doi.org/10.1136/bjsports-2015-095788
12. Garber CE, Blissmer B, Deschenes MR, et al. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. *Med Sci Sports Exerc*. 2011;43:1334-1359. https://doi.org/10.1249/MSS.0b013e318213fefb
13. Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol*. 2000;29:722-729. https://doi.org/10.1093/ije/29.4.722
14. Hamner SR, Delp SL. Muscle contributions to fore-aft and vertical body mass center accelerations over a range of running speeds. *J Biomech*. 2013;46:780-787. https://doi.org/10.1016/j.jbiomech.2012.11.024
15. Hespanhol LC, Jr., van Mechelen W, Verhagen E.

Effectiveness of online tailored advice to prevent running-related injuries and promote preventive behaviour in Dutch trail runners: a pragmatic randomised controlled trial. *Br J Sports Med*. 2018;52:851-858. https://doi.org/10.1136/bjsports-2016-097025

16. Hespanhol Junior LC, Costa LO, Carvalho AC, Lopes AD. A description of training characteristics and its association with previous musculoskeletal injuries in recreational runners: a cross-sectional study. *Rev Bras Fisioter*. 2012;16:46-53. https://doi.org/10.1590/S1413-35552012000100009
17. Hespanhol Junior LC, Costa LO, Lopes AD. Previous injuries and some training characteristics predict running-related injuries in recreational runners: a prospective cohort study. *J Physiother*. 2013;59:263-269. https://doi.org/10.1016/S1836-9553(13)70203-0
18. Hespanhol Junior LC, Pillay JD, van Mechelen W, Verhagen E. Meta-analyses of the effects of habitual running on indices of health in physically inactive adults. *Sports Med*. 2015;45:1455-1468. https://doi.org/10.1007/s40279-015-0359-y
19. Hreljac A. Etiology, prevention, and early intervention of overuse injuries in runners: a biomechanical perspective. *Phys Med Rehabil Clin N Am*. 2005;16:651-667. https://doi.org/10.1016/j.pmr.2005.02.002
20. Hulme A, Nielsen RO, Timpka T, Verhagen E, Finch C. Risk and protective factors for middle- and long-distance running-related injury. *Sports Med*. 2017;47:869-886. https://doi.org/10.1007/s40279-016-0636-4
21. Kjaersgaard MI, Parner ET. Instrumental variable method for time-to-event data using a pseudo-observation approach. *Biometrics*. 2016;72:463-472.
22. Knechtle B, Barandun U, Knechtle P, Zingg MA, Rosemann T, Rüst CA. Prediction of half-marathon race time in recreational female and male runners. *Springerplus*. 2014;3:248. https://doi.org/10.1186/2193-1801-3-248
23. Koplan JP, Rothenberg RB, Jones EL. The natural history of exercise: a 10-yr follow-up of a cohort of runners. *Med Sci Sports Exerc*. 1995;27:1180-1184.
24. Lopes AD, Hespanhol LC, Jr., Yeung SS, Costa LO. What are the main running-related musculoskeletal injuries? A systematic review. *Sports Med*. 2012;42:891-905. https://doi.org/10.2165/11631170-000000000-00000
25. Meeuwisse WH, Tyreman H, Hagel B, Emery C. A dynamic model of etiology in sport injury: the recursive nature of risk and causation. *Clin J Sport Med*. 2007;17:215-219. https://doi.org/10.1097/JSM.0b013e3180592a48
26. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. *J Clin Epidemiol*. 2010;63:e1-e37. https://doi.org/10.1016/j.jclinepi.2010.03.004
27. Nielsen RO, Buist I, Parner ET, et al. Predictors of running-related injuries among 930 novice runners: a 1-year prospective follow-up study. *Orthop J Sports Med*. 2013;1:2325967113487316. https://doi.org/10.1177/2325967113487316
28. Nielsen RO, Buist I, Sørensen H, Lind M, Rasmussen S. Training errors and running related injuries: a systematic review. *Int J Sports Phys Ther*. 2012;7:58-75.
29. Nielsen RØ, Malisoux L, Møller M, Theisen D, Parner ET. Shedding light on the etiology of sports injuries: a look behind the scenes of time-to-event analyses. *J Orthop Sports Phys Ther*. 2016;46:300-311. https://doi.org/10.2519/jospt.2016.6510
30. Nielsen RO, Nohr EA, Rasmussen S, Sørensen H. Classifying running-related injuries based upon etiology, with emphasis on volume and pace. *Int J Sports Phys Ther*. 2013;8:172-179.
31. Nielsen RØ, Parner ET, Nohr EA, Sørensen H, Lind M, Rasmussen S. Excessive progression in weekly running distance and risk of running-related injuries: an association which varies according to type of injury. *J Orthop Sports Phys Ther*. 2014;44:739-747. https://doi.org/10.2519/jospt.2014.5164
32. Nielsen RO, Rønnow L, Rasmussen S, Lind M. A prospective study on time to recovery in 254 injured novice runners. *PLoS One*. 2014;9:e99877. https://doi.org/10.1371/journal.pone.0099877
33. Nielsen RO, Videbaek S, Hansen M, Parner ET, Rasmussen S, Langberg H. Does running with or without diet changes reduce fat mass in novice runners? A 1-year prospective study. *J Sports Med Phys Fitness*. 2016;56:105-113.
34. Parner ET, Andersen PK. Regression analysis of censored data using pseudo-observations. *Stata J*. 2010;10:408-422.
35. Petersen J, Nielsen RO, Rasmussen S, Sørensen H. Comparisons of increases in knee and ankle joint moments following an increase in running speed from 8 to 12 to 16 km·h^{-1}. *Clin Biomech (Bristol, Avon)*. 2014;29:959-964. https://doi.org/10.1016/j.clinbiomech.2014.09.003
36. Petersen J, Sørensen H, Nielsen RØ. Cumulative loads increase at the knee joint with slow-speed running compared to faster running: a biomechanical study. *J Orthop Sports Phys Ther*. 2015;45:316-322. https://doi.org/10.2519/jospt.2015.5469
37. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: competing risks and multi-state models. *Stat Med*. 2007;26:2389-2430.
38. Ramskov D, Nielsen RO, Sørensen H, Parner E, Lind M, Rasmussen S. The design of the Run Clever randomized trial: running volume, -intensity and running-related injuries. *BMC Musculoskelet Disord*. 2016;17:177. https://doi.org/10.1186/s12891-016-1020-0
39. Ramskov D, Rasmussen S, Sørensen H, Parner ET, Lind M, Nielsen RO. Run Clever – no difference in risk of injury when comparing progression in running volume and running intensity in recreational runners: a randomised trial. *BMJ Open Sport Exerc Med*. 2018;4:e000333. https://doi.org/10.1136/bmjsem-2017-000333
40. Saragiotto BT, Yamato TP, Lopes AD. What do recreational runners think about risk factors for running injuries? A descriptive study of their beliefs and opinions. *J Orthop Sports Phys Ther*. 2014;44:733-738. https://doi.org/10.2519/jospt.2014.5710
41. Schulz KF, Altman DG, Moher D, Consort Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c332. https://doi.org/10.1136/bmj.c332
42. Shipway R, Holloway I. Health and the running body: notes from an ethnography. *Int Rev Sociol Sport*. 2016;51:78-96. https://doi.org/10.1177/1012690213509807
43. Soligard T, Schwellnus M, Alonso JM, et al. How much is too much? (Part 1) International Olympic Committee consensus statement on load in sport and risk of injury. *Br J Sports Med*. 2016;50:1030-1041. https://doi.org/10.1136/bjsports-2016-096581
44. Taunton JE, Ryan MB, Clement DB, McKenzie DC, Lloyd-Smith DR, Zumbo BD. A retrospective case-control analysis of 2002 running injuries. *Br J Sports Med*. 2002;36:95-101. https://doi.org/10.1136/bjsm.36.2.95
45. Wendel-Vos GC, Schuit AJ, Saris WH, Kromhout D. Reproducibility and relative validity of the Short QUestionnaire to ASsess Health-enhancing physical activity. *J Clin Epidemiol*. 2003;56:1163-1169. https://doi.org/10.1016/S0895-4356(03)00220-8
46. Windt J, Gabbett TJ. How do training and competition workloads relate to injury? The workload–injury aetiology model. *Br J Sports Med*. 2017;51:428-435. https://doi.org/10.1136/bjsports-2016-096040

APPENDIX A

RISK OF INTENSITY- AND VOLUME-SPECIFIC RRIS: WEEKS AS TIME SCALE*

Time/Risk/Analysis	Reference Risk†	RD‡	*P* Value
Week 2			
Risk of intensity injuries			
ITT	1.4	−0.4 (−2.5, 1.6)	.68
PP	0.8	−0.8 (−2.5, 0.9)	.34
IV	1.4	−0.8 (−5.0, 3.4)	.70
Risk of volume injuries			
ITT	0.9	0.5 (−1.6, 2.5)	.66
PP	0.0	0.8 (−0.9, 2.5)	.34
IV	1.4	−0.9 (−5.0, 3.2)	.65
Week 4			
Risk of intensity injuries			
ITT	2.4	−0.4 (−3.3, 2.4)	.77
PP	2.6	−1.7 (−5.1, 1.8)	.34
IV	2.5	−0.8 (−6.7, 5.1)	.79
Risk of volume injuries			
ITT	1.4	1.0 (−1.7, 3.6)	.48
PP	0.0	0.8 (−0.9, 2.5)	.34
IV	2.5	−2.0 (−7.5, 3.5)	.47
Week 8			
Risk of intensity injuries			
ITT	3.5	−1.0 (−4.4, 2.4)	.56
PP	4.3	−2.5 (−7.1, 2.1)	.28
IV	3.7	−2.0 (−9.2, 5.1)	.57
Risk of volume injuries			
ITT	1.4	1.5 (−1.3, 4.4)	.30
PP	0.0	1.7 (−0.7, 4.2)	.17
IV	3.1	−3.2 (−9.1, 2.7)	.29
Week 16			
Risk of intensity injuries			
ITT	6.6	−2.1 (−7.1, 3.0)	.42
PP	8.4	−3.5 (−10.4, 3.5)	.33
IV	7.5	−5.1 (−16.5, 6.3)	.38
Risk of volume injuries			
ITT	3.2	1.6 (−2.7, 5.9)	.45
PP	1.0	3.4 (−1.1, 7.9)	.13
IV	5.3	−3.4 (−13.2, 6.2)	.49

Abbreviations: ITT, intention to treat; IV, instrumental variable; PP, per protocol; RD, risk difference; RRI, running-related injury; S-I, schedule intensity; S-V, schedule volume.

**Values are absolute percentage points unless otherwise indicated. When testing difference in risk of injuries related to intensity (S-I), S-V was used as the reference group. When testing difference in risk of injuries related to volume (S-V), S-I was used as the reference group.*

†The risk of sustaining an injury among runners randomized to either the S-V or S-I group.

‡Values in parentheses are 95% confidence interval. A negative RD is protective, whereas a positive value is more injurious.

APPENDIX B

RISK OF INTENSITY- AND VOLUME-SPECIFIC RRIS: DAYS AS TIME SCALE*

Time/Risk/Analysis	Reference Risk†	RD‡	*P* Value
Day 14			
Risk of intensity injuries			
ITT	1.4	−1.0 (−2.8, 0.9)	.32
PP	0.8	−0.8 (−2.5, 0.9)	.34
IV	1.4	−1.9 (−5.7, 1.9)	.32
Risk of volume injuries			
ITT	0.9	0.5 (−1.5, 2.5)	.62
PP	0.0	0.8 (−0.9, 2.6)	.33
IV	1.4	−1.0 (−5.2, 3.0)	.61
Day 28			
Risk of intensity injuries			
ITT	1.9	−1.0 (−3.3, 1.4)	.42
PP	1.7	−1.7 (−4.1, 0.7)	.17
IV	2.0	−1.9 (−6.7, 2.8)	.43
Risk of volume injuries			
ITT	0.9	1.0 (−1.2, 3.3)	.38
PP	0.0	0.8 (−0.9, 2.6)	.33
IV	2.0	−2.1 (−6.7, 2.5)	.36
Day 56			
Risk of intensity injuries			
ITT	3.0	−0.5 (−3.7, 2.8)	.78
PP	3.5	−1.6 (−5.9, 2.6)	.45
IV	3.1	−1.0 (−7.7, 5.9)	.80
Risk of volume injuries			
ITT	1.4	1.6 (−1.3, 4.4)	.28
PP	0.0	1.7 (−0.7, 4.2)	.17
IV	3.1	−3.3 (−9.3, 2.6)	.27
Day 112			
Risk of intensity injuries			
ITT	5.7	−1.2 (−5.9, 3.5)	.61
PP	7.3	−2.4 (−8.9, 4.2)	.48
IV	6.4	−3.0 (−14.0, 7.6)	.58
Risk of volume injuries			
ITT	2.1	2.8 (−1.0, 6.7)	.15
PP	1.0	3.5 (−1.1, 8.0)	.13
IV	5.5	−6.5 (−15.0, 2.0)	.14

Abbreviations: ITT, intention to treat; IV, instrumental variable; PP, per protocol; RD, risk difference; RRI, running-related injury; S-I, schedule intensity; S-V, schedule volume.
**Values are absolute percentage points unless otherwise indicated. When testing difference in risk of injuries related to intensity (S-I), S-V was used as the reference group. When testing difference in risk of injuries related to volume (S-V), S-I was used as the reference group.*
†The risk of sustaining an injury among runners randomized to either the S-V or S-I group.
‡Values in parentheses are 95% confidence interval. A negative RD is protective, whereas a positive value is more injurious.

APPENDIX C

SENSITIVITY ANALYSIS OF RISK OF INTENSITY- AND VOLUME-SPECIFIC RRIS, ACCORDING TO NIELSEN ET AL[31]*

	Week 2			Week 4			Week 8			Week 16		
Risk/Analysis	Reference Risk†	RD‡	*P* Value	Reference Risk†	RD‡	*P* Value	Reference Risk†	RD‡	*P* Value	Reference Risk†	RD‡	*P* Value
Risk of intensity injuries												
ITT	1.4	0.0 (−2.2, 2.3)	.98	2.9	0.0 (−3.3, 3.3)	.99	4.0	−0.5 (−4.3, 3.2)	.78	7.7	−0.4 (−6.1, 5.4)	.90
PP	0.8	−0.8 (−2.5, 0.9)	.34	3.4	−2.5 (−6.4, 1.3)	.20	5.2	−3.4 (−8.3, 1.5)	.17	10.2	−2.3 (−10.2, 5.6)	.57
IV	1.4	0.1 (−4.5, 4.6)	.98	3.0	0.1 (−6.7, 6.9)	.98	4.2	−1.2 (−9.0, 6.6)	.77	8.6	−1.6 (−14.3, 11.1)	.81
Risk of volume injuries												
ITT	4.2	−1.9 (−5.2, 1.4)	.26	5.2	−1.4 (−5.4, 2.6)	.49	6.4	−1.4 (−5.9, 3.1)	.53	8.3	5.7 (−3.9, 14.9)	.23
PP	0.9	0.0 (−2.5, 2.3)	.95	0.9	0.8 (−2.2, 3.8)	.60	2.8	0.7 (−4.0, 5.3)	.78	4.1	11.4 (−1.0, 23.8)	.07
IV	2.3	3.9 (−2.9, 10.7)	.26	3.9	2.8 (−5.4, 10.9)	.50	5.1	2.8 (−6.5, 12.1)	.56	15.4	−13.8 (−34.7, 7.0)	.19
	Day 14			**Day 28**			**Day 56**			**Day 112**		
Risk of intensity injuries												
ITT	1.4	−0.5 (−2.6, 1.6)	.65	1.9	0.0 (−2.7, 2.8)	.99	3.5	0.0 (−3.6, 3.6)	.99	6.9	0.5 (−5.0, 6.0)	.86
PP	0.8	−0.8 (−2.5, 0.9)	.34	1.7	−1.7 (−4.1, 0.7)	.17	4.3	−2.5 (−7.1, 2.1)	.28	9.1	−1.2 (−8.7, 6.4)	.77
IV	1.4	−1.0 (−5.2, 3.3)	.66	2.0	0.1 (−5.4, 5.6)	.98	3.7	0.0 (−7.6, 7.6)	1.00	7.7	0.5 (−11.6, 12.7)	.93
Risk of volume injuries												
ITT	3.7	−1.4 (−4.6, 1.8)	.39	4.2	−1.4 (−4.9, 2.1)	.43	5.2	−0.9 (−5.0, 3.3)	.68	7.1	3.6 (−2.3, 9.5)	.23
PP	1.0	−0.1 (−2.5, 2.4)	.96	1.0	−0.1 (−2.5, 2.4)	.96	0.1	1.7 (−1.8, 5.2)	.33	3.9	7.0 (−0.1, 14.1)	.05
IV	2.3	2.8 (−3.7, 9.4)	.40	2.8	2.8 (−4.3, 9.9)	.44	4.5	1.6 (−6.9, 10.1)	.71	11.7	−8.8 (−21.4, 3.7)	.17

Abbreviations: ITT, intention to treat; IV, instrumental variable; PP, per protocol; RD, risk difference; RRI, running-related injury; S-I, schedule intensity; S-V, schedule volume.

**Values are absolute percentage points unless otherwise indicated. When testing difference in risk of injuries related to intensity (S-I), S-V was used as the reference group. When testing difference in risk of injuries related to volume (S-V), S-I was used as the reference group. Five end-point categories are included in this analysis: volume-related injuries (patellofemoral pain, iliotibial band syndrome, medial tibial stress syndrome, patellar tendinopathy, gluteus medius injury, greater trochanteric bursitis, and injury to the tensor fascia latae), intensity-related injuries (plantar fasciitis, Achilles tendinopathy, tibial stress fracture, hamstring injuries, iliopsoas injuries, and injuries to the triceps surae muscles), other overuse injuries (medial meniscus, other stress fractures, and other overuse injuries not included in the volume-related or intensity-related injury categories), acute injuries (ankle inversion injuries), and injuries not related to running.*

†The risk of sustaining an injury among runners randomized to either the S-I or S-V group.

‡Values in parentheses are 95% confidence interval. A negative RD is protective, whereas a positive value is more injurious.

APPENDIX D

SENSITIVITY ANALYSIS OF RISK OF INTENSITY- AND VOLUME-SPECIFIC RRIS, ACCORDING TO ANATOMICAL REGION*

	Week 2			Week 4			Week 8			Week 16		
Risk/Analysis	Reference Risk†	RD‡	*P* Value	Reference Risk†	RD‡	*P* Value	Reference Risk†	RD‡	*P* Value	Reference Risk†	RD‡	*P* Value
Risk of intensity injuries												
ITT	1.8	1.4 (−1.5, 4.4)	.34	2.9	2.5 (−1.3, 6.3)	.21	4.0	2.5 (−1.9, 6.8)	.26	9.0	−0.5 (−6.6, 5.5)	.86
PP	0.8	0.1 (−2.3, 2.5)	.95	2.6	0.2 (−4.1, 4.4)	.94	4.3	0.3 (−5.2, 5.7)	.92	11.2	−3.5 (−11.6, 4.6)	.40
IV	1.8	3.0 (−3.1, 9.1)	.34	2.9	5.1 (−2.9, 13.0)	.21	4.1	5.0 (−4.1, 14.0)	.28	10.0	−2.7 (−15.9, 10.6)	.69
Risk of volume injuries												
ITT	1.4	0.0 (−2.2, 2.2)	.99	2.4	1.5 (−1.9, 4.9)	.37	2.4	3.2 (−0.7, 7.0)	.10	6.9	3.1 (−3.0, 9.1)	.32
PP	0.0	0.8 (−0.9, 2.5)	.34	0.0	3.5 (0.0, 6.9)	.05	0.1	6.1 (1.6, 10.7)	.008	5.1	6.4 (−1.1, 13.9)	.10
IV	1.4	0.0 (−4.5, 4.4)	.99	4.1	−3.2 (−10.2, 3.7)	.36	5.8	−6.8 (−14.7, 1.2)	.10	10.8	−6.3 (−19.6, 7.0)	.35
	Day 14			**Day 28**			**Day 56**			**Day 112**		
Risk of intensity injuries												
ITT	1.9	1.0 (−1.9, 3.8)	.51	2.4	1.5 (−1.9, 4.8)	.39	3.4	2.5 (−1.6, 6.6)	.24	7.5	1.0 (−4.7, 6.7)	.73
PP	0.8	0.1 (−2.3, 2.5)	.95	1.7	0.1 (−3.3, 3.6)	.95	3.5	0.2 (−4.7, 5.1)	.94	9.2	−1.4 (−9.0, 6.1)	.71
IV	1.9	2.0 (−3.9, 7.9)	.51	2.4	3.0 (−3.8, 9.9)	.39	3.6	5.0 (−3.6, 13.6)	.25	8.3	1.1 (−11.3, 13.4)	.87
Risk of volume injuries												
ITT	1.4	0.0 (−2.2, 2.3)	.97	1.9	0.6 (−2.2, 3.3)	.69	2.4	2.7 (−1.0, 6.4)	.17	5.9	4.2 (−1.5, 10.0)	.15
PP	0.0	1.0 (−0.9, 2.6)	.33	0.0	1.7 (−0.7, 4.2)	.17	0.0	5.3 (1.0, 9.5)	.01	5.2	6.8 (−1.2, 14.0)	.10
IV	1.4	−0.1 (−4.6, 4.4)	.96	2.5	−1.2 (−6.9, 4.5)	.68	5.3	−5.7 (−13.4, 2.0)	.15	11.0	−9.1 (−22.0, 3.6)	.16

Abbreviations: ITT, intention to treat; IV, instrumental variable; PP, per protocol; RD, risk difference; RRI, running-related injury; S-I, schedule intensity; S-V, schedule volume.

**Values are absolute percentage points unless otherwise indicated. When testing difference in risk of injuries sustained in the foot, ankle, or lower leg and related to intensity, S-I used S-V as the reference group. When testing difference in risk of injuries sustained in the thigh or knee and related to volume, S-V used S-I as the reference group. Six end-point categories are included in this analysis: foot/ankle/lower leg (Achilles tendinopathy, medial tibial stress syndrome, peroneal tendinopathy, plantar fasciitis, stress fracture-metatarsal, and injuries to the triceps surae muscles), thigh/knee (patellofemoral pain, hamstring injury, quadriceps injury, meniscal injury, pes anserinus injury, iliotibial band syndrome, patellar tendinopathy, and stress fracture-femur), hip (gluteus medius injury, greater trochanteric bursitis, and iliopsoas injury), acute injuries (ankle inversion injuries), other RRI (instability problem, hallux valgus, and lumbal columna injury), and injuries not related to running.*

†*The risk of sustaining an injury among runners randomized to either the S-I or S-V group.*

‡*Values in parentheses are 95% confidence interval. A negative RD is protective, whereas a positive value is more injurious.*

COLETTE RIDEHALGH, PT, PhD[1] • OLIVER P. SANDY-HINDMARCH, MSc[2] • ANNINA B. SCHMID, PT, PhD[2]

Validity of Clinical Small-Fiber Sensory Testing to Detect Small–Nerve Fiber Degeneration

BACKGROUND: Small–nerve fiber, or small-fiber, degeneration commonly occurs in patients with peripheral neuropathies, resulting in a deterioration of nerve function. Currently, the gold standard to identify small-fiber degeneration is through skin biopsy. Simple clinical tests aim to identify small-fiber degeneration, but their validity remains unknown.

OBJECTIVES: To examine the validity of clinical tests to assess small–nerve fiber degeneration, using carpal tunnel syndrome as a model neuropathy.

METHODS: One hundred seven participants (22 healthy, 85 with carpal tunnel syndrome) in this prospective, cross-sectional diagnostic accuracy study underwent pinprick testing of the index finger and were assessed for cold detection threshold and warm detection threshold using quantitative sensory testing. In a subgroup of patients with carpal tunnel syndrome (n = 51), cold and warm sensations were also tested, using coins at room and body temperature, respectively. The validity of these clinical tests was established against intraepidermal nerve fiber density measured in skin biopsies from the index finger.

RESULTS: Optimal validity occurred with clusters of tests. Specifically, normal warm or cold sensation is highly sensitive to rule out small-fiber degeneration (sensitivity, 0.98; 95% confidence interval [CI]: 0.85, 0.99), but has a low specificity (0.20; 95% CI: 0.03, 0.52). By contrast, a reduction in pinprick is highly specific (0.88; 95% CI: 0.72, 0.95), and so can be used to rule in small-fiber degeneration. For quantitative sensory testing, the highest specificity (0.83) occurs for warm detection threshold and the highest sensitivity (0.84; 95% CI: 0.72, 0.91) for cold detection threshold or warm detection threshold.

CONCLUSION: Pinprick testing, followed by warm and cold tests if pinprick is normal, is a valid and cost-effective method to detect small-fiber degeneration. For quantitative sensory testing, warm detection threshold is useful for ruling in small-fiber degeneration. To rule out small-fiber degeneration, both cold detection threshold and warm detection threshold must be negative.

LEVEL OF EVIDENCE: Diagnosis, level 2. *J Orthop Sports Phys Ther 2018;48(10):767-774. Epub 22 Jun 2018. doi:10.2519/jospt.2018.8230*

KEY WORDS: *bedside sensory testing, peripheral neuropathy, sensitivity, small-fiber degeneration, specificity*

Patients with suspected peripheral neuropathies such as carpal tunnel syndrome, radiculopathy, or diabetic neuropathy are often examined clinically to determine the extent of nerve function loss. Traditionally, such exams have included light touch, muscle strength, and reflex testing.[1,9] These tests focus solely on the function of large-diameter nerve fibers; however, recent work suggests that small-diameter fibers are often affected in peripheral neuropathies[18,26] and may deteriorate before a compromise in large-fiber function becomes apparent.[21,24]

Small-fiber pathology can be attributed to metabolic, neurotoxic, immune-mediated, or genetic factors; however, the underlying cause remains unknown in 30% to 50% of patients.[25] Importantly, neglecting to examine small-fiber function may result in underreporting sensory changes in people with suspected peripheral neuropathies. The examination of small fibers is, therefore, an integral part of the recently published grading system for neuropathic pain.[7]

The determination of intraepidermal nerve fiber density from skin biopsy is the gold standard for evaluating structural small-fiber degeneration.[4,13] However, this method is invasive and requires access to a specialized histology facility that often is not readily available in clinical practice.

The functional properties of small fibers can be assessed with quantitative sensory testing.[20] Specifically, warm detection threshold, cold detection threshold, and the perception of pinprick stimuli are known to examine the function of the C- and A-delta fibers.[20] These tests use specialized equipment that allows quantification of the respective detection thresholds. The warm and cold detection thresholds were found to

[1]School of Health Sciences, University of Brighton, Eastbourne, United Kingdom. [2]Nuffield Department of Clinical Neurosciences, Oxford University, Oxford, United Kingdom. The London Riverside Ethics Committee approved this study (REC reference 10/H0706/35). Dr Schmid was supported by a Neil Hamilton Fairley Fellowship from the National Health and Medical Research Council Australia (APP1053058) and an advanced postdoctoral mobility fellowship from the Swiss National Science Foundation (P00P3-158835). This research was supported by the National Institute for Health Research (NIHR), Oxford Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the Swiss National Science Foundation, the NIHR, or the Department of Health. The authors certify that they have no affiliations with or financial involvement in any organization or entity with a direct financial interest in the subject matter or materials discussed in the article. Address correspondence to Dr Annina B. Schmid, Nuffield Department of Clinical Neurosciences, Oxford University, Oxford OX39DU UK. E-mail: annina.schmid@neuro-research.ch

correlate with intraepidermal nerve fiber loss in patients with peripheral nerve injury of the lower limb.[22] However, the clinimetric properties of these tests were not analyzed.

Another study[11] described the clinical features of patients with small-fiber neuropathy of unknown diagnosis using intraepidermal nerve fiber density and some quantitative sensory testing parameters. While examining the diagnostic accuracy of quantitative sensory testing was not the aim of the study, it is possible to calculate the sensitivity and specificity from the data provided in the manuscript. Sensitivity and specificity of cold detection thresholds compared to intraepidermal nerve fiber density collected from the foot, were 0.7 and 0.71, respectively. Only 27 participants were included though, and only cold detection threshold was examined, thus limiting the interpretation of these results.[11] Further research is required to investigate the validity of quantitative sensory testing parameters in comparison to intraepidermal nerve fiber density as a gold standard of small-fiber degeneration.

Quantitative sensory testing allows a detailed analysis of small-fiber function; however, the required equipment is costly and often not available to clinicians outside specialized health centers. Therefore, simple bedside tests to determine small-fiber degeneration are warranted. To the authors' knowledge, only 1 study has evaluated the validity of simple bedside small-fiber tests in patients with peripheral neuropathies.[14] These data suggest frequent disagreement between bedside sensory tests and standardized quantitative sensory testing in patients with partial peripheral nerve injury. To date, no study has investigated the validity of simple bedside sensory tests to detect structural small-fiber degeneration as determined by skin biopsies.

Thus, the current study aimed to determine the validity of simple bedside tests for small-fiber function (warm/cold detection using a coin at body/room temperature, and pinprick using a Neurotip [Owen Mumford Ltd, Woodstock, UK] and toothpick), and of quantitative sensory tests (warm and cold detection thresholds). This was achieved in individuals with carpal tunnel syndrome as a model system, because carpal tunnel syndrome is the most common peripheral neuropathy and a proportion of patients have small-fiber degeneration.[21] Intraepidermal nerve fiber density was used as a gold standard of structural small-fiber degeneration.

Specific research objectives were (1) to assess the validity of clinical bedside tests for identifying small-fiber degeneration in people with carpal tunnel syndrome and in healthy individuals, and (2) to assess the validity of warm and cold detection thresholds for identifying small-fiber degeneration in people with carpal tunnel syndrome and in healthy individuals.

METHODS

Design

This study is a prospective, cross-sectional diagnostic accuracy study and includes data from the Oxford carpal tunnel syndrome cohort from May 2013 to August 2016. Data on a subgroup of this cohort have been published previously.[1,21] The present study was approved by the London Riverside Ethics Committee (REC reference 10/H0706/35). All participants provided written informed consent prior to participating, and their rights were protected.

Participants and Centers

Eighty-five consecutive participants with clinical signs and symptoms of carpal tunnel syndrome were recruited from the neurophysiology and hand surgery departments of participating hospitals, as well as through advertisements on local notice boards. The presence of isolated carpal tunnel syndrome with electrodiagnostic tests for the median nerve was confirmed using the grading scale proposed by Bland.[2] In the presence of bilateral carpal tunnel syndrome (73% of patients), the authors included the subjectively more severe hand, which was identified by asking the patient which hand he or she thought was more severely affected. Exclusion criteria included the presence of peripheral neuropathies other than carpal tunnel syndrome, any other medical condition that could affect the upper extremity or neck, and a history of previous surgery or significant trauma to the upper limb or neck. Patients who were pregnant or had diabetes were also excluded.

Twenty-two healthy participants who satisfied eligibility criteria and agreed to participate were included in this study. Participants with subclinical abnormalities upon electrodiagnostic testing of any of the evaluated nerves (median, ulnar, radial) were excluded from the study. All participants attended a single appointment at the Nuffield Department of Clinical Neurosciences, University of Oxford. All testing procedures were performed by the same researcher, who had 16 years of experience and specialized in musculoskeletal physical therapy.

Bedside Clinical Tests

Several simple and cost-effective clinical tests were used to evaluate small-fiber function. First, a Neurotip (Owen Mumford Ltd) was used to establish the ability to detect sharp stimuli. The Neurotip was first applied to the ventral forearm (innervated by the median nerve proximal to the carpal tunnel) and then to the palmar tip of the index finger (affected median nerve territory). The Neurotip was applied with pressure sufficient to produce blanching of the skin, but without penetration. The participants were asked whether the sharpness of these 2 stimulations was comparable. A reduction in sharpness sensation at the fingertip was rated as a reduced mechanical pain threshold. A sharper prick experience at the fingertip was recorded (n = 2), but because the researchers were specifically looking for loss of function in this test, this finding was graded as "normal" for subsequent analysis.

In a subgroup of patients (n = 51 consecutively recruited patients), a toothpick

was used to determine the ability to discriminate sharp sensations. The toothpick was gently pressed over the lateral upper arm innervated by the radial nerve, and then over the palmar aspect of the index fingertip. Participants were asked to compare the sharpness of these 2 pricks. Comparable to the Neurotip, a reduced sharpness in the fingertip was rated as a reduced mechanical pain threshold.

In the same subgroup, a metal coin was used to determine the ability to discriminate thermal sensations.[7] A coin held at room temperature was placed over the lateral upper arm. The coin was then placed over the palmar aspect of the index fingertip, and the patients were asked whether the temperature of the coin was comparable between the 2 sites. Patients were asked to compare the perceived temperature of the coin at the fingertip to that at the lateral upper arm (the same, colder, or warmer). Metal is a good heat conductor and is perceived as "cold" at room temperature.[17] Thus, a perception of "warmer or less cold" at the fingertip was rated as a deficit in cold detection.

The same procedure was repeated with a coin that was placed in the investigator's pocket for at least 30 minutes. The researchers have found, clinically, that this coin is perceived as neutral or slightly warm in a healthy population. A perception of "colder or less warm" over the palmar tip of the index finger compared to the lateral upper arm was interpreted as a deficit in warm detection.

Quantitative Sensory Testing

Quantitative sensory testing was performed for thermal detection thresholds according to a previously published protocol from the German Research Network on Neuropathic Pain.[20] Cold and warm detection thresholds were measured with a Thermotester (25 × 50-mm thermode; Somedic SenseLab AB, Sösdala, Sweden) over the palmar aspect of the index finger. Both warm and cold detection thresholds were obtained using ramped stimuli of 1°C/s, starting at 32°C, which was ceased when the participant pressed the button. Cutoff temperatures were 4°C and 50°C. Participants were familiarized with cold and warm detection thresholds prior to these thresholds being collected on the back of their hand. Measures were taken 3 times, and the mean was used for the analysis.

Skin Biopsy

Skin biopsies of 3 mm in diameter were collected from the ventroradial aspect of the proximal phalanx of the index finger following local subcutaneous anesthetics with 1% lidocaine. No adverse events were associated with the performance of skin biopsies. The skin biopsy was fixed in fresh periodate-lysine-paraformaldehyde (2%) for 30 minutes, followed by washes in 0.1 M phosphate buffer and incubation in 15% sucrose in 0.1 M phosphate buffer for 2 to 3 days. Skin samples were then embedded in optimal cutting temperature gel, frozen, and stored at −80°C. Sections of 50-μm thickness were cut on a cryostat and stained using a free-floating protocol previously described.[21] Blocking was performed with 5% fish gelatine for 1 hour before an overnight incubation with protein gene product (PGP) 9.5 as a primary antibody.

At the start of the study, the researchers used the PGP 9.5 antibody from Ultraclone Ltd (Isle of Wight, UK) (1:800), which was discontinued while the study was still running. Benchmarking of the PGP 9.5 antibody from Zytomed Systems (Berlin, Germany) (1:200) revealed comparable nerve fiber counts; therefore, this antibody was used for the remainder of the study. On the second day, sections were washed 3 times for 1 hour in 0.1% Triton X in phosphate-buffered saline (PBS), followed by an overnight incubation with the secondary antibody (Cy3; Stratech Scientific Ltd, Ely, UK) (1:1000). On the third day, sections were washed 3 times for 1 hour in PBS containing 0.1% Triton X and for 1 hour in PBS alone. Sections were mounted on glass slides for confocal analysis.

The skin samples were coded, and intraepidermal nerve fiber density was established by a blinded investigator with vast experience in the interpretation of skin biopsies (greater than 400 samples). Fibers crossing the dermal-epidermal border were counted down the microscope (Axio LSM 700; ZEISS, Oberkochen, Germany), strictly following the principles set out by published guidelines.[22] Intraepidermal nerve fiber density was expressed as fibers per millimeter epidermis, and the average of 3 sections was used for analysis.

In a previous study, the authors determined a cutoff of 7.1 or fewer fibers per mm epidermis to be highly sensitive (0.89) and specific (0.85; area under the curve [AUC] = 0.88; 95% confidence interval [CI]: 0.78, 0.98; P<.001) in discriminating patients with carpal tunnel syndrome from healthy volunteers.[21] As a result, in the present study, this same cutoff of 7.1 or fewer small fibers was used to divide the participants into those with and without small-fiber degeneration. Intratester reliability to determine intraepidermal nerve fiber density was examined by the same investigator processing and counting 23 skin samples twice. The intraclass correlation coefficient (model 2,1) revealed excellent intratester agreement for intraepidermal nerve fiber density determination, with an intraclass correlation coefficient of 0.91 (95% CI: 0.79, 0.96; P<.001).

Data Analysis

Sample size was determined using tables devised by Flahault et al[8] for diagnostic accuracy studies. For an estimated sensitivity and specificity of 0.8, these tables required a minimal sample size of 60 participants with small–nerve fiber degeneration to achieve a minimum accepted lower-bound estimate of the 95% CI of 0.60.

For electrodiagnostic test results, Kruskal-Wallis tests were conducted to assess differences between healthy participants and those with carpal tunnel syndrome. During data analysis, absent sensory and motor recordings were replaced with values of zero for amplitudes but excluded from analysis of latencies and nerve con-

duction velocities to prevent inflated results (n = 35 and n = 3, respectively).

For the bedside clinical tests, a 2-by-2 contingency table was created, and sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio, and negative likelihood ratio were calculated. Positive likelihood ratios were calculated using the formula sensitivity/(1 – specificity), and negative likelihood ratios with the formula (1 – sensitivity)/specificity. For pinprick sensitivity testing using a Neurotip in the larger cohort, further analysis was performed by dividing the sample by severity of the electrodiagnostic tests (mild, grades 1-2; moderate, grade 3; severe, grades 4-6).

In addition, logistic regression was used to identify factors most likely to predict small-fiber degeneration using an enter model. Only factors with an accepted level of accuracy were entered into the model. Age and sex were entered into the model as well as the specific tests of interest. The Hosmer-Lemeshow summary-of-goodness-of-fit statistic was used to assess the fit of the model. *P* values of less than .02 were accepted.[6]

For the quantitative sensory testing measures, receiver operating characteristic curves were drawn in SPSS Statistics Version 22 (IBM Corporation, Armonk, NY), and the AUC was calculated for both warm and cold detection thresholds. This provided a cutoff temperature that yielded the best sensitivity and specificity, and likelihood ratios were then calculated from these figures. An AUC of less than 0.5 suggested no discrimination, 0.7 to 0.8 acceptable, 0.8 to 0.9 excellent, and 0.9 and above outstanding discrimination.[12] Using these cutoff measures, validity of test combinations was then calculated.

TABLE 1 Demographic Data for All Participants*

	CTS (Neurotip, QST)	CTS (Warm/Cold Coin, Toothpick)	Asymptomatic (Neurotip, QST)
n	85	51	22
Age, y	61 ± 13	62 ± 13	46 ± 16
Sex, n			
Female	51	34	17
Male	34	17	5
Height, cm	169 ± 14	170 ± 15	168 ± 16
Weight, kg	76 ± 19	74 ± 21	70 ± 21
Duration of symptoms, mo	63 ± 86	70 ± 103	NA
Electrodiagnostic grade†	3.0 (2.0)	3.0 (1.5)	NA
1, %	9.41	9.80	
2, %	22.35	15.69	
3, %	30.59	35.29	
4, %	16.47	17.65	
5, %	17.65	17.65	
6, %	3.53	3.92	

Abbreviations: CTS, carpal tunnel syndrome; NA, not applicable; QST, quantitative sensory testing.
**Values are mean ± SD unless otherwise indicated.*
†*Values are median (interquartile range). Grades (Bland[2]): 1, very mild; 2, mild; 3, moderate; 4, severe; 5, very severe; and 6, extremely severe.*

RESULTS

Flow of Participants

Demographic data are shown in **TABLE 1**. Eighty-five participants with carpal tunnel syndrome and 22 asymptomatic participants were included in the analysis of pinprick sensitivity with the Neurotip and with quantitative sensory testing (warm detection threshold, cold detection threshold). For the subcohort examining additional bedside clinical tests, 51 participants with carpal tunnel syndrome were included. Applying the diagnostic cutoff for small-fiber degeneration, 67 of the larger cohort and 40 of the smaller cohort of patients with carpal tunnel syndrome had small-fiber degeneration. Patients with carpal tunnel syndrome had significantly altered sensory and motor electrodiagnostic test properties of the median nerve compared to healthy participants (**APPENDIX A**, available at www.jospt.org).

Validity of Bedside Clinical Small-Fiber Tests

TABLE 2 provides the sensitivity values, specificity values, positive predictive values and negative predictive values, and the likelihood ratios of the individual bedside clinical tests and clusters of tests compared to intraepidermal nerve fiber density.

The highest sensitivity values were from a positive cold or warm test (0.98; 95% CI: 0.85, 0.99), with a corresponding low negative likelihood ratio (0.14; 95% CI: 0.01, 2.10). The highest specificity values (and corresponding positive likelihood ratio) were demonstrated from pinprick using a Neurotip, with control participants included in the analysis (specificity, 0.88; 95% CI: 0.72, 0.95; positive likelihood ratio = 3.9; 95% CI: 1.7, 9.3).

APPENDIX B (available at www.jospt.org) reports the sensitivity, specificity, positive and negative predictive values, and the likelihood ratios for the larger cohort (n = 85), subgrouped by the severity of the electrodiagnostic test results. The highest specificity was seen in participants with more severe nerve conduction changes, whereas the highest sensitivity was apparent in participants with less severe nerve conduction changes.

Logistic Regression

Entering the bedside clinical tests into a logistic regression model, single tests or clusters of tests were not significant (*P*>.02), apart from pinprick using a Neurotip (**TABLE 3**). Age and sex entered into the model

TABLE 2 Validity Results for Clinical Tests to Detect Small-Fiber Degeneration*

Modality	Sensitivity	Specificity	PPV	NPV	+LR	–LR
Neurotip (including healthy controls)	0.49 (0.37, 0.62)	0.88 (0.72, 0.95)	0.87 (0.71, 0.95)	0.51 (0.39, 0.63)	3.94 (1.68, 9.27)	0.58 (0.45, 0.74)
Neurotip without healthy controls	0.49 (0.37, 0.62)	0.72 (0.46, 0.89)	0.87 (0.71, 0.95)	0.28 (0.16, 0.43)	1.77 (0.81, 3.88)	0.70 (0.53, 0.93)
Cold coin	0.85 (0.69, 0.94)	0.27 (0.07, 0.61)	0.81 (0.65, 0.91)	0.33 (0.09, 0.69)	1.17 (0.80, 1.72)	0.55 (0.16, 1.90)
Warm coin	0.75 (0.58, 0.87)	0.55 (0.25, 0.82)	0.86 (0.69, 0.95)	0.38 (0.16, 0.64)	1.65 (0.84, 3.23)	0.46 (0.23, 0.91)
Toothpick	0.20 (0.10, 0.36)	0.64 (0.32, 0.88)	0.67 (0.35, 0.89)	0.18 (0.08, 0.34)	0.55 (0.20, 1.49)	1.30 (0.97, 1.60)
Cold coin or warm coin	0.98 (0.85, 0.99)	0.20 (0.03, 0.52)	0.82 (0.67, 0.91)	0.67 (0.17, 0.95)	1.19 (0.90, 1.58)	0.14 (0.01, 2.10)
Cold coin and warm coin	0.60 (0.43, 0.75)	0.64 (0.32, 0.88)	0.86 (0.66, 0.95)	0.30 (0.14, 0.53)	1.65 (0.73, 3.75)	0.63 (0.39, 1.00)
Neurotip or toothpick	0.63 (0.46, 0.77)	0.45 (0.18, 0.75)	0.81 (0.62, 0.92)	0.25 (0.10, 0.50)	1.15 (0.63, 2.07)	0.83 (0.45, 1.50)

Abbreviations: –LR, negative likelihood ratio; +LR, positive likelihood ratio; NPV, negative predictive value; PPV, positive predictive value.
**Values in parentheses are 95% confidence interval.*

did not significantly predict the presence of small-fiber degeneration (P>.1).

Quantitative Sensory Tests

The receiver operating characteristic curve analysis results for thermal quantitative sensory testing measures can be seen in **TABLE 4**, including the best cutoff values yielding the highest sensitivity and specificity results of the thermal thresholds compared to intraepidermal nerve fiber density.

To establish whether clusters of tests enhanced the validity, further testing was performed using the cutoff temperature values (warm detection threshold of 35.5° with controls and 36° without, and cold detection threshold of 28.9° with controls and 28.4° without). **TABLE 5** provides the results for these combinations of tests. Warm detection threshold without controls had the best accuracy in terms of specificity and positive likelihood ratio (specificity, 0.83; positive likelihood ratio = 3.8) (**TABLE 4**). Cold detection threshold or warm detection threshold with control participants had the best sensitivity and negative likelihood ratio (sensitivity, 0.84; 95% CI: 0.72, 0.91; negative likelihood ratio = 0.26; 95% CI: 0.15, 0.47).

TABLE 3 Logistic Regression Analysis for Pinprick Predicting Small-Fiber Degeneration*

	B Coefficient	Odds Ratio
Constant	–0.029 (–0.52, 0.45)	
Intervention	1.92 (1.00, 3.46)	6.8 (2.4, 19.5)

R^2 = 0.14 (Cox and Snell), 0.19 (Nagelkerke); χ^2 = 16.2, P<.01. Values in parentheses are 95% confidence interval.

TABLE 4 Receiver Operating Characteristic Curve Analysis of QST Measures

QST	Cutoff, °C	AUC*	Sensitivity	Specificity	+LR	–LR
WDT with controls	35.5	0.79 (0.70, 0.88)	0.70	0.77	3.00	0.40
WDT without controls	36.0	0.72 (0.59, 0.86)	0.63	0.83	3.80	0.44
CDT with controls	28.9	0.74 (0.64, 0.84)	0.64	0.72	2.26	0.50
CDT without controls	28.4	0.56 (0.42, 0.69)	0.46	0.72	1.64	0.75

Abbreviations: AUC, area under the curve; CDT, cold detection threshold; –LR, negative likelihood ratio; +LR, positive likelihood ratio; QST, quantitative sensory testing; WDT, warm detection threshold.
**Values in parentheses are 95% confidence interval.*

DISCUSSION

Small–nerve fiber loss and dysfunction is apparent in individuals with peripheral neuropathies, including those with carpal tunnel syndrome and diabetes,[3,18,21,23,24] and may even precede abnormalities detectable by large-fiber tests, such as nerve conduction tests.[21,24] Therefore, tests for small-fiber function (pinprick, warm and cold sensation) may be considered an essential aspect of neurological testing. The primary aim of this study was to determine whether clinical bedside tests can accurately detect small-fiber degeneration. This study evaluated intraepidermal nerve fiber density in skin biopsies and used a previously ascertained cutoff value to establish whether small-fiber degeneration was present.

Using intraepidermal nerve fiber density as a gold standard, the results suggested that none of the clinical tests yielded sufficient validity in isolation. However, the presence of reduced pinprick sensation exhibited an accuracy of 88% to detect small-fiber degeneration. When pinprick sensation was normal, further testing with a cold and a warm coin was required. When cold and warm

TABLE 5 Validity of Combinations of WDT and CDT to Detect Small-Fiber Degeneration*

QST Combinations	Sensitivity	Specificity	PPV	NPV	+LR	–LR
CDT or WDT (with healthy controls)	0.84 (0.72, 0.91)	0.62 (0.45, 0.76)	0.79 (0.68, 0.88)	0.69 (0.51, 0.83)	2.18 (1.50, 3.30)	0.26 (0.15, 0.47)
CDT or WDT (without healthy controls)	0.75 (0.63, 0.88)	0.65 (0.39, 0.85)	0.89 (0.78, 0.96)	0.39 (0.22, 0.59)	2.13 (1.10, 4.10)	0.39 (0.24, 0.62)

Abbreviations: CDT, cold detection threshold; –LR, negative likelihood ratio; +LR, positive likelihood ratio; NPV, negative predictive value; PPV, positive predictive value; QST, quantitative sensory testing; WDT, warm detection threshold.
**Values in parentheses are 95% confidence interval.*

tests were both normal, there was a 98% probability that there was no small-fiber degeneration.

It has been suggested that likelihood ratios, in addition to sensitivity and specificity values, are important, as they consider the pretest probability of the presence or absence of the condition.[9] The negative likelihood ratio for warm or cold tests being positive was moderate at ruling out small-fiber degeneration, supporting the high sensitivity of these tests. The positive likelihood ratio for pinprick (3.94) demonstrated a small ability to rule in small-fiber degeneration, with CIs ranging from a small and rarely important (1.68) to a moderate shift in probability (9.27). However, this may still represent an important shift in probability of having the condition[9] and, together with the results of the logistic regression, suggests that a reduction in pinprick sensation may be useful to identify small-fiber degeneration. Therefore, pinprick testing, followed by cold- and warm-coin testing when pinprick is unaffected, may be a valid and cost-effective method to detect small-fiber degeneration.

Subgroup analysis of pinprick validity, according to the electrodiagnostic test severity, revealed the highest specificity in patients with more severe electrodiagnostic findings, whereas the highest sensitivity was seen in participants with less severe electrodiagnostic changes. However, these findings should be interpreted with caution due to the relatively small number of participants in each subgroup and the large range of CIs. For instance, in the group with severe electrodiagnostic findings, only 2 patients had negative pinprick testing (as expected with increasing severity).

To the authors' knowledge, no other studies have examined the validity of bedside clinical examination compared to intraepidermal nerve fiber density as a gold standard to identify small-fiber degeneration. However, the validity of bedside tests compared to standardized quantitative sensory testing has been examined.[14] The results of that study indicated substantial disagreement between some of the clinical tests and the quantitative sensory testing measures. For example, while 76% of patients felt reduced cold sensation to a cold roller, only 50% had reduced sensation using quantitative sensory testing. In contrast, warm detection threshold showed more consistency between the 2 types of tests (69% with the warm roller compared to 71% with quantitative sensory testing). However, because Leffler and Hansson[14] focused on small-fiber dysfunction and the current study focused on structural degeneration, it is difficult to make direct comparisons.

In the current study, testing pinprick with a toothpick had poorer sensitivity and specificity than testing pinprick with a Neurotip. Combining the 2 did not enhance validity, suggesting that the method of toothpick testing used in this study may not be a useful way to assess small-fiber degeneration. There may be several reasons for this.

First, the area of comparison varied between the Neurotip and toothpick. Using a Neurotip, the perception of sharpness was compared between the index finger and the forearm. In contrast, the sharpness of the toothpick on the index finger was compared to that on the lateral upper arm. While both control regions are innervated by unaffected nerves in patients with carpal tunnel syndrome, the innervation patterns of these anatomical locations differ. The intraepidermal nerve fiber density tends to be higher in the upper arm as compared to the forearm and palm.[15,16] It could be argued that this difference in intraepidermal nerve fiber density might have led to an increased number of false-positive findings when compared to the upper arm, potentially lowering the specificity values. In addition, the Neurotip has a finer, sharper point than a toothpick, and, as such, testing with a toothpick might have resulted in a reduced sensitivity to detect subtle changes. Further studies are needed to clarify the optimal comparison sites. Until such time, the Neurotip may be preferred over the toothpick for testing.

With respect to the thermal quantitative sensory testing in the current study, the highest sensitivity values and negative likelihood ratios were found when combining cold and warm detection thresholds. The findings indicated that if neither cold nor warm detection thresholds were abnormal, then there was an 84% probability that there was no small-fiber degeneration. A warm detection threshold demonstrated the highest specificity, suggesting that an elevated warm detection threshold would strongly indicate small-fiber degeneration. The positive and negative likelihood ratios of these combinations of tests indicated only a small, but potentially important,

shift in the probability to rule the presence of small-fiber degeneration in or out. These results suggest that warm detection threshold may be a useful test to rule in small-fiber degeneration, whereas normal cold detection threshold and warm detection threshold may be useful in ruling out small-fiber degeneration.

The use of control participants in this study might have affected the results by yielding higher specificity values and positive likelihood ratios. While this might have inflated the values, it also indicates that warm detection threshold may be able to discriminate asymptomatic participants without small-fiber degeneration and those with carpal tunnel syndrome with degeneration.

Findings in cold detection threshold from the quantitative sensory testing battery have previously been compared to intraepidermal nerve fiber density.[11] That study did not report findings on validity, but it was possible to calculate them. The cold detection threshold was found to have a sensitivity of 0.70 (95% CI: 0.46, 0.88), specificity of 0.71 (95% CI: 0.29, 0.96), positive likelihood ratio of 2.45 (95% CI: 0.73, 8.18), and negative likelihood ratio of 0.42 (95% CI: 0.19, 0.95) in individuals with idiopathic small-fiber neuropathies. These diagnostic values for cold detection threshold are similar to those found in the current study. The current study is the first to report the findings of warm detection threshold and a combination of warm and cold detection thresholds in comparison to intraepidermal nerve fiber density in patients with peripheral neuropathy, and suggests that both warm and cold detection thresholds may be required to enhance the validity to determine small-fiber degeneration.

There are some limitations of the present study. The researchers included patients with carpal tunnel syndrome as a model neuropathy, as it is well established that a subset of these patients have small-fiber degeneration.[21] Further studies are required to establish whether our results are generalizable to other peripheral neuropathies where small-fiber degeneration has been found. Most importantly, it remains to be determined whether the suggested clinical bedside tests have comparable clinimetric properties for other commonly affected body regions, such as the feet, where innervation density is much lower compared to the hand.

The current study examined the validity of these tests; however, reliability is another essential clinimetric property to ensure that a test is appropriate for use in clinical practice. While both intertester and intratester reliability are well established for the validated quantitative sensory testing protocol,[5,10] future studies are needed to determine the reliability of the bedside clinical small-fiber tests.

CONCLUSION

Our data suggest that simple bedside tests are valid in the identification of small-fiber degeneration in individuals with carpal tunnel syndrome. A finding of reduced pinprick sensation tested with a Neurotip may indicate the presence of small-fiber degeneration. A normal pinprick sensation, however, warrants further testing with a cold and a warm coin. A similar sensation to both temperatures compared to a reference area indicates that there may be no small-fiber degeneration.

In addition, our study found that quantitative sensory testing thresholds have some validity in identifying small-fiber degeneration. If warm detection threshold is elevated, then it is highly likely that small-fiber degeneration is present. If warm detection threshold is normal, followed by a normal cold detection threshold, then it is highly unlikely that small-fiber degeneration is present. ◉

KEY POINTS

FINDINGS: The use of pinprick followed by thermal testing, or warm detection followed by cold detection thresholds, may be valid methods of assessing small-fiber degeneration.

IMPLICATIONS: As small fibers constitute a substantial proportion of peripheral nerves and have been shown to deteriorate in common peripheral neuropathies, the ability to test these fibers accurately in a clinical setting is of great importance. This study suggests that simple and cost-effective tools may be valid to detect small-fiber degeneration in a clinical setting.

CAUTION: Only participants with a diagnosis of carpal tunnel syndrome and asymptomatic participants were included in this study. Further research is needed in patients with mixed presentations or diagnoses.

ACKNOWLEDGMENTS: *The input of Professor David Bennett in the design of the Oxford carpal tunnel syndrome study and the assistance of Shemane Murtagh with the logistic regression analysis are gratefully acknowledged.*

REFERENCES

1. Baselgia LT, Bennett DL, Silbiger RM, Schmid AB. Negative neurodynamic tests do not exclude neural dysfunction in patients with entrapment neuropathies. *Arch Phys Med Rehabil*. 2017;98:480-486. https://doi.org/10.1016/j.apmr.2016.06.019
2. Bland JD. A neurophysiological grading scale for carpal tunnel syndrome. *Muscle Nerve*. 2000;23:1280-1283. https://doi.org/10.1002/1097-4598(200008)23:8<1280::AID-MUS20>3.0.CO;2-Y
3. de la Llave-Rincón AI, Fernández-de-las-Peñas C, Laguarta-Val S, et al. Increased pain sensitivity is not associated with electrodiagnostic findings in women with carpal tunnel syndrome. *Clin J Pain*. 2011;27:747-754. https://doi.org/10.1097/AJP.0b013e31821c29d3
4. Devigili G, Tugnoli V, Penza P, et al. The diagnostic criteria for small fibre neuropathy: from symptoms to neuropathology. *Brain*. 2008;131:1912-1925. https://doi.org/10.1093/brain/awn093
5. Felix ER, Widerström-Noga EG. Reliability and validity of quantitative sensory testing in persons with spinal cord injury and neuropathic pain. *J Rehabil Res Dev*. 2009;46:69-83. https://doi.org/10.1682/JRRD.2008.04.0058
6. Field A. *Discovering Statistics Using IBM SPSS Statistics*. 4th ed. London, UK: SAGE; 2013.
7. Finnerup NB, Haroutounian S, Kamerman P, et al. Neuropathic pain: an updated grading system for research and clinical practice. *Pain*. 2016;157:1599-1606. https://doi.org/10.1097/j.

pain.0000000000000492
8. Flahault A, Cadilhac M, Thomas G. Sample size calculation should be performed for design accuracy in diagnostic test studies. *J Clin Epidemiol*. 2005;58:859-862. https://doi.org/10.1016/j.jclinepi.2004.12.009
9. Fritz JM, Wainner RS. Examining diagnostic tests: an evidence-based perspective. *Phys Ther*. 2001;81:1546-1564. https://doi.org/10.1093/ptj/81.9.1546
10. Geber C, Klein T, Azad S, et al. Test-retest and interobserver reliability of quantitative sensory testing according to the protocol of the German Research Network on Neuropathic Pain (DFNS): a multi-centre study. *Pain*. 2011;152:548-556. https://doi.org/10.1016/j.pain.2010.11.013
11. Holland NR, Crawford TO, Hauer P, Cornblath DR, Griffin JW, McArthur JC. Small-fiber sensory neuropathies: clinical course and neuropathology of idiopathic cases. *Ann Neurol*. 1998;44:47-59. https://doi.org/10.1002/ana.410440111
12. Hosmer D, Lemeshow S. *Applied Logistic Regression*. 2nd ed. New York, NY: Wiley-Interscience; 2000.
13. Lauria G, Bakkers M, Schmitz C, et al. Intraepidermal nerve fiber density at the distal leg: a worldwide normative reference study. *J Peripher Nerv Syst*. 2010;15:202-207. https://doi.org/10.1111/j.1529-8027.2010.00271.x
14. Leffler AS, Hansson P. Painful traumatic peripheral partial nerve injury-sensory dysfunction profiles comparing outcomes of bedside examination and quantitative sensory testing. *Eur J Pain*. 2008;12:397-402. https://doi.org/10.1016/j.ejpain.2007.08.009
15. Ling L, Xue J, Liu Y, et al. Quantitative and morphological study of intraepidermal nerve fibre in healthy individuals. *Neurol Res*. 2015;37:974-978. https://doi.org/10.1179/1743132815Y.0000000082
16. Liu Y, Fan X, Wei Y, Piao Z, Jiang X. Intraepidermal nerve fiber density of healthy human. *Neurol Res*. 2014;36:911-914. https://doi.org/10.1179/1743132814Y.0000000377
17. Marchettini P, Marangoni C, Lacerenza M, Formaglio F. The Lindblom roller. *Eur J Pain*. 2003;7:359-364. https://doi.org/10.1016/S1090-3801(03)00052-1
18. Petersen KL, Rice FL, Farhadi M, Reda H, Rowbotham MC. Natural history of cutaneous innervation following herpes zoster. *Pain*. 2010;150:75-82. https://doi.org/10.1016/j.pain.2010.04.002
19. Petty NJ. *Neuromusculoskeletal Examination and Assessment. A Handbook for Therapists*. 4th ed. Edinburgh, UK: Elsevier/Churchill Livingstone; 2011.
20. Rolke R, Magerl W, Campbell KA, et al. Quantitative sensory testing: a comprehensive protocol for clinical trials. *Eur J Pain*. 2006;10:77-88. https://doi.org/10.1016/j.ejpain.2005.02.003
21. Schmid AB, Bland JD, Bhat MA, Bennett DL. The relationship of nerve fibre pathology to sensory function in entrapment neuropathy. *Brain*. 2014;137:3186-3199. https://doi.org/10.1093/brain/awu288
22. Schüning J, Scherens A, Haussleiter IS, et al. Sensory changes and loss of intraepidermal nerve fibers in painful unilateral nerve injury. *Clin J Pain*. 2009;25:683-690. https://doi.org/10.1097/AJP.0b013e3181a1260e
23. Sveen KA, Karimé B, Jørum E, et al. Small- and large-fiber neuropathy after 40 years of type 1 diabetes: associations with glycemic control and advanced protein glycation: the Oslo Study. *Diabetes Care*. 2013;36:3712-3717. https://doi.org/10.2337/dc13-0788
24. Tamburin S, Cacciatori C, Praitano ML, et al. Median nerve small- and large-fiber damage in carpal tunnel syndrome: a quantitative sensory testing study. *J Pain*. 2011;12:205-212. https://doi.org/10.1016/j.jpain.2010.06.010
25. Terkelsen AJ, Karlsson P, Lauria G, Freeman R, Finnerup NB, Jensen TS. The diagnostic challenge of small fibre neuropathy: clinical presentations, evaluations, and causes. *Lancet Neurol*. 2017;16:934-944. https://doi.org/10.1016/S1474-4422(17)30329-0
26. Themistocleous AC, Ramirez JD, Shillo PR, et al. The Pain in Neuropathy Study (PiNS): a cross-sectional observational study determining the somatosensory phenotype of painful and painless diabetic neuropathy. *Pain*. 2016;157:1132-1145. https://doi.org/10.1097/j.pain.0000000000000491

APPENDIX A

NEUROPHYSIOLOGY DATA OF THE MEDIAN NERVE IN PATIENTS WITH CARPAL TUNNEL SYNDROME AND HEALTHY VOLUNTEERS*

	Mild	Moderate	Severe	Controls	*P* Value
SNAP, μV	7.02 (4.57)	6.59 (6.23)	Absent	14.47 (14.96)	<.001†
Sensory NCV, m/s	36.4 (5.36)	31.83 (6.54)	NA	51.30 (8.81)	<.001‡
DML, ms	4.2 (0.49)	5.08 (0.98)	7.02 (3.02)	3.13 (0.61)	<.001§
CMAP, mV	7.53 (4.36)	5.28 (4.28)	3.54 (2.9)	7.25 (3.23)	<.001‖

Abbreviations: CMAP, compound motor action potential; DML, distal motor latency; NA, not applicable; NCV, nerve conduction velocity; SNAP, sensory nerve action potential.

**Values are median (interquartile range).*

†*Post hoc testing: asymptomatic versus mild, P = .08; asymptomatic versus moderate, P = .005.*

‡*Post hoc testing: all pairwise comparisons, P<.001.*

§*Post hoc testing: asymptomatic versus mild, P = .047; asymptomatic versus moderate, P<.001; asymptomatic versus severe, P<.001.*

‖*Post hoc testing: asymptomatic versus mild, P>.05; asymptomatic versus moderate, P>.05; asymptomatic versus severe, P<.001.*

APPENDIX B

VALIDITY RESULTS FOR PINPRICK (NEUROTIP) DIVIDED BY SEVERITY BASED ON ELECTRODIAGNOSTIC TEST GRADES*

Grade	n	Sensitivity	Specificity	PPV	NPV	+LR	–LR
Mild (grades 1-2)	27	0.60 (0.36, 0.81)	0.57 (0.18, 0.90)	0.80 (0.61, 0.91)	0.33 (0.18, 0.54)	1.40 (0.55, 3.54)	0.70 (0.30, 1.62)
Moderate (grade 3)	26	0.53 (0.29, 0.76)	0.86 (0.42, 1.00)	0.91 (0.61, 0.98)	0.40 (0.28, 0.54)	3.68 (0.57, 23.76)	0.55 (0.31, 0.97)
Severe (grades 4-6)	32	0.43 (0.25, 0.63)	1.00 (0.16, 1.00)	1.00 (1.00, 1.00)	0.11 (0.08, 0.14)	Not defined†	0.57 (0.41, 0.77)

Abbreviations: –LR, negative likelihood ratio; +LR, positive likelihood ratio; NPV, negative predictive value; PPV, positive predictive value.
**Values in parentheses are 95% confidence interval.*
†*The calculation of +LR is sensitivity/(1 – specificity), and specificity is 1; therefore, sensitivity is divided by zero and hence +LR is undefined.*

COLETTE RIDEHALGH, PT, PhD[1] • OLIVER P. SANDY-HINDMARCH, MSc[2] • ANNINA B. SCHMID, PT, PhD[2]

Validity of Clinical Small-Fiber Sensory Testing to Detect Small–Nerve Fiber Degeneration

BACKGROUND: Small–nerve fiber, or small-fiber, degeneration commonly occurs in patients with peripheral neuropathies, resulting in a deterioration of nerve function. Currently, the gold standard to identify small-fiber degeneration is through skin biopsy. Simple clinical tests aim to identify small-fiber degeneration, but their validity remains unknown.

OBJECTIVES: To examine the validity of clinical tests to assess small–nerve fiber degeneration, using carpal tunnel syndrome as a model neuropathy.

METHODS: One hundred seven participants (22 healthy, 85 with carpal tunnel syndrome) in this prospective, cross-sectional diagnostic accuracy study underwent pinprick testing of the index finger and were assessed for cold detection threshold and warm detection threshold using quantitative sensory testing. In a subgroup of patients with carpal tunnel syndrome (n = 51), cold and warm sensations were also tested, using coins at room and body temperature, respectively. The validity of these clinical tests was established against intraepidermal nerve fiber density measured in skin biopsies from the index finger.

RESULTS: Optimal validity occurred with clusters of tests. Specifically, normal warm or cold sensation is highly sensitive to rule out small-fiber degeneration (sensitivity, 0.98; 95% confidence interval [CI]: 0.85, 0.99), but has a low specificity (0.20; 95% CI: 0.03, 0.52). By contrast, a reduction in pinprick is highly specific (0.88; 95% CI: 0.72, 0.95), and so can be used to rule in small-fiber degeneration. For quantitative sensory testing, the highest specificity (0.83) occurs for warm detection threshold and the highest sensitivity (0.84; 95% CI: 0.72, 0.91) for cold detection threshold or warm detection threshold.

CONCLUSION: Pinprick testing, followed by warm and cold tests if pinprick is normal, is a valid and cost-effective method to detect small-fiber degeneration. For quantitative sensory testing, warm detection threshold is useful for ruling in small-fiber degeneration. To rule out small-fiber degeneration, both cold detection threshold and warm detection threshold must be negative.

LEVEL OF EVIDENCE: Diagnosis, level 2. *J Orthop Sports Phys Ther 2018;48(10):767-774. Epub 22 Jun 2018. doi:10.2519/jospt.2018.8230*

KEY WORDS: *bedside sensory testing, peripheral neuropathy, sensitivity, small-fiber degeneration, specificity*

Patients with suspected peripheral neuropathies such as carpal tunnel syndrome, radiculopathy, or diabetic neuropathy are often examined clinically to determine the extent of nerve function loss. Traditionally, such exams have included light touch, muscle strength, and reflex testing.[1,9] These tests focus solely on the function of large-diameter nerve fibers; however, recent work suggests that small-diameter fibers are often affected in peripheral neuropathies[18,26] and may deteriorate before a compromise in large-fiber function becomes apparent.[21,24]

Small-fiber pathology can be attributed to metabolic, neurotoxic, immune-mediated, or genetic factors; however, the underlying cause remains unknown in 30% to 50% of patients.[25] Importantly, neglecting to examine small-fiber function may result in underreporting sensory changes in people with suspected peripheral neuropathies. The examination of small fibers is, therefore, an integral part of the recently published grading system for neuropathic pain.[7]

The determination of intraepidermal nerve fiber density from skin biopsy is the gold standard for evaluating structural small-fiber degeneration.[4,13] However, this method is invasive and requires access to a specialized histology facility that often is not readily available in clinical practice.

The functional properties of small fibers can be assessed with quantitative sensory testing.[20] Specifically, warm detection threshold, cold detection threshold, and the perception of pinprick stimuli are known to examine the function of the C- and A-delta fibers.[20] These tests use specialized equipment that allows quantification of the respective detection thresholds. The warm and cold detection thresholds were found to

[1]School of Health Sciences, University of Brighton, Eastbourne, United Kingdom. [2]Nuffield Department of Clinical Neurosciences, Oxford University, Oxford, United Kingdom. The London Riverside Ethics Committee approved this study (REC reference 10/H0706/35). Dr Schmid was supported by a Neil Hamilton Fairley Fellowship from the National Health and Medical Research Council Australia (APP1053058) and an advanced postdoctoral mobility fellowship from the Swiss National Science Foundation (P00P3-158835). This research was supported by the National Institute for Health Research (NIHR), Oxford Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the Swiss National Science Foundation, the NIHR, or the Department of Health. The authors certify that they have no affiliations with or financial involvement in any organization or entity with a direct financial interest in the subject matter or materials discussed in the article. Address correspondence to Dr Annina B. Schmid, Nuffield Department of Clinical Neurosciences, Oxford University, Oxford OX39DU UK. E-mail: annina.schmid@neuro-research.ch

correlate with intraepidermal nerve fiber loss in patients with peripheral nerve injury of the lower limb.[22] However, the clinimetric properties of these tests were not analyzed.

Another study[11] described the clinical features of patients with small-fiber neuropathy of unknown diagnosis using intraepidermal nerve fiber density and some quantitative sensory testing parameters. While examining the diagnostic accuracy of quantitative sensory testing was not the aim of the study, it is possible to calculate the sensitivity and specificity from the data provided in the manuscript. Sensitivity and specificity of cold detection thresholds compared to intraepidermal nerve fiber density collected from the foot, were 0.7 and 0.71, respectively. Only 27 participants were included though, and only cold detection threshold was examined, thus limiting the interpretation of these results.[11] Further research is required to investigate the validity of quantitative sensory testing parameters in comparison to intraepidermal nerve fiber density as a gold standard of small-fiber degeneration.

Quantitative sensory testing allows a detailed analysis of small-fiber function; however, the required equipment is costly and often not available to clinicians outside specialized health centers. Therefore, simple bedside tests to determine small-fiber degeneration are warranted. To the authors' knowledge, only 1 study has evaluated the validity of simple bedside small-fiber tests in patients with peripheral neuropathies.[14] These data suggest frequent disagreement between bedside sensory tests and standardized quantitative sensory testing in patients with partial peripheral nerve injury. To date, no study has investigated the validity of simple bedside sensory tests to detect structural small-fiber degeneration as determined by skin biopsies.

Thus, the current study aimed to determine the validity of simple bedside tests for small-fiber function (warm/cold detection using a coin at body/room temperature, and pinprick using a Neurotip [Owen Mumford Ltd, Woodstock, UK] and toothpick), and of quantitative sensory tests (warm and cold detection thresholds). This was achieved in individuals with carpal tunnel syndrome as a model system, because carpal tunnel syndrome is the most common peripheral neuropathy and a proportion of patients have small-fiber degeneration.[21] Intraepidermal nerve fiber density was used as a gold standard of structural small-fiber degeneration.

Specific research objectives were (1) to assess the validity of clinical bedside tests for identifying small-fiber degeneration in people with carpal tunnel syndrome and in healthy individuals, and (2) to assess the validity of warm and cold detection thresholds for identifying small-fiber degeneration in people with carpal tunnel syndrome and in healthy individuals.

METHODS

Design

This study is a prospective, cross-sectional diagnostic accuracy study and includes data from the Oxford carpal tunnel syndrome cohort from May 2013 to August 2016. Data on a subgroup of this cohort have been published previously.[1,21] The present study was approved by the London Riverside Ethics Committee (REC reference 10/H0706/35). All participants provided written informed consent prior to participating, and their rights were protected.

Participants and Centers

Eighty-five consecutive participants with clinical signs and symptoms of carpal tunnel syndrome were recruited from the neurophysiology and hand surgery departments of participating hospitals, as well as through advertisements on local notice boards. The presence of isolated carpal tunnel syndrome with electrodiagnostic tests for the median nerve was confirmed using the grading scale proposed by Bland.[2] In the presence of bilateral carpal tunnel syndrome (73% of patients), the authors included the subjectively more severe hand, which was identified by asking the patient which hand he or she thought was more severely affected. Exclusion criteria included the presence of peripheral neuropathies other than carpal tunnel syndrome, any other medical condition that could affect the upper extremity or neck, and a history of previous surgery or significant trauma to the upper limb or neck. Patients who were pregnant or had diabetes were also excluded.

Twenty-two healthy participants who satisfied eligibility criteria and agreed to participate were included in this study. Participants with subclinical abnormalities upon electrodiagnostic testing of any of the evaluated nerves (median, ulnar, radial) were excluded from the study. All participants attended a single appointment at the Nuffield Department of Clinical Neurosciences, University of Oxford. All testing procedures were performed by the same researcher, who had 16 years of experience and specialized in musculoskeletal physical therapy.

Bedside Clinical Tests

Several simple and cost-effective clinical tests were used to evaluate small-fiber function. First, a Neurotip (Owen Mumford Ltd) was used to establish the ability to detect sharp stimuli. The Neurotip was first applied to the ventral forearm (innervated by the median nerve proximal to the carpal tunnel) and then to the palmar tip of the index finger (affected median nerve territory). The Neurotip was applied with pressure sufficient to produce blanching of the skin, but without penetration. The participants were asked whether the sharpness of these 2 stimulations was comparable. A reduction in sharpness sensation at the fingertip was rated as a reduced mechanical pain threshold. A sharper prick experience at the fingertip was recorded (n = 2), but because the researchers were specifically looking for loss of function in this test, this finding was graded as "normal" for subsequent analysis.

In a subgroup of patients (n = 51 consecutively recruited patients), a toothpick

was used to determine the ability to discriminate sharp sensations. The toothpick was gently pressed over the lateral upper arm innervated by the radial nerve, and then over the palmar aspect of the index fingertip. Participants were asked to compare the sharpness of these 2 pricks. Comparable to the Neurotip, a reduced sharpness in the fingertip was rated as a reduced mechanical pain threshold.

In the same subgroup, a metal coin was used to determine the ability to discriminate thermal sensations.[7] A coin held at room temperature was placed over the lateral upper arm. The coin was then placed over the palmar aspect of the index fingertip, and the patients were asked whether the temperature of the coin was comparable between the 2 sites. Patients were asked to compare the perceived temperature of the coin at the fingertip to that at the lateral upper arm (the same, colder, or warmer). Metal is a good heat conductor and is perceived as "cold" at room temperature.[17] Thus, a perception of "warmer or less cold" at the fingertip was rated as a deficit in cold detection.

The same procedure was repeated with a coin that was placed in the investigator's pocket for at least 30 minutes. The researchers have found, clinically, that this coin is perceived as neutral or slightly warm in a healthy population. A perception of "colder or less warm" over the palmar tip of the index finger compared to the lateral upper arm was interpreted as a deficit in warm detection.

Quantitative Sensory Testing

Quantitative sensory testing was performed for thermal detection thresholds according to a previously published protocol from the German Research Network on Neuropathic Pain.[20] Cold and warm detection thresholds were measured with a Thermotester (25 × 50-mm thermode; Somedic SenseLab AB, Sösdala, Sweden) over the palmar aspect of the index finger. Both warm and cold detection thresholds were obtained using ramped stimuli of 1°C/s, starting at 32°C, which was ceased when the participant pressed the button. Cutoff temperatures were 4°C and 50°C. Participants were familiarized with cold and warm detection thresholds prior to these thresholds being collected on the back of their hand. Measures were taken 3 times, and the mean was used for the analysis.

Skin Biopsy

Skin biopsies of 3 mm in diameter were collected from the ventroradial aspect of the proximal phalanx of the index finger following local subcutaneous anesthetics with 1% lidocaine. No adverse events were associated with the performance of skin biopsies. The skin biopsy was fixed in fresh periodate-lysine-paraformaldehyde (2%) for 30 minutes, followed by washes in 0.1 M phosphate buffer and incubation in 15% sucrose in 0.1 M phosphate buffer for 2 to 3 days. Skin samples were then embedded in optimal cutting temperature gel, frozen, and stored at −80°C. Sections of 50-μm thickness were cut on a cryostat and stained using a free-floating protocol previously described.[21] Blocking was performed with 5% fish gelatine for 1 hour before an overnight incubation with protein gene product (PGP) 9.5 as a primary antibody.

At the start of the study, the researchers used the PGP 9.5 antibody from Ultraclone Ltd (Isle of Wight, UK) (1:800), which was discontinued while the study was still running. Benchmarking of the PGP 9.5 antibody from Zytomed Systems (Berlin, Germany) (1:200) revealed comparable nerve fiber counts; therefore, this antibody was used for the remainder of the study. On the second day, sections were washed 3 times for 1 hour in 0.1% Triton X in phosphate-buffered saline (PBS), followed by an overnight incubation with the secondary antibody (Cy3; Stratech Scientific Ltd, Ely, UK) (1:1000). On the third day, sections were washed 3 times for 1 hour in PBS containing 0.1% Triton X and for 1 hour in PBS alone. Sections were mounted on glass slides for confocal analysis.

The skin samples were coded, and intraepidermal nerve fiber density was established by a blinded investigator with vast experience in the interpretation of skin biopsies (greater than 400 samples). Fibers crossing the dermal-epidermal border were counted down the microscope (Axio LSM 700; ZEISS, Oberkochen, Germany), strictly following the principles set out by published guidelines.[22] Intraepidermal nerve fiber density was expressed as fibers per millimeter epidermis, and the average of 3 sections was used for analysis.

In a previous study, the authors determined a cutoff of 7.1 or fewer fibers per mm epidermis to be highly sensitive (0.89) and specific (0.85; area under the curve [AUC] = 0.88; 95% confidence interval [CI]: 0.78, 0.98; P<.001) in discriminating patients with carpal tunnel syndrome from healthy volunteers.[21] As a result, in the present study, this same cutoff of 7.1 or fewer small fibers was used to divide the participants into those with and without small-fiber degeneration. Intratester reliability to determine intraepidermal nerve fiber density was examined by the same investigator processing and counting 23 skin samples twice. The intraclass correlation coefficient (model 2,1) revealed excellent intratester agreement for intraepidermal nerve fiber density determination, with an intraclass correlation coefficient of 0.91 (95% CI: 0.79, 0.96; P<.001).

Data Analysis

Sample size was determined using tables devised by Flahault et al[8] for diagnostic accuracy studies. For an estimated sensitivity and specificity of 0.8, these tables required a minimal sample size of 60 participants with small–nerve fiber degeneration to achieve a minimum accepted lower-bound estimate of the 95% CI of 0.60.

For electrodiagnostic test results, Kruskal-Wallis tests were conducted to assess differences between healthy participants and those with carpal tunnel syndrome. During data analysis, absent sensory and motor recordings were replaced with values of zero for amplitudes but excluded from analysis of latencies and nerve con-

duction velocities to prevent inflated results (n = 35 and n = 3, respectively).

For the bedside clinical tests, a 2-by-2 contingency table was created, and sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio, and negative likelihood ratio were calculated. Positive likelihood ratios were calculated using the formula sensitivity/(1 – specificity), and negative likelihood ratios with the formula (1 – sensitivity)/specificity. For pinprick sensitivity testing using a Neurotip in the larger cohort, further analysis was performed by dividing the sample by severity of the electrodiagnostic tests (mild, grades 1-2; moderate, grade 3; severe, grades 4-6).

In addition, logistic regression was used to identify factors most likely to predict small-fiber degeneration using an enter model. Only factors with an accepted level of accuracy were entered into the model. Age and sex were entered into the model as well as the specific tests of interest. The Hosmer-Lemeshow summary-of-goodness-of-fit statistic was used to assess the fit of the model. *P* values of less than .02 were accepted.[6]

For the quantitative sensory testing measures, receiver operating characteristic curves were drawn in SPSS Statistics Version 22 (IBM Corporation, Armonk, NY), and the AUC was calculated for both warm and cold detection thresholds. This provided a cutoff temperature that yielded the best sensitivity and specificity, and likelihood ratios were then calculated from these figures. An AUC of less than 0.5 suggested no discrimination, 0.7 to 0.8 acceptable, 0.8 to 0.9 excellent, and 0.9 and above outstanding discrimination.[12] Using these cutoff measures, validity of test combinations was then calculated.

TABLE 1 Demographic Data for All Participants*

	CTS (Neurotip, QST)	CTS (Warm/Cold Coin, Toothpick)	Asymptomatic (Neurotip, QST)
n	85	51	22
Age, y	61 ± 13	62 ± 13	46 ± 16
Sex, n			
Female	51	34	17
Male	34	17	5
Height, cm	169 ± 14	170 ± 15	168 ± 16
Weight, kg	76 ± 19	74 ± 21	70 ± 21
Duration of symptoms, mo	63 ± 86	70 ± 103	NA
Electrodiagnostic grade†	3.0 (2.0)	3.0 (1.5)	NA
1, %	9.41	9.80	
2, %	22.35	15.69	
3, %	30.59	35.29	
4, %	16.47	17.65	
5, %	17.65	17.65	
6, %	3.53	3.92	

Abbreviations: CTS, carpal tunnel syndrome; NA, not applicable; QST, quantitative sensory testing.
**Values are mean ± SD unless otherwise indicated.*
†*Values are median (interquartile range). Grades (Bland[2]): 1, very mild; 2, mild; 3, moderate; 4, severe; 5, very severe; and 6, extremely severe.*

RESULTS

Flow of Participants

Demographic data are shown in **TABLE 1**. Eighty-five participants with carpal tunnel syndrome and 22 asymptomatic participants were included in the analysis of pinprick sensitivity with the Neurotip and with quantitative sensory testing (warm detection threshold, cold detection threshold). For the subcohort examining additional bedside clinical tests, 51 participants with carpal tunnel syndrome were included. Applying the diagnostic cutoff for small-fiber degeneration, 67 of the larger cohort and 40 of the smaller cohort of patients with carpal tunnel syndrome had small-fiber degeneration. Patients with carpal tunnel syndrome had significantly altered sensory and motor electrodiagnostic test properties of the median nerve compared to healthy participants (**APPENDIX A**, available at www.jospt.org).

Validity of Bedside Clinical Small-Fiber Tests

TABLE 2 provides the sensitivity values, specificity values, positive predictive values and negative predictive values, and the likelihood ratios of the individual bedside clinical tests and clusters of tests compared to intraepidermal nerve fiber density.

The highest sensitivity values were from a positive cold or warm test (0.98; 95% CI: 0.85, 0.99), with a corresponding low negative likelihood ratio (0.14; 95% CI: 0.01, 2.10). The highest specificity values (and corresponding positive likelihood ratio) were demonstrated from pinprick using a Neurotip, with control participants included in the analysis (specificity, 0.88; 95% CI: 0.72, 0.95; positive likelihood ratio = 3.9; 95% CI: 1.7, 9.3).

APPENDIX B (available at www.jospt.org) reports the sensitivity, specificity, positive and negative predictive values, and the likelihood ratios for the larger cohort (n = 85), subgrouped by the severity of the electrodiagnostic test results. The highest specificity was seen in participants with more severe nerve conduction changes, whereas the highest sensitivity was apparent in participants with less severe nerve conduction changes.

Logistic Regression

Entering the bedside clinical tests into a logistic regression model, single tests or clusters of tests were not significant (*P*>.02), apart from pinprick using a Neurotip (**TABLE 3**). Age and sex entered into the model

TABLE 2 Validity Results for Clinical Tests to Detect Small-Fiber Degeneration*

Modality	Sensitivity	Specificity	PPV	NPV	+LR	–LR
Neurotip (including healthy controls)	0.49 (0.37, 0.62)	0.88 (0.72, 0.95)	0.87 (0.71, 0.95)	0.51 (0.39, 0.63)	3.94 (1.68, 9.27)	0.58 (0.45, 0.74)
Neurotip without healthy controls	0.49 (0.37, 0.62)	0.72 (0.46, 0.89)	0.87 (0.71, 0.95)	0.28 (0.16, 0.43)	1.77 (0.81, 3.88)	0.70 (0.53, 0.93)
Cold coin	0.85 (0.69, 0.94)	0.27 (0.07, 0.61)	0.81 (0.65, 0.91)	0.33 (0.09, 0.69)	1.17 (0.80, 1.72)	0.55 (0.16, 1.90)
Warm coin	0.75 (0.58, 0.87)	0.55 (0.25, 0.82)	0.86 (0.69, 0.95)	0.38 (0.16, 0.64)	1.65 (0.84, 3.23)	0.46 (0.23, 0.91)
Toothpick	0.20 (0.10, 0.36)	0.64 (0.32, 0.88)	0.67 (0.35, 0.89)	0.18 (0.08, 0.34)	0.55 (0.20, 1.49)	1.30 (0.97, 1.60)
Cold coin or warm coin	0.98 (0.85, 0.99)	0.20 (0.03, 0.52)	0.82 (0.67, 0.91)	0.67 (0.17, 0.95)	1.19 (0.90, 1.58)	0.14 (0.01, 2.10)
Cold coin and warm coin	0.60 (0.43, 0.75)	0.64 (0.32, 0.88)	0.86 (0.66, 0.95)	0.30 (0.14, 0.53)	1.65 (0.73, 3.75)	0.63 (0.39, 1.00)
Neurotip or toothpick	0.63 (0.46, 0.77)	0.45 (0.18, 0.75)	0.81 (0.62, 0.92)	0.25 (0.10, 0.50)	1.15 (0.63, 2.07)	0.83 (0.45, 1.50)

Abbreviations: –LR, negative likelihood ratio; +LR, positive likelihood ratio; NPV, negative predictive value; PPV, positive predictive value.
**Values in parentheses are 95% confidence interval.*

did not significantly predict the presence of small-fiber degeneration (P>.1).

Quantitative Sensory Tests

The receiver operating characteristic curve analysis results for thermal quantitative sensory testing measures can be seen in **TABLE 4**, including the best cutoff values yielding the highest sensitivity and specificity results of the thermal thresholds compared to intraepidermal nerve fiber density.

To establish whether clusters of tests enhanced the validity, further testing was performed using the cutoff temperature values (warm detection threshold of 35.5° with controls and 36° without, and cold detection threshold of 28.9° with controls and 28.4° without). **TABLE 5** provides the results for these combinations of tests. Warm detection threshold without controls had the best accuracy in terms of specificity and positive likelihood ratio (specificity, 0.83; positive likelihood ratio = 3.8) (**TABLE 4**). Cold detection threshold or warm detection threshold with control participants had the best sensitivity and negative likelihood ratio (sensitivity, 0.84; 95% CI: 0.72, 0.91; negative likelihood ratio = 0.26; 95% CI: 0.15, 0.47).

TABLE 3 Logistic Regression Analysis for Pinprick Predicting Small-Fiber Degeneration*

	B Coefficient	Odds Ratio
Constant	–0.029 (–0.52, 0.45)	
Intervention	1.92 (1.00, 3.46)	6.8 (2.4, 19.5)

R^2 = 0.14 (Cox and Snell), 0.19 (Nagelkerke); χ^2 = 16.2, P<.01. Values in parentheses are 95% confidence interval.

TABLE 4 Receiver Operating Characteristic Curve Analysis of QST Measures

QST	Cutoff, °C	AUC*	Sensitivity	Specificity	+LR	–LR
WDT with controls	35.5	0.79 (0.70, 0.88)	0.70	0.77	3.00	0.40
WDT without controls	36.0	0.72 (0.59, 0.86)	0.63	0.83	3.80	0.44
CDT with controls	28.9	0.74 (0.64, 0.84)	0.64	0.72	2.26	0.50
CDT without controls	28.4	0.56 (0.42, 0.69)	0.46	0.72	1.64	0.75

Abbreviations: AUC, area under the curve; CDT, cold detection threshold; –LR, negative likelihood ratio; +LR, positive likelihood ratio; QST, quantitative sensory testing; WDT, warm detection threshold.
**Values in parentheses are 95% confidence interval.*

DISCUSSION

Small–nerve fiber loss and dysfunction is apparent in individuals with peripheral neuropathies, including those with carpal tunnel syndrome and diabetes,[3,18,21,23,24] and may even precede abnormalities detectable by large-fiber tests, such as nerve conduction tests.[21,24] Therefore, tests for small-fiber function (pinprick, warm and cold sensation) may be considered an essential aspect of neurological testing. The primary aim of this study was to determine whether clinical bedside tests can accurately detect small-fiber degeneration. This study evaluated intraepidermal nerve fiber density in skin biopsies and used a previously ascertained cutoff value to establish whether small-fiber degeneration was present.

Using intraepidermal nerve fiber density as a gold standard, the results suggested that none of the clinical tests yielded sufficient validity in isolation. However, the presence of reduced pinprick sensation exhibited an accuracy of 88% to detect small-fiber degeneration. When pinprick sensation was normal, further testing with a cold and a warm coin was required. When cold and warm

TABLE 5 Validity of Combinations of WDT and CDT to Detect Small-Fiber Degeneration*

QST Combinations	Sensitivity	Specificity	PPV	NPV	+LR	–LR
CDT or WDT (with healthy controls)	0.84 (0.72, 0.91)	0.62 (0.45, 0.76)	0.79 (0.68, 0.88)	0.69 (0.51, 0.83)	2.18 (1.50, 3.30)	0.26 (0.15, 0.47)
CDT or WDT (without healthy controls)	0.75 (0.63, 0.88)	0.65 (0.39, 0.85)	0.89 (0.78, 0.96)	0.39 (0.22, 0.59)	2.13 (1.10, 4.10)	0.39 (0.24, 0.62)

Abbreviations: CDT, cold detection threshold; –LR, negative likelihood ratio; +LR, positive likelihood ratio; NPV, negative predictive value; PPV, positive predictive value; QST, quantitative sensory testing; WDT, warm detection threshold.
**Values in parentheses are 95% confidence interval.*

tests were both normal, there was a 98% probability that there was no small-fiber degeneration.

It has been suggested that likelihood ratios, in addition to sensitivity and specificity values, are important, as they consider the pretest probability of the presence or absence of the condition.[9] The negative likelihood ratio for warm or cold tests being positive was moderate at ruling out small-fiber degeneration, supporting the high sensitivity of these tests. The positive likelihood ratio for pinprick (3.94) demonstrated a small ability to rule in small-fiber degeneration, with CIs ranging from a small and rarely important (1.68) to a moderate shift in probability (9.27). However, this may still represent an important shift in probability of having the condition[9] and, together with the results of the logistic regression, suggests that a reduction in pinprick sensation may be useful to identify small-fiber degeneration. Therefore, pinprick testing, followed by cold- and warm-coin testing when pinprick is unaffected, may be a valid and cost-effective method to detect small-fiber degeneration.

Subgroup analysis of pinprick validity, according to the electrodiagnostic test severity, revealed the highest specificity in patients with more severe electrodiagnostic findings, whereas the highest sensitivity was seen in participants with less severe electrodiagnostic changes. However, these findings should be interpreted with caution due to the relatively small number of participants in each subgroup and the large range of CIs. For instance, in the group with severe electrodiagnostic findings, only 2 patients had negative pinprick testing (as expected with increasing severity).

To the authors' knowledge, no other studies have examined the validity of bedside clinical examination compared to intraepidermal nerve fiber density as a gold standard to identify small-fiber degeneration. However, the validity of bedside tests compared to standardized quantitative sensory testing has been examined.[14] The results of that study indicated substantial disagreement between some of the clinical tests and the quantitative sensory testing measures. For example, while 76% of patients felt reduced cold sensation to a cold roller, only 50% had reduced sensation using quantitative sensory testing. In contrast, warm detection threshold showed more consistency between the 2 types of tests (69% with the warm roller compared to 71% with quantitative sensory testing). However, because Leffler and Hansson[14] focused on small-fiber dysfunction and the current study focused on structural degeneration, it is difficult to make direct comparisons.

In the current study, testing pinprick with a toothpick had poorer sensitivity and specificity than testing pinprick with a Neurotip. Combining the 2 did not enhance validity, suggesting that the method of toothpick testing used in this study may not be a useful way to assess small-fiber degeneration. There may be several reasons for this.

First, the area of comparison varied between the Neurotip and toothpick. Using a Neurotip, the perception of sharpness was compared between the index finger and the forearm. In contrast, the sharpness of the toothpick on the index finger was compared to that on the lateral upper arm. While both control regions are innervated by unaffected nerves in patients with carpal tunnel syndrome, the innervation patterns of these anatomical locations differ. The intraepidermal nerve fiber density tends to be higher in the upper arm as compared to the forearm and palm.[15,16] It could be argued that this difference in intraepidermal nerve fiber density might have led to an increased number of false-positive findings when compared to the upper arm, potentially lowering the specificity values. In addition, the Neurotip has a finer, sharper point than a toothpick, and, as such, testing with a toothpick might have resulted in a reduced sensitivity to detect subtle changes. Further studies are needed to clarify the optimal comparison sites. Until such time, the Neurotip may be preferred over the toothpick for testing.

With respect to the thermal quantitative sensory testing in the current study, the highest sensitivity values and negative likelihood ratios were found when combining cold and warm detection thresholds. The findings indicated that if neither cold nor warm detection thresholds were abnormal, then there was an 84% probability that there was no small-fiber degeneration. A warm detection threshold demonstrated the highest specificity, suggesting that an elevated warm detection threshold would strongly indicate small-fiber degeneration. The positive and negative likelihood ratios of these combinations of tests indicated only a small, but potentially important,

shift in the probability to rule the presence of small-fiber degeneration in or out. These results suggest that warm detection threshold may be a useful test to rule in small-fiber degeneration, whereas normal cold detection threshold and warm detection threshold may be useful in ruling out small-fiber degeneration.

The use of control participants in this study might have affected the results by yielding higher specificity values and positive likelihood ratios. While this might have inflated the values, it also indicates that warm detection threshold may be able to discriminate asymptomatic participants without small-fiber degeneration and those with carpal tunnel syndrome with degeneration.

Findings in cold detection threshold from the quantitative sensory testing battery have previously been compared to intraepidermal nerve fiber density.[11] That study did not report findings on validity, but it was possible to calculate them. The cold detection threshold was found to have a sensitivity of 0.70 (95% CI: 0.46, 0.88), specificity of 0.71 (95% CI: 0.29, 0.96), positive likelihood ratio of 2.45 (95% CI: 0.73, 8.18), and negative likelihood ratio of 0.42 (95% CI: 0.19, 0.95) in individuals with idiopathic small-fiber neuropathies. These diagnostic values for cold detection threshold are similar to those found in the current study. The current study is the first to report the findings of warm detection threshold and a combination of warm and cold detection thresholds in comparison to intraepidermal nerve fiber density in patients with peripheral neuropathy, and suggests that both warm and cold detection thresholds may be required to enhance the validity to determine small-fiber degeneration.

There are some limitations of the present study. The researchers included patients with carpal tunnel syndrome as a model neuropathy, as it is well established that a subset of these patients have small-fiber degeneration.[21] Further studies are required to establish whether our results are generalizable to other peripheral neuropathies where small-fiber degeneration has been found. Most importantly, it remains to be determined whether the suggested clinical bedside tests have comparable clinimetric properties for other commonly affected body regions, such as the feet, where innervation density is much lower compared to the hand.

The current study examined the validity of these tests; however, reliability is another essential clinimetric property to ensure that a test is appropriate for use in clinical practice. While both intertester and intratester reliability are well established for the validated quantitative sensory testing protocol,[5,10] future studies are needed to determine the reliability of the bedside clinical small-fiber tests.

CONCLUSION

Our data suggest that simple bedside tests are valid in the identification of small-fiber degeneration in individuals with carpal tunnel syndrome. A finding of reduced pinprick sensation tested with a Neurotip may indicate the presence of small-fiber degeneration. A normal pinprick sensation, however, warrants further testing with a cold and a warm coin. A similar sensation to both temperatures compared to a reference area indicates that there may be no small-fiber degeneration.

In addition, our study found that quantitative sensory testing thresholds have some validity in identifying small-fiber degeneration. If warm detection threshold is elevated, then it is highly likely that small-fiber degeneration is present. If warm detection threshold is normal, followed by a normal cold detection threshold, then it is highly unlikely that small-fiber degeneration is present. ◉

KEY POINTS

FINDINGS: The use of pinprick followed by thermal testing, or warm detection followed by cold detection thresholds, may be valid methods of assessing small-fiber degeneration.

IMPLICATIONS: As small fibers constitute a substantial proportion of peripheral nerves and have been shown to deteriorate in common peripheral neuropathies, the ability to test these fibers accurately in a clinical setting is of great importance. This study suggests that simple and cost-effective tools may be valid to detect small-fiber degeneration in a clinical setting.

CAUTION: Only participants with a diagnosis of carpal tunnel syndrome and asymptomatic participants were included in this study. Further research is needed in patients with mixed presentations or diagnoses.

ACKNOWLEDGMENTS: *The input of Professor David Bennett in the design of the Oxford carpal tunnel syndrome study and the assistance of Shemane Murtagh with the logistic regression analysis are gratefully acknowledged.*

REFERENCES

1. Baselgia LT, Bennett DL, Silbiger RM, Schmid AB. Negative neurodynamic tests do not exclude neural dysfunction in patients with entrapment neuropathies. *Arch Phys Med Rehabil*. 2017;98:480-486. https://doi.org/10.1016/j.apmr.2016.06.019
2. Bland JD. A neurophysiological grading scale for carpal tunnel syndrome. *Muscle Nerve*. 2000;23:1280-1283. https://doi.org/10.1002/1097-4598(200008)23:8<1280::AID-MUS20>3.0.CO;2-Y
3. de la Llave-Rincón AI, Fernández-de-las-Peñas C, Laguarta-Val S, et al. Increased pain sensitivity is not associated with electrodiagnostic findings in women with carpal tunnel syndrome. *Clin J Pain*. 2011;27:747-754. https://doi.org/10.1097/AJP.0b013e31821c29d3
4. Devigili G, Tugnoli V, Penza P, et al. The diagnostic criteria for small fibre neuropathy: from symptoms to neuropathology. *Brain*. 2008;131:1912-1925. https://doi.org/10.1093/brain/awn093
5. Felix ER, Widerström-Noga EG. Reliability and validity of quantitative sensory testing in persons with spinal cord injury and neuropathic pain. *J Rehabil Res Dev*. 2009;46:69-83. https://doi.org/10.1682/JRRD.2008.04.0058
6. Field A. *Discovering Statistics Using IBM SPSS Statistics*. 4th ed. London, UK: SAGE; 2013.
7. Finnerup NB, Haroutounian S, Kamerman P, et al. Neuropathic pain: an updated grading system for research and clinical practice. *Pain*. 2016;157:1599-1606. https://doi.org/10.1097/j.

pain.0000000000000492
8. Flahault A, Cadilhac M, Thomas G. Sample size calculation should be performed for design accuracy in diagnostic test studies. *J Clin Epidemiol*. 2005;58:859-862. https://doi.org/10.1016/j.jclinepi.2004.12.009
9. Fritz JM, Wainner RS. Examining diagnostic tests: an evidence-based perspective. *Phys Ther*. 2001;81:1546-1564. https://doi.org/10.1093/ptj/81.9.1546
10. Geber C, Klein T, Azad S, et al. Test-retest and interobserver reliability of quantitative sensory testing according to the protocol of the German Research Network on Neuropathic Pain (DFNS): a multi-centre study. *Pain*. 2011;152:548-556. https://doi.org/10.1016/j.pain.2010.11.013
11. Holland NR, Crawford TO, Hauer P, Cornblath DR, Griffin JW, McArthur JC. Small-fiber sensory neuropathies: clinical course and neuropathology of idiopathic cases. *Ann Neurol*. 1998;44:47-59. https://doi.org/10.1002/ana.410440111
12. Hosmer D, Lemeshow S. *Applied Logistic Regression*. 2nd ed. New York, NY: Wiley-Interscience; 2000.
13. Lauria G, Bakkers M, Schmitz C, et al. Intraepidermal nerve fiber density at the distal leg: a worldwide normative reference study. *J Peripher Nerv Syst*. 2010;15:202-207. https://doi.org/10.1111/j.1529-8027.2010.00271.x
14. Leffler AS, Hansson P. Painful traumatic peripheral partial nerve injury-sensory dysfunction profiles comparing outcomes of bedside examination and quantitative sensory testing. *Eur J Pain*. 2008;12:397-402. https://doi.org/10.1016/j.ejpain.2007.08.009
15. Ling L, Xue J, Liu Y, et al. Quantitative and morphological study of intraepidermal nerve fibre in healthy individuals. *Neurol Res*. 2015;37:974-978. https://doi.org/10.1179/1743132815Y.0000000082
16. Liu Y, Fan X, Wei Y, Piao Z, Jiang X. Intraepidermal nerve fiber density of healthy human. *Neurol Res*. 2014;36:911-914. https://doi.org/10.1179/1743132814Y.0000000377
17. Marchettini P, Marangoni C, Lacerenza M, Formaglio F. The Lindblom roller. *Eur J Pain*. 2003;7:359-364. https://doi.org/10.1016/S1090-3801(03)00052-1
18. Petersen KL, Rice FL, Farhadi M, Reda H, Rowbotham MC. Natural history of cutaneous innervation following herpes zoster. *Pain*. 2010;150:75-82. https://doi.org/10.1016/j.pain.2010.04.002
19. Petty NJ. *Neuromusculoskeletal Examination and Assessment. A Handbook for Therapists*. 4th ed. Edinburgh, UK: Elsevier/Churchill Livingstone; 2011.
20. Rolke R, Magerl W, Campbell KA, et al. Quantitative sensory testing: a comprehensive protocol for clinical trials. *Eur J Pain*. 2006;10:77-88. https://doi.org/10.1016/j.ejpain.2005.02.003
21. Schmid AB, Bland JD, Bhat MA, Bennett DL. The relationship of nerve fibre pathology to sensory function in entrapment neuropathy. *Brain*. 2014;137:3186-3199. https://doi.org/10.1093/brain/awu288
22. Schüning J, Scherens A, Haussleiter IS, et al. Sensory changes and loss of intraepidermal nerve fibers in painful unilateral nerve injury. *Clin J Pain*. 2009;25:683-690. https://doi.org/10.1097/AJP.0b013e3181a1260e
23. Sveen KA, Karimé B, Jørum E, et al. Small- and large-fiber neuropathy after 40 years of type 1 diabetes: associations with glycemic control and advanced protein glycation: the Oslo Study. *Diabetes Care*. 2013;36:3712-3717. https://doi.org/10.2337/dc13-0788
24. Tamburin S, Cacciatori C, Praitano ML, et al. Median nerve small- and large-fiber damage in carpal tunnel syndrome: a quantitative sensory testing study. *J Pain*. 2011;12:205-212. https://doi.org/10.1016/j.jpain.2010.06.010
25. Terkelsen AJ, Karlsson P, Lauria G, Freeman R, Finnerup NB, Jensen TS. The diagnostic challenge of small fibre neuropathy: clinical presentations, evaluations, and causes. *Lancet Neurol*. 2017;16:934-944. https://doi.org/10.1016/S1474-4422(17)30329-0
26. Themistocleous AC, Ramirez JD, Shillo PR, et al. The Pain in Neuropathy Study (PiNS): a cross-sectional observational study determining the somatosensory phenotype of painful and painless diabetic neuropathy. *Pain*. 2016;157:1132-1145. https://doi.org/10.1097/j.pain.0000000000000491

APPENDIX A

NEUROPHYSIOLOGY DATA OF THE MEDIAN NERVE IN PATIENTS WITH CARPAL TUNNEL SYNDROME AND HEALTHY VOLUNTEERS*

	Mild	Moderate	Severe	Controls	*P* Value
SNAP, μV	7.02 (4.57)	6.59 (6.23)	Absent	14.47 (14.96)	<.001†
Sensory NCV, m/s	36.4 (5.36)	31.83 (6.54)	NA	51.30 (8.81)	<.001‡
DML, ms	4.2 (0.49)	5.08 (0.98)	7.02 (3.02)	3.13 (0.61)	<.001§
CMAP, mV	7.53 (4.36)	5.28 (4.28)	3.54 (2.9)	7.25 (3.23)	<.001‖

Abbreviations: CMAP, compound motor action potential; DML, distal motor latency; NA, not applicable; NCV, nerve conduction velocity; SNAP, sensory nerve action potential.

**Values are median (interquartile range).*

†*Post hoc testing: asymptomatic versus mild, P = .08; asymptomatic versus moderate, P = .005.*

‡*Post hoc testing: all pairwise comparisons, P<.001.*

§*Post hoc testing: asymptomatic versus mild, P = .047; asymptomatic versus moderate, P<.001; asymptomatic versus severe, P<.001.*

‖*Post hoc testing: asymptomatic versus mild, P>.05; asymptomatic versus moderate, P>.05; asymptomatic versus severe, P<.001.*

APPENDIX B

VALIDITY RESULTS FOR PINPRICK (NEUROTIP) DIVIDED BY SEVERITY BASED ON ELECTRODIAGNOSTIC TEST GRADES*

Grade	n	Sensitivity	Specificity	PPV	NPV	+LR	–LR
Mild (grades 1-2)	27	0.60 (0.36, 0.81)	0.57 (0.18, 0.90)	0.80 (0.61, 0.91)	0.33 (0.18, 0.54)	1.40 (0.55, 3.54)	0.70 (0.30, 1.62)
Moderate (grade 3)	26	0.53 (0.29, 0.76)	0.86 (0.42, 1.00)	0.91 (0.61, 0.98)	0.40 (0.28, 0.54)	3.68 (0.57, 23.76)	0.55 (0.31, 0.97)
Severe (grades 4-6)	32	0.43 (0.25, 0.63)	1.00 (0.16, 1.00)	1.00 (1.00, 1.00)	0.11 (0.08, 0.14)	Not defined†	0.57 (0.41, 0.77)

Abbreviations: –LR, negative likelihood ratio; +LR, positive likelihood ratio; NPV, negative predictive value; PPV, positive predictive value.
**Values in parentheses are 95% confidence interval.*
†*The calculation of +LR is sensitivity/(1 – specificity), and specificity is 1; therefore, sensitivity is divided by zero and hence +LR is undefined.*

DEYDRE S. TEYHEN, DPT, PhD[1] • **STEPHEN L. GOFFAR,** DPT, PhD[2] • **SCOTT W. SHAFFER,** MPT, PhD[3,4]
KYLE KIESEL, MPT, PhD[5] • **ROBERT J. BUTLER,** DPT, PhD[6] • **ANN-MAREE TEDALDI,** DPT[7]
JENNIFER C. PRYE, MEd, ATC[8] • **DANIEL I. RHON,** DPT, DSc[4,9] • **PHILLIP J. PLISKY,** MPT, DSc[5]

Incidence of Musculoskeletal Injury in US Army Unit Types: A Prospective Cohort Study

BACKGROUND: Musculoskeletal injuries during military service are a primary source of disability, resulting in 2.4 million annual health care visits and 25 million limited-duty days. While the injury incidence during basic training is well documented, there is little understanding of injury distribution by organization type in the US Army following initial training.

OBJECTIVE: To compare injury incidence, distribution, and impact across various military units.

METHODS: In this prospective observational cohort study, comprehensive injury data from subject questionnaires and medical chart reviews were collected over 12 months for 1430 initially healthy Army personnel, representing combat, combat support, combat service support, and ranger units. Health care utilization and time loss due to injury were also collected.

RESULTS: Of 1430 soldiers, 481 (33.6%) had time-loss injury, 222 (15.5%) were injured without limited work, 60 (4.2%) reported an injury but did not seek medical care, and 667 (46.6%) were uninjured. Across the whole sample, injuries were responsible for 5.9 ± 14.4 medical visits per soldier, 21 902 days of limited work, and $1 337 000 ($1901 ± $6535 per soldier) in medical costs. Considering only those reporting injury, each person averaged 36.3 ± 59.7 limited-work days. The injury incidence was highest in combat service support units (65.6%), with a risk ratio 1.60 times that of the reference group (combat, 41.1%).

CONCLUSION: Combat support and combat service support personnel were more likely to have 1 or more injuries compared to rangers and combat personnel. The higher relative risk of injury in support units should be explored further. *J Orthop Sports Phys Ther 2018;48(10):749-757. Epub 22 May 2018. doi:10.2519/jospt.2018.7979*

KEY WORDS: *injury incidence, medical costs, musculoskeletal, overuse injury, pain*

Within the US Department of Defense, approximately 1.6 million musculoskeletal injuries occur annually.[7] Musculoskeletal injuries are the leading contributor to disability for the US military, accounting for approximately 2.4 million medical visits and $548 million in direct patient care costs (personal communication with Scott Gregg, who collected and analyzed FY2007-2008 M2 data from the Military Health System Data Repository). This translates into in excess of 25 million limited-duty days and more than 900 000 service members affected each year.[2,4] To optimize performance and effectiveness of the US military, efforts must be made to decrease the impact of musculoskeletal injuries and maximize the health of the all-volunteer force.

Military duties are intrinsically hazardous; however, surprisingly, the leading cause of musculoskeletal injuries is not due to combat-related injuries.[4,7,9] Between 2001 and 2013, 31% to 34% of medical evacuations from Iraq and Afghanistan were for nonbattle musculoskeletal injuries.[8] Regardless of setting, the vast majority of injuries (82%) were classified as overuse (unpublished report, Army Medical Surveillance Activity, US Army Center for Health Promotion and Preventive Medicine, 2005),[1] and injuries were often related to participation in recreational sports and physical training.[9,21,24] Only a small number of incidences require detailed reporting of a specific cause and setting of injury (eg, parachute injuries). While these facts highlight the importance of this problem, they omit crucial details, such as severity of injury in terms of cost and limited-duty days and the prevalences of musculoskeletal injuries between different types of military units.

[1]Walter Reed Army Institute of Research (WRAIR), Silver Spring, MD. [2]School of Physical Therapy, University of the Incarnate Word, San Antonio, TX. [3]Graduate School, US Army Medical Department Center and School, Ft Sam Houston, TX. [4]Graduate Program in Physical Therapy, US Army-Baylor University, Waco, TX. [5]School of Physical Therapy, University of Evansville, Evansville, IN. [6]St Louis Cardinals, St Louis, MO. [7]Rocky Mountain University of Health Professions, Provo, UT. [8]Health Leadership High School, Albuquerque, NM. [9]Center for the Intrepid, Brooke Army Medical Center, San Antonio, TX. The Institutional Review Board at Madigan Army Medical Center (Joint Base Lewis-McChord, WA) approved this study. The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or reflecting the views of the US Army Medical Department, the US Army Office of the Surgeon General, the US Army, the Department of Defense, or the Federal Government. This project was supported financially by the Army Medical Department Advanced Medical Technology Initiative, the Telemedicine and Advanced Technology Research Center, and the Defense Medical Research and Development Program, Military Operational Medicine Research Program from the US Army Medical Research and Materiel Command, Ft Detrick, MD. The authors certify that they have no affiliations with or financial involvement in any organization or entity with a direct financial interest in the subject matter or materials discussed in the article. Address correspondence to Dr Stephen L. Goffar, 4301 Broadway, CPO 412, San Antonio, TX 78250. E-mail: Goffar@uiwtx.edu

Although the occurrence of musculoskeletal injuries in military recruits in basic or initial entry training (IET) has been well documented,[14,15] limited data exist regarding the incidence and type of musculoskeletal injuries across active-duty soldiers in military units following IET. The limited availability of data on soldiers after IET is problematic, because there is scant evidence to direct prevention efforts or distribute medical resources. Additionally, research on the diversity of occupational demands across military specialties is inadequate, thus hampering individual and occupational injury prevention and management approaches. The purpose of this study was to compare injury incidence rates, body regions, health care utilization costs, and limited-work days among soldiers in 4 different primary military unit types: combat soldiers from an infantry unit, combat support soldiers who provide direct support to the infantry, combat service support soldiers who provide logistical support (supply, transportation, health services, etc), and elite soldiers from a ranger unit. A better understanding of these variables can help improve our ability to prioritize the injury types, body regions, and military units where injury prevention and medical management should be focused.

METHODS

Subjects

Participants were active-duty US soldiers, aged 18 to 45 years, who could speak and read English and had no current physical limitations (able to participate fully in military training and sport) (**FIGURE**). All soldiers were enrolled in the Military Health System and were excluded if they had less than 9 months of continuous eligibility in the benefits system from the date of enrollment. Participants were recruited from across the spectrum of military units (rangers, combat, combat support, and combat service support). Participants with a current or recent injury were excluded. A current or recent injury was defined as an injury in the last 30 days in which the individual sought medical care. Ethics approval was granted by the Institutional Review Board at Madigan Army Medical Center (Joint Base Lewis-McChord, WA). All soldiers provided written informed consent prior to their participation.

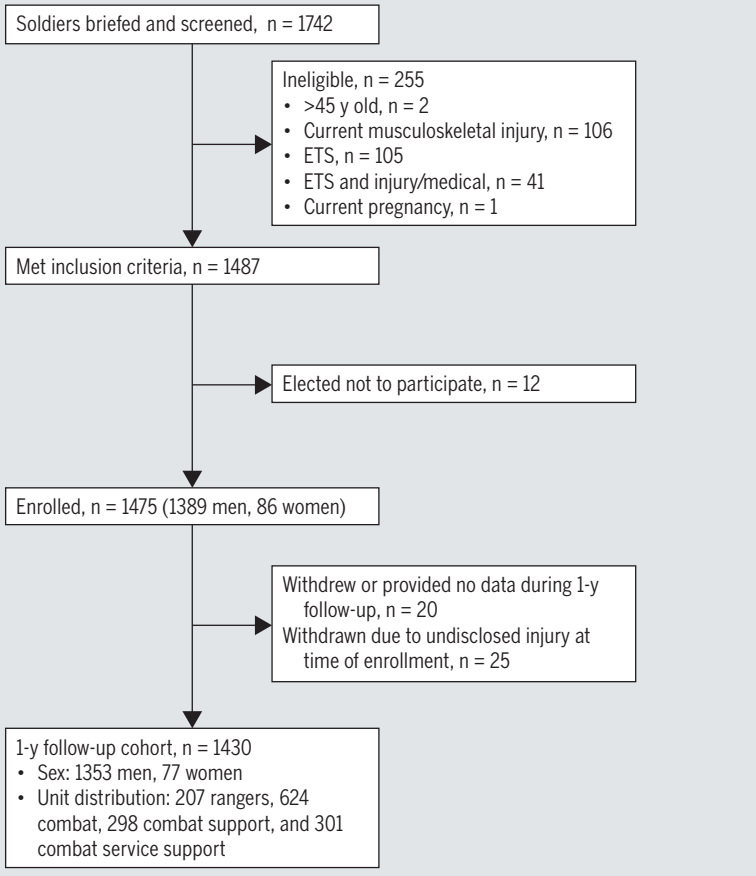


FIGURE. Flow of participants. Abbreviation: ETS, expiration of term of service.

Procedures

Healthy soldiers were recruited from large group briefings, and then volunteers were consented and screened for inclusion in the study. Enrollment of subjects occurred between May and August 2011. Subjects agreed to prospective injury surveillance, consisting of 3 different approaches: (1) a monthly survey e-mailed to each subject, (2) a comprehensive review of medical records for occurrence of an injury, and (3) mining of injury diagnosis billing codes, based on health care utilization data extracted from the Military Health System Management Analysis and Reporting Tool (M2). The M2 data dictionary is publicly available at www.health.mil.

Variables of Interest

The primary variables of interest included injury incidence, body region affected, injury-related health care costs, and time loss due to injury. Beyond simple incidence of injuries, a series of 4 indices were calculated to allow a more specific representation of musculoskeletal injuries alone (comprehensive), musculoskeletal injuries based on cause (acute or overuse), and musculoskeletal injuries that resulted in time loss. These indices also allowed for comparison across organizations of different sizes and previous

reports. To ensure the greatest possibility of capturing all data of interest, a variety of sources were explored.

Data Sources

Monthly Injury Surveillance Survey Using a web-based automated system, a monthly injury surveillance survey was used to collect self-reported injury data (see the **APPENDIX** for details, available at www.jospt.org). Each survey consisted of 3 general questions. When participants indicated that they had a new injury, they answered 8 injury-specific questions. Every 30 days over a 1-year period, starting from the date of enrollment, participants were sent an e-mail containing the link to the survey. Participants who had not completed the survey within 7 days were sent a second reminder e-mail and placed on a list for telephone reminders.

Medical Record Reviews Reviews of medical records were conducted to identify injuries that resulted in care seeking during the 12-month period that began from the date of enrollment. Two different reviewers used a structured format previously described to identify injury occurrence.[10,29] The goals of the medical record review were to triangulate injury occurrence, obtain detailed information related to health care utilization, and assist in the classification of injuries as acute or overuse. This classification facilitated inclusion in the injury indices.

Health Care Utilization Database The US Defense Health Agency uses the M2 database to track health care utilization. It delivers summary and detailed clinical, population, and financial data. It includes all person-level encounters within the Military Health System and TRICARE civilian referral network. This includes health care provided on combat deployments to areas such as Afghanistan and Iraq. The M2 database allows extraction of health care utilization based on the International Classification of Diseases-Ninth Revision (ICD-9). The ICD-9 codes of interest were limited to those in the musculoskeletal injuries series (ICD-9 700-900) and categorized into an injury index as described below. Previous studies have leveraged the M2 database for similar analyses of health care utilization and costs.[3,5]

Health care utilization for each participant was extracted from the M2 database. All musculoskeletal injuries were identified and categorized first as upper extremity, lower extremity, or spine, and then into 8 more specific body-region categories (upper back, head, and neck; shoulder; elbow, wrist, and hand; low back; hip and thigh; knee; lower leg; and foot and ankle). For each injury identified, medical costs, number of health care visits, and time loss were identified for data analysis. Injuries were further divided into $500 increments, based on the costs associated with the medical management of the injury.

Time-Loss Accounting System

Information concerning the number of days of limited duty or work time loss due to injury was extracted from a central digital repository using e-Profile software (Military Operational Data System). A secondary system, generally used for shorter-duration time loss, is maintained in hard copy at the unit level. To ensure the greatest possible accuracy, the researchers queried the digital system and performed a physical count of time-loss documentation kept by each military unit. Duration of time loss was calculated for each injury.

Data Validation

Collecting injury data from 3 sources provided the best means of accurately capturing all instances of musculoskeletal injuries. If a participant omitted an injury from the monthly e-mail or phone survey for which he or she sought care in the Military Health System, his or her data would be captured through the data pulled from M2. Injuries that occurred but for which no care was sought (n = 60) were captured only through the self-report e-mail or phone surveys. All reported injuries, regardless of source, were cross-matched across all other data sources to ensure there were no omissions or duplications of reported injuries. Data points reported through the e-mail or phone survey but not present in the health care utilization database had no health care utilization or medical cost associated with them.

Injury Indices

To classify injuries identified during data collection and enable comparisons to other population studies, 4 injury indices were employed. The comprehensive injury index (CII), overuse injury index (OII), and acute injury index (AII) were calculated from the incidence of specific ICD-9 codes found in the medical record, as described by Knapik et al[13] in 2004. The OII includes musculoskeletal injuries associated with cumulative microtrauma, while the AII includes higher-energy trauma or acute injuries. The CII includes both the OII and AII, plus several musculoskeletal conditions not typically categorized as acute or overuse. An additional index, the time-loss injury index (TLII), was developed to account for injuries that resulted in lost work days per 1000 person-days. If an injury resulted in any lost time from work, it was included as a TLII injury.

Data Analysis

Incidence (percent) of soldiers incurring an injury attributable to each index is reported in total as well as for Army unit type. This value does not consider how many injuries per soldier occurred. The relative risk, or risk ratio (RR), of an individual incurring a musculoskeletal injury was calculated for each unit and index using the lowest-incidence military unit as the referent. Within each injury index, injury incidence rates (IIRs) were calculated by dividing the net occurrence of injury per 1000 person-days in total and in each unit. The IIR includes all injuries incurred and accounts for the occurrence of multiple injuries per individual. Rate ratio was calculated by dividing the IIR of the military unit of interest by the IIR of the military unit with the lowest IIR.

The resultant value allows a comparison of total injury occurrence by injury type across varying sizes of military organization units within a common period of 1000 person-days.

RESULTS

Demographics

Of 2315 soldiers briefed about this study, 1475 active-duty soldiers were eligible and enrolled. Throughout the 1-year follow-up period, 20 (1.35%) participants requested to be withdrawn from the study or provided no data. The medical chart review revealed that 25 (1.7%) participants were injured and sought medical care during enrollment and were withdrawn from the study, resulting in a final sample of 1430 (mean ± SD age, 24.6 ± 5.0 years; body mass index, 26.7 ± 3.4 kg/m^2; 1353 male, 77 female) (**FIGURE**). Participants included soldiers in combat (n = 624, 43.6%), combat support (n = 298, 20.8%), combat service support (n = 301, 21.0%), and ranger (n = 207, 14.5%) military units. Time in the US Army included less than 1 year (n = 273, 19.1%), at least 1 but less than 3 years (n = 418, 29.2%), at least 3 years but less than 5 years (n = 362, 25.3%), at least 5 but less than 10 years (n = 261, 18.3%), and 10 years or greater (n = 116, 8.1%). Over half of the sample (n = 866, 60.6%) reported serving on a deployment within the 5 years prior to study enrollment.

Overall, 667 (46.6%) participants remained uninjured, and 763 (53.4%) sustained a musculoskeletal injury. The total number of injuries sustained was 1181 (0.8 ± 1.0 injuries per participant). Of those who sustained a musculoskeletal injury, 63.2% had a single injury, while 36.8% sustained multiple injuries (range, 2-6 injuries) (**TABLE 1**). Of the total 1181 injuries incurred, 648 were classified as comprehensive. Of those 648 injuries, 81 were further subclassified as acute and 449 as overuse. The remaining 533 musculoskeletal injuries reported were not classified under one of the indices. Diagnoses excluded from consideration were musculoskeletal conditions unrelated to injury, such as ICD-9 735.00 Hallux valgus.

Injury Incidence

Across all unit types, 50.6% of soldiers incurred a musculoskeletal injury captured in the CII, including 40.2% in the OII and 10.8% in the AII (**TABLE 2**). The rate was highest in the combat service support unit and lowest in the combat unit, with CII rates at 65.6% versus 41.1%, OII rates at 57.0% versus 30.2%, and AII rates at 16.1% versus 8.4%, respectively. With the combat unit as the referent, risk of injury was greater in all other units: rangers (RR = 1.1; 95% confidence interval [CI]: 1.04, 1.16), combat support (RR = 1.43; 95% CI: 1.35, 1.51), and combat service support (RR = 1.60; 95% CI: 1.52, 1.68).

Across all military units, the IIR for all injuries (CII) was 2.02, with 1.49 for overuse injuries (OII) and 0.32 for acute injuries (AII). The IIR for the CII was the lowest in rangers (0.33), who were used as the reference group. Compared to the referent, the IIRs for all 4 injury indices were greatest in the combat unit and lowest in the combat support unit (**TABLE 2**).

Body Region

The distribution of injuries within each unit did not vary substantially from the population as a whole, except for the incidence of foot and ankle injuries in the combat unit (20.7%). Generally, foot and ankle (20.5%), lower back (17.8%), and knee (15.9%) conditions were the most prevalent, but the entire lower extremity accounted for just under 50% of musculoskeletal injuries (**TABLE 3**).

Time Loss

Of those who sought medical care for their injury, 481 (68.4%) also had at least 1 time-loss injury (**TABLE 4**). The incidence of time-loss injuries ranged from 18.9% in the rangers to 54.0% in combat service support (**TABLE 2**). Overall, the TLII IIR was 0.98 injuries per 1000 person-days, with the combat unit reporting a TLII IIR essentially double those of other unit types (**TABLE 2**). The total sum of limited-work days for the 1430 participating soldiers was 21902 (36.3 ± 59.7 days per injured participant with documented days lost). The median duration of work limitation in this population was 14 days.

Cost

Of the 763 (53.4%) soldiers who sustained a musculoskeletal injury, 60 (4.2% of total

TABLE 1 Injuries Per Participant*

				Number of Injuries Sustained by Participant					
	Noninjured	Injured	Total Injuries	1 Injury	2 Injuries	3 Injuries	4 Injuries	5 Injuries	6 Injuries
Rangers (n = 207)	103	104 (50.2)	141	77 (74.0)	18 (17.3)	8 (7.7)	1 (1.0)	0 (0.0)	0 0.0)
Combat (n = 624)	347	277 (44.4)	381	198 (71.5)	58 (20.9)	18 (6.5)	2 (0.7)	1 (0.4)	0 (0.0)
Combat support (n = 298)	118	180 (60.4)	293	104 (57.8)	49 (27.2)	20 (11.1)	5 (2.8)	1 (0.6)	1 (0.6)
Combat service support (n = 301)	99	202 (67.1)	366	103 (51.0)	53 (26.2)	33 (16.3)	9 (4.5)	2 (0.1)	2 (0.1)
Total (n = 1430)	667	763 (53.4)	1181	482	178	79	17	4	3

Values are n or n (percent) of unit type total.

sample) reported a physical complaint that did not require medical care. The remaining 703 (49.2% of total sample) sought medical care for their musculoskeletal injuries, which was captured by the electronic medical record. Injured participants attended a mean ± SD of 5.9 ± 14.4 medical encounters at an overall cost of $1 337 000 ($1901 ± $6535 per injured participant). Across all military units, only 15.5% of injuries resulted in injury-related medical costs that exceeded $2500 (**TABLE 4**). The percent of ranger (25.6%) and combat (41.1%) unit members incurring medical costs greater than $500 was less than that incurred by combat support (60.6%) and combat service support (59.4%) members.

DISCUSSION

Several studies have reported injury rates for military service personnel.[6,14,16,18] However, this study is unique in that it included robust injury surveillance methodology, categorized different types of injuries (overuse, acute, and time loss), and analyzed the impact of time-loss injuries and costs of these injuries based on the type of military unit. Over half (53.4%) of the soldiers in this prospective study sustained a musculoskeletal injury. The incidence of injuries and subsequent time loss were greatest in combat support and combat service support units. Injuries to the lower extremity accounted for almost half of all injuries, while injuries to the spine and upper extremity represented 31% and 20%, respectively.

Injury Incidence

This research effort is one of a few that have captured injury incidence in multiple military specialty categories in a

TABLE 2 Injury Incidences and IIRs for Each Military Unit

	Total (n = 1430)*	Ranger (n = 207)†		Combat (n = 624)‖		Combat Support (n = 298)¶		Combat Service Support (n = 301)#	
	Injury Incidence	Injury Incidence	Risk Ratio‡§	Injury Incidence	Risk Ratio‡§	Injury Incidence	Risk Ratio‡§	Injury Incidence	Risk Ratio‡§
CII	50.6%	45.2%	1.10 (1.04, 1.16)	41.1%	1.00	58.7%	1.43 (1.35, 1.51)	65.6%	1.60 (1.52, 1.68)
OII	40.2%	31.8%	1.05 (0.99, 1.11)	30.2%	1.00	50.4%	1.67 (1.57, 1.77)	57.0%	1.89 (1.78, 2.00)
AII	10.8%	13.4%	1.60 (1.49, 1.71)	8.4%	1.00	10.6%	1.26 (1.17, 1.35)	16.1%	1.92 (1.78, 2.06)
TLII	33.6%	18.9%	0.71 (0.67, 0.75)	26.5%	1.00	49.8%	1.88 (1.77, 1.99)	54.0%	2.04 (1.92, 2.16)
	IIR**	IIR**	Rate Ratio‡††	IIR**	Rate Ratio‡††	IIR**	Rate Ratio‡††	IIR**	Rate Ratio‡††
CII	2.02	0.33	1.00	0.87	2.64 (2.50, 2.78)	0.63	1.90 (1.80, 2.00)	0.72	2.18 (2.06, 2.30)
OII	1.49	0.19	1.00	0.57	3.00 (2.82, 3.18)	0.47	2.47 (2.33, 2.61)	0.51	2.68 (2.52, 2.84)
AII	0.32	0.07	1.00	0.13	1.86 (1.73, 1.99)	0.06	0.86 (0.80, 0.92)	0.08	1.14 (1.06, 1.22)
TLII	0.98	0.27	1.00	0.45	1.97 (1.84, 2.10)	0.18	0.67 (0.63, 0.71)	0.21	0.79 (0.74, 0.84)

Abbreviations: AII, acute injury index; CII, comprehensive injury index; IIR, injury incidence rate; OII, overuse injury index; TLII, time-loss injury index.
**n = 763 (53.4% injured).*
†n = 104 (50.2% injured).
‡Values in parentheses are 95% confidence interval.
§Risk ratio was defined by using the unit with the lowest injury incidence (combat) as the referent.
‖n = 277 (44.4% injured).
¶n = 180 (60.4% injured).
#n = 202 (67.1% injured).
***Injuries per 1000 person-days.*
††Rate ratio was defined by using the unit with the lowest IIR (rangers) as the referent.

TABLE 3 Injuries by Body Region and Military Unit*

	UBHN	Shoulder	EWH	Low Back	HT	Knee	LL	FA	Unknown
Ranger (n = 207; 141 injuries)	24 (17.0)	12 (8.5)	17 (12.1)	15 (10.6)	7 (5.0)	27 (19.1)	5 (3.5)	34 (24.1)	0 (0.0)
Combat (n = 624; 381 injuries)	50 (13.1)	32 (8.4)	42 (11.0)	73 (19.2)	16 (4.2)	67 (17.6)	22 (5.8)	79 (20.7)	0 (0.0)
Combat support (n = 298; 293 injuries)	28 (9.6)	36 (12.3)	30 (10.2)	53 (18.1)	18 (6.1)	46 (15.7)	17 (5.8)	65 (22.2)	0 (0.0)
Combat service support (n = 301; 366 injuries)	53 (14.5)	31 (8.5)	36 (9.8)	69 (18.9)	34 (9.3)	48 (13.1)	30 (8.2)	64 (17.5)	1 (0.3)
Total (n = 1430; 1181 injuries)	155 (13.1)	111 (9.4)	125 (10.6)	210 (17.8)	75 (6.4)	188 (15.9)	74 (6.3)	242 (20.5)	1 (0.01)

Abbreviations: EWH, elbow, wrist, and hand; FA, foot and ankle; HT, hip and thigh; LL, lower leg; UBHN, upper back, head, and neck.
**Values are n (percent) of injuries per military unit.*

given time frame (2011-2012). While several researchers have looked at single military occupational specialties,[14,16,20,29] this project took a broader perspective, examining larger clusters of military specialties grouped by unit (combat, combat support, combat service support, and rangers). In general, the findings for each unit were similar to those reported previously for specific occupational specialties within each military unit,[14,16,17,29] but the identified differences between major units indicate that there is variation in risk that must be better understood if prevention efforts are to be effective (**TABLE 3**).

In 2013, Knapik et al[14] examined the injury incidence of military police soldiers during IET (basic and advanced individual training) and found that 34% of men and 66.7% of women incurred some type of musculoskeletal injury (CII), with overuse injury (OII) occurrences of 23.6% (IIR = 1.79) and 58.4% (IIR = 4.51), respectively. These injury incidences are lower for men when compared to the injury incidence the present study found in combat support soldiers (58.7% for overall injuries and 50.4% for overuse injuries). Combat service support units reported injuries in this study at 65.6%, slightly lower than the rate reported for mechanics working in combat service support units.[16] While overuse injury and traumatic injury incidence have been reported to be as high as 30%,[16] the authors noted a much higher overuse injury incidence of 57.0% and a lower traumatic (acute) injury incidence of 16.1% in combat service support soldiers. These differences may relate to the variance in demands and location of unit type, as previous work was conducted at an airborne installation with greater risk of parachute-related injuries (9% of overall injuries).[16]

Nondeployment injury incidence over a 12-month period has been reported in 3 types of combat units (infantry, 48.6%; artillery, 66%; special forces, 53%) and 1 type of combat support unit (combat engineers, 86%).[20] These reported injury incidences were slightly higher than the combat (41.1%), combat support (58.7%), and ranger (45.2%) injury incidences seen in the present study. Of note, consistent between studies is the variation in injury incidence observed across units; that is, combat units reported lower incidence than did combat service units. The higher incidences observed in this study may be associated with increased demands associated with multiple deployments.

The findings reported here and previously by other researchers further delineate the occupational risk differences between military organization subgroups. Contrary to common speculation by soldiers and military leaders, the risk of overall injury and overuse injury is greater in combat service support and combat support units, not among those in combat arms.

Body Regions

Like previous reports, the current study identified that most injuries involved the lower extremities and lower back.[7,9,23,24] Hauret et al[7] reported that a total of 743 547 musculoskeletal injuries occurred in service members in 2006, with 22% in the knee and lower leg, 20% in the lumbar spine, and 13% in the foot and ankle. In comparison, this research found an almost identical injury occurrence in the knee and lower leg (22.2%) and a similar occurrence in the low back (17.8%). Interestingly, the results of this study showed a slightly higher foot and ankle injury occurrence (20.5%) compared to that reported by Hauret et al[7] (only 13%), but they closely followed foot and ankle injuries reported in US Army Brigade Combat Team soldiers serving in Afghanistan (19%).[23] Collectively, these results reinforce that the majority of musculoskeletal injuries in the military occur in the lower extremity and low back (66.9% in this cohort); as a result, further resources and critical analysis to mitigate and enhance recovery from these specific injuries appear warranted.

One of the novel aspects of the present study was the specific inclusion and analysis of military specialties by unit type. These results demonstrated limited variability in the distribution of injury by body area between the military units, except for low back injuries for rangers (10.6%). These results are surprising, as

TABLE 4 Medical Costs and Time Loss for Soldiers Who Sustained a Musculoskeletal Injury*

		Time Loss for Those Who Sought Medical Care					Medical Costs		
	Injured, n†	None	1-7 d	8-30 d	>30 d	ND	≤$500	>$500 and ≤$2500	>$2500
Total (n = 1430)	60	222	49	162	171	99	359	235	109
Rangers (n = 207)	14	52 (57.8)	3 (3.3)	10 (11.1)	11 (12.2)	14 (15.6)	67 (74.4)	15 (16.7)	8 (8.9)
Combat (n = 624)	29	87 (35.1)	20 (8.1)	53 (21.4)	52 (21.0)	36 (14.5)	146 (58.9)	76 (30.6)	26 (10.5)
Combat support (n = 298)	7	39 (22.5)	16 (9.2)	47 (27.2)	54 (31.2)	17 (9.8)	68 (39.3)	62 (35.8)	43 (24.9)
Combat service support (n = 301)	10	44 (22.9)	10 (5.2)	52 (27.1)	54 (28.1)	32 (16.7)	78 (40.6)	82 (42.7)	32 (16.7)

Abbreviation: ND, not documented in medical records or in e-Profile.
**Values are n (percent) of unit total seeking care, unless otherwise indicated.*
†Self-report only (not seeking care).

previous research in athletic and military populations has exhibited variability in the incidence of injured body regions among various sports and military occupations.[11,22] The authors expected to find a distribution of injuries reflective of the diversity of occupational specialty demands and the training associated with preparation for those demands. Additional investigation into sports- and military-specific occupational demands and specific screening is ongoing and required to mitigate future injuries and enhance performance.[29]

Time Loss

Although various studies have examined injury-related work loss in relationship to injury type or mechanism of injury,[22,24] this research examined limited-duty days and health care costs among groups of military specialties. The results of this study demonstrated limited time-loss utilization, lost-duty days, and health care costs for US Army rangers and combat soldiers compared to combat support and combat service support soldiers (**TABLE 4**). There are several potential variables (eg, environment, training demands, and access to health care) for increased time-loss injury and health care utilization among combat support and combat service support units that require additional examination.

Several authors have advocated for increased personnel and a sports medicine approach that includes injury prevention, enhanced access to musculoskeletal expertise, and human performance optimization.[2,19,22,28] Currently, the sports medicine team for US Army rangers includes physical therapists and strength and conditioning coaches, with a ratio of approximately 1 physical therapist per 600 to 1000 soldiers.[28] In contrast, and despite increased incidence of time-loss injuries and health care costs, combat units have a ratio of 1 physical therapist per 3500 soldiers and do not have any assigned strength and conditioning coaches. Soldiers in combat support and combat service support units do not have sports medicine providers integrated into their units. Additional analysis into the optimal staffing models required to mitigate injuries, enhance return to duty, and optimize performance is needed.

Costs

Musculoskeletal injuries are the most common reason service members seek health care, regardless of the setting (home base or deployed).[7,12] Beyond the reduction in military unit effectiveness resulting from time loss due to injury, musculoskeletal injuries result in a significant financial burden on the military health care system. Costs associated with the long-term medical care of an injury may also be used as a measure of severity,[25,26] and health care utilization can help assess the impact of the injury on public health.[27] Some injuries do not result in any utilization of health care in the medical system, as evidenced by this cohort (**TABLE 4**). This indicates that there is a subset of patients who do not need to seek any care beyond an initial evaluation and treatment (based on the median cost of injuries being less than $500). These subgroups of patients are important to identify and distinguish from those that use large amounts of health care resources. Future research should focus on identifying predictors of high health care utilizers and the potential mediators of health care costs after injury in each of these military units.

While important to the understanding of injury distribution, development of injury prevention strategies, and distribution of health care resources in the US Army, this study has limitations. First, the results may have a narrow generalizability to the other military services. The US Air Force, Navy, and Marine Corps each have job classifications that overlap with those explored in this research, but the training demands, physical standards, and settings vary significantly. Second, despite efforts made to recruit an appropriate representation of female soldiers, our sample does not specifically reflect the ratio of men to women in the US Army at the time of the study.

CONCLUSION

This study is unique in that it reports injury rates in soldiers from different US Army unit types based on type and location of injuries, along with associated health care costs and the amount of work restriction due to injury. Injury incidence was high overall, affecting as much as two thirds of a military unit's personnel. Combat support and combat service support personnel were more likely to be injured and more likely to have multiple injuries when compared to rangers and combat personnel. While the percent of soldiers injured was lowest in combat units, this same group had the highest incidence of injuries per 1000 person-days.

Injuries to the lower extremity accounted for almost half of all injuries, while injuries to the spine represented 31%. Future studies can use this information for focused research on injuries that are costlier, result in more days of lost work, and occur in more common body regions. Additionally, the influence of the sports medicine team on IIRs in each unit type should be explored further. ◉

KEY POINTS

FINDINGS: Over half (53.4%) of the military members in this prospective study sustained a musculoskeletal injury. The incidence of injuries was greatest in units supporting combat, compared to US Army ranger and combat units. Injuries to the lower extremity accounted for almost half of all injuries, while injuries to the spine and upper extremity represented 31% and 20%, respectively.

IMPLICATIONS: This study suggests a disparity in injury rates among military members of different organization types. This may indicate that a standard injury prevention approach may not optimize the health of the fighting force. It also provides potential areas of future investigation and supportive clinical care to address the diversity of injuries and associated health care costs among various military units.

CAUTION: While this study addresses a wide variety of US Army unit types and their distinct physical implications, generalization of these findings to the other military services (Navy, Marine Corps, Air Force) and first responders should be made with caution. Additional research examining injury rates among various military units and military occupations is needed to validate and expand on this study's conclusions.

ACKNOWLEDGMENTS: *This study was conducted in collaboration with physical therapy researchers and administrative assistants from US Army-Baylor University, US Army Medical Department Center and School, San Antonio, TX; the University of Evansville, Evansville, IN; Duke University, Durham, NC; and the University of Puget Sound, Tacoma, WA. Additional research support was provided by Joint Base Lewis-McChord's Physical Therapy Clinic, Madigan Army Medical Center, and the 2nd Ranger Battalion, 75th Ranger Regiment. The authors would like to acknowledge the following staff for their administrative and technical support: Josh Miller; Julie Koepp; Nadine Shewell; Norman Gardner; Liz Reilly-Hess; Celestina Wildman; Miyoko D. Sasakura, MSPH; Gail Baudendistel, PT, CPA, MS, MAFM; Tom Kurtz, PT; and Jeff Wilson. Technology support included Bob Humphrey, Dalene Becka, Jack Brower, and Jared Willard. Additional data logistical support was provided by Theresa Leatherwood, CPT; Carrie Hoppes, SFC; Sean Stephens, SFC; Ryan Murphy, SSG; Christopher Villarreal, CPO; Ferdinand Ajel; Paul Gorman, ATC, PTA, CSCS; Luck Chanthabandith; Phil Borja; Matthew Buchanan; and Valeria DeAngelis.*

REFERENCES

1. Almeida SA, Williams KM, Shaffer RA, Brodine SK. Epidemiological patterns of musculoskeletal injuries and physical training. *Med Sci Sports Exerc*. 1999;31:1176-1182.
2. Cameron KL, Owens BD, DeBerardino TM. Incidence of ankle sprains among active-duty members of the United States Armed Services from 1998 through 2006. *J Athl Train*. 2010;45:29-38. https://doi.org/10.4085/1062-6050-45.1.29
3. Childs JD, Fritz JM, Wu SS, et al. Implications of early and guideline adherent physical therapy for low back pain on utilization and costs. *BMC Health Serv Res*. 2015;15:150. https://doi.org/10.1186/s12913-015-0830-3
4. Cohen SP, Griffith S, Larkin TM, Villena F, Larkin R. Presentation, diagnoses, mechanisms of injury, and treatment of soldiers injured in Operation Iraqi Freedom: an epidemiological study conducted at two military pain management centers. *Anesth Analg*. 2005;101:1098-1103. https://doi.org/10.1213/01.ane.0000169332.45209.cf
5. George SZ, Childs JD, Teyhen DS, et al. Brief psychosocial education, not core stabilization, reduced incidence of low back pain: results from the Prevention of Low Back Pain in the Military (POLM) cluster randomized trial. *BMC Med*. 2011;9:128. https://doi.org/10.1186/1741-7015-9-128
6. Grier TL, Morrison S, Knapik JJ, Canham-Chervak M, Jones BH. Risk factors for injuries in the U.S. Army Ordnance School. *Mil Med*. 2011;176:1292-1299. https://doi.org/10.7205/MILMED-D-11-00215
7. Hauret KG, Jones BH, Bullock SH, Canham-Chervak M, Canada S. Musculoskeletal injuries: description of an under-recognized injury problem among military personnel. *Am J Prev Med*. 2010;38:S61-S70. https://doi.org/10.1016/j.amepre.2009.10.021
8. Hauret KG, Pacha L, Taylor BJ, Jones BH. Surveillance of disease and nonbattle injuries during US Army operations in Afghanistan and Iraq. *US Army Med Dep J*. 2016:15-23.
9. Hauret KG, Taylor BJ, Clemmons NS, Block SR, Jones BH. Frequency and causes of nonbattle injuries air evacuated from Operations Iraqi Freedom and Enduring Freedom, U.S. Army, 2001-2006. *Am J Prev Med*. 2010;38:S94-S107. https://doi.org/10.1016/j.amepre.2009.10.022
10. Henderson NE, Knapik JJ, Shaffer SW, McKenzie TH, Schneider GM. Injuries and injury risk factors among men and women in U.S. Army Combat Medic advanced individual training. *Mil Med*. 2000;165:647-652. https://doi.org/10.1093/milmed/165.9.647
11. Hootman JM, Dick R, Agel J. Epidemiology of collegiate injuries for 15 sports: summary and recommendations for injury prevention initiatives. *J Athl Train*. 2007;42:311-319.
12. Jones BH, Canham-Chervak M, Canada S, Mitchener TA, Moore S. Medical surveillance of injuries in the U.S. Military: descriptive epidemiology and recommendations for improvement. *Am J Prev Med*. 2010;38:S42-S60. https://doi.org/10.1016/j.amepre.2009.10.014
13. Knapik JJ, Darakjy S, Scott S, Hauret KG, Canada S. Evaluation of Two Army Fitness Programs: The TRADOC Standardized Physical Training Program for Basic Combat Training and the Fitness Assessment Program. Aberdeen Proving Ground, MD: Army Center for Health Promotion and Preventive Medicine; 2004.
14. Knapik JJ, Graham B, Cobbs J, Thompson D, Steelman R, Jones BH. A prospective investigation of injury incidence and injury risk factors among Army recruits in military police training. *BMC Musculoskelet Disord*. 2013;14:32. https://doi.org/10.1186/1471-2474-14-32
15. Knapik JJ, Graham B, Cobbs J, Thompson D, Steelman R, Jones BH. A prospective investigation of injury incidence and risk factors among Army recruits in combat engineer training. *J Occup Med Toxicol*. 2013;8:5. https://doi.org/10.1186/1745-6673-8-5
16. Knapik JJ, Jones SB, Darakjy S, et al. Injury rates and injury risk factors among U.S. Army wheel vehicle mechanics. *Mil Med*. 2007;172:988-996. https://doi.org/10.7205/MILMED.172.9.988
17. Knapik JJ, McCollam R, Canham-Chervak M, et al. Injuries and injury prevention among senior military officers at the Army War College. *Mil Med*. 2002;167:593-599. https://doi.org/10.1093/milmed/167.7.593
18. Knapik JJ, Spiess A, Grier T, et al. Injuries before and after deployments to Afghanistan and Iraq. *Public Health*. 2012;126:498-506. https://doi.org/10.1016/j.puhe.2012.01.031
19. Nindl BC, Williams TJ, Deuster PA, Butler NL, Jones BH. Strategies for optimizing military physical readiness and preventing musculoskeletal injuries in the 21st century. *US Army Med Dep J*. 2013:5-23.
20. Reynolds K, Cosio-Lima L, Bovill M, Tharion W, Williams J, Hodges T. A comparison of injuries, limited-duty days, and injury risk factors in infantry, artillery, construction engineers, and special forces soldiers. *Mil Med*. 2009;174:702-708. https://doi.org/10.7205/MILMED-D-02-2008
21. Rhon DI. A physical therapist experience, observation, and practice with an infantry brigade combat team in support of Operation Iraqi Freedom. *Mil Med*. 2010;175:442-447. https://doi.org/10.7205/MILMED-D-09-00097
22. Roy TC. Diagnoses and mechanisms of musculoskeletal injuries in an infantry brigade combat team deployed to Afghanistan evaluated by the brigade physical therapist. *Mil Med*. 2011;176:903-908. https://doi.org/10.7205/MILMED-D-11-00006
23. Roy TC, Ritland BM, Sharp MA. A description of injuries in men and women while serving in Afghanistan. *Mil Med*. 2015;180:126-131. https://doi.org/10.7205/MILMED-D-14-00321
24. Ruscio BA, Jones BH, Bullock SH, et al. A process to identify military injury prevention priorities based on injury type and limited duty days. *Am J Prev Med*. 2010;38:S19-S33. https://doi.org/10.1016/j.amepre.2009.10.004
25. Sears JM, Blanar L, Bowman SM, Adams D, Silverstein BA. Predicting work-related disability and medical cost outcomes: estimating injury severity scores from workers' compensation

data. *J Occup Rehabil*. 2013;23:19-31. https://doi.org/10.1007/s10926-012-9377-x
26. Sears JM, Bowman SM, Rotert M, Hogg-Johnson S. A new method to classify injury severity by diagnosis: validation using workers' compensation and trauma registry data. *J Occup Rehabil*. 2015;25:742-751. https://doi.org/10.1007/s10926-015-9582-5
27. Segui-Gomez M, MacKenzie EJ. Measuring the public health impact of injuries. *Epidemiol Rev*. 2003;25:3-19. https://doi.org/10.1093/epirev/mxg007
28. Shaffer SW, Moore JH. US Army physical therapist roles and contributions in Operations Enduring Freedom and Iraqi Freedom. *US Army Med Dep J*. 2016:52-57.
29. Teyhen DS, Shaffer SW, Butler RJ, et al. What risk factors are associated with musculoskeletal injury in US Army Rangers? A prospective prognostic study. *Clin Orthop Relat Res*. 2015;473:2948-2958. https://doi.org/10.1007/s11999-015-4342-6

MORE INFORMATION
WWW.JOSPT.ORG

APPENDIX

MONTHLY FOLLOW-UP SURVEYS

Introduction

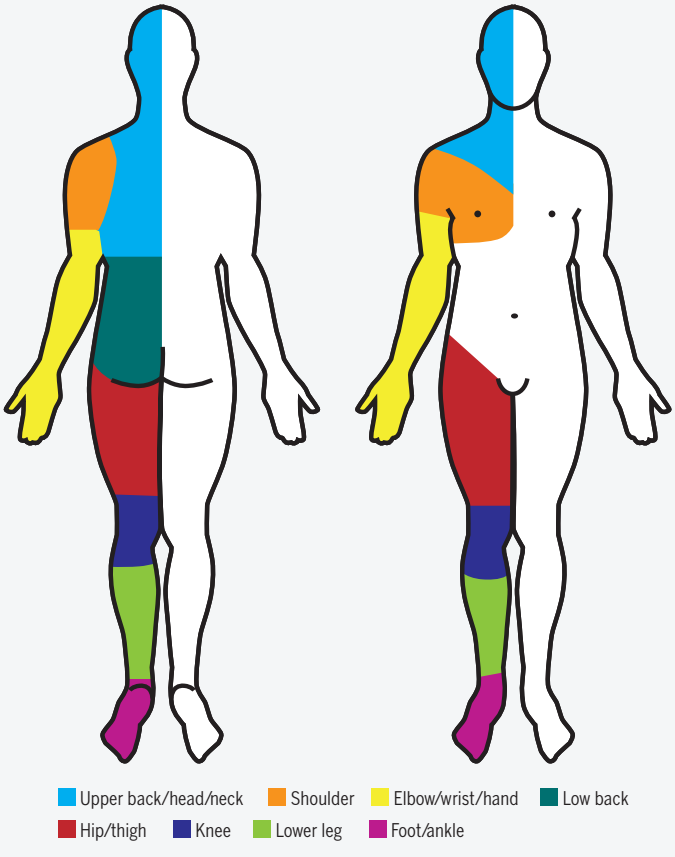
Thank you for your support of the MP3 study. These data and your feedback are critical for us to be able to predict injuries and hopefully target treatments that will reduce injuries in active-duty service members. Your results are confidential and will not be shared with your command group. The survey should take less than 10 minutes to complete.

Introductory Questions

1. Since the last time you answered an MP3 study survey, which option best describes your military status? Are you:
   A. Active duty
   B. Activated or mobilized reservist or National Guard
   C. Drilling reservist or National Guard
   D. No longer in the service
2. Which option best describes your current location? Are you:
   A. Stateside
   B. Overseas without combat pay
   C. Overseas with combat pay
3. Over the last month (or since the last time you completed one of the MP3 study surveys), have you had any injury that resulted in pain or decreased function lasting more than 48 hours or resulted in medical care?
   A. Yes
   B. No

(Survey stops if answering no to question 3)

4. Select the region of the body that was injured/causing pain. You may select more than 1 region.



APPENDIX

Region-Specific Questions

For each body region selected, answer the following 8 questions.

1. What month did your [insert region] injury (pain) start?
   A. January
   B. February
   C. March
   D. April
   E. May
   F. June
   G. July
   H. August
   I. September
   J. October
   K. November
   L. December

2. Have you received any medical treatment for your [insert region] injury (pain and/or decreased function) since completing the last survey? (Medical treatment includes both traditional medicine and alternative medicine treatments provided by a professional practitioner.)
   A. No
   B. Yes

3. Have you been placed on profile due to your [insert region] injury (pain and/or decreased function) since completing your last survey?
   A. No
   B. Yes

4. How long did this [insert region] injury (pain) cause you to restrict your own activity or result in the official restriction of activity through a profile/limited-duty status?
   A. Less than 1 week
   B. 8 to 21 days
   C. 22 to 30 days
   D. Greater than 30 days but less than 90 days
   E. Greater than 90 days but less than 180 days
   F. Greater than 180 days but less than 365 days
   G. Greater than 365 days

5. Was your [insert region] injury (pain) caused by:
   Upper back, neck, or head region?
      A. Trauma (ie, whiplash, car accident, fall)
      B. Combat trauma (ie, gunshot wound, blast injury, vehicle rollover from explosion)
      C. Overuse: "developed over time" (ie, neck pain, headaches)
   Shoulder region?
      A. Trauma (ie, dislocation, subluxation, fracture)
      B. Combat trauma (ie, gunshot wound, blast injury, vehicle rollover from explosion)
      C. Overuse: "developed over time" (ie, bursitis, tendinitis, rotator cuff injury)
   Elbow, wrist, or hand region?
      A. Trauma (ie, dislocation, subluxation, fracture)
      B. Combat trauma (ie, gunshot wound, blast injury, vehicle rollover from explosion)
      C. Overuse: "developed over time" (ie, bursitis, tendinitis, carpal tunnel syndrome)
   Lower back region?
      A. Trauma (ie, car accident, fall)
      B. Combat trauma (ie, gunshot wound, blast injury, vehicle rollover from explosion)
      C. Overuse: "developed over time" (ie, mechanical low back pain, chronic low back pain, sacroiliac joint pain)

APPENDIX

Hip and thigh region?
- A. Trauma (ie, dislocation, subluxation, fracture)
- B. Combat trauma (ie, gunshot wound, blast injury, vehicle rollover from explosion)
- C. Overuse: "developed over time" (ie, bursitis, tendinitis, stress fracture)

Knee region?
- A. Trauma (ie, ligamentous injury, meniscal injury, fracture)
- B. Combat trauma (ie, gunshot wound, blast injury, vehicle rollover from explosion)
- C. Overuse: "developed over time" (ie, bursitis, tendinitis, anterior knee pain)

Lower leg region?
- A. Trauma (ie, fracture, acute compartment syndrome)
- B. Combat trauma (ie, gunshot wound, blast injury, vehicle rollover from explosion)
- C. Overuse: "developed over time" (ie, stress fractures, shin splints, tendinitis, nerve injury)

Foot and ankle region?
- A. Trauma (ie, fracture, sprain, strain)
- B. Combat trauma (ie, gunshot wound, blast injury, vehicle rollover from explosion)
- C. Overuse: "developed over time" (ie, stress fracture, tendinitis, plantar fasciitis, heel pain)

6. Select what you think caused your injury:
   - A. Motor vehicle accident
   - B. Fall
   - C. Physical training
   - D. Sports (not associated with physical training)
   - E. Increased load carriage
   - F. Other, work related (not physical training or load carriage)
   - G. Other, not listed

7. On a scale from 0 to 10, with 0 being no pain at all and 10 being the worst pain imaginable, where would you rate your worst pain since you last filled out this survey?
   __/10

8. On a scale from 0 to 100, with 0 being no function at all and 100 being full normal function for you, please rate your ability to function using your [insert region] right now.
   ___/100

Thank you for your time!