DEREK CLEWLEY, DPT, PhD[1] • **YUSRA IFTIKHAR**, DPT[1]
MAGGIE E. HORN, DPT, PhD, MPH[1,2] • **DANIEL I. RHON**, DPT, DSc[1,3]

Do the Number of Visits and the Cost of Musculoskeletal Care Improve Outcomes? More May Not Be Better

Minimizing costs while maximizing health-related outcomes characterizes value-based health care. Value-based health care is the magnitude of a health-related outcome achieved per dollar spent.[1] The relationship between the amount of medical services provided and the change in a health-related outcome is more important than addressing each in isolation when determining the value of health care.[20,21]

Patient satisfaction scores are highly valued by health care systems. The paradox is that patients are often more satisfied when they receive additional health care, even if it does not directly improve outcomes.[7] Clinical practice variation between providers is related to overuse.[32,33] A medical provider may perceive a need to choose between what satisfies a patient and what is backed by evidence and concordant with practice guidelines.[29] These choices can cause conflict, especially when, in some cases, no care at all could be equally effective.[3,22,28,31]

Pressure from patients was one of the top influencers behind clinical decisions in a large survey of physicians in the United States, and the pressure to improve patient satisfaction was a driver of inappropriate medical care.[15] Thus, measuring outcomes in the context of the volume of health care use for a given condition is necessary. At the same time, there is likely more to prognosis and outcomes than the type and volume of the intervention alone.

Shoulder pain is one of the most common musculoskeletal conditions, with a reported lifetime point prevalence of 67%.[14] Shoulder pain accounts for one third of physician office visits for musculoskeletal pain, and the most frequent culprit is subacromial pain syndrome (SAPS).[18,34,36] Patients with persistent pain account for 74% of the total costs associated with management of shoulder pain.[12]

We provide an example of how value could be assessed in clinical practice. We assessed the relationship between health care use and the magnitude of change in patient outcomes (pain and disability) in

OBJECTIVES: To determine the relationship between health care use and the magnitude of change in patient-reported outcomes in individuals who received treatment for subacromial pain syndrome. The secondary objective was to determine the value of care, as measured by change in pain and disability per dollar spent.

DESIGN: Secondary analysis of a randomized clinical trial that investigated the effects of nonsurgical care for subacromial pain syndrome.

METHODS: Two groups of treatment responders were created, based on 1-year change in Shoulder Pain and Disability Index (SPADI) score (high, 46.83 points; low, 8.21 points). Regression analysis was performed to determine the association between health care use and 1-year change in SPADI score. Baseline SPADI score was used as a covariate in the regression analysis. Value was measured by comparing health care visits and costs expended per SPADI 1-point change between responder groups.

RESULTS: Ninety-eight patients were included; 38 were classified as high responders (mean 1-year SPADI change score, 46.83 points) and 60 were classified as low responders (1-year SPADI change score, 8.21 points). Neither unadjusted medical visits (5.89; 95% confidence interval [CI]: 4.35, 7.44 versus 6.30; 95% CI: 5.14, 7.46) nor medical costs ($1404.86; 95% CI: $1109.34, $1779.09 versus $1679.26; 95% CI: $1391.54, $2026.48) were significantly different between high and low responders, respectively.

CONCLUSION: Neither the number of visits nor the financial cost of nonsurgical shoulder-related care was associated with improvement in shoulder pain and disability at 1 year. *J Orthop Sports Phys Ther 2020;50(11):642-648. doi:10.2519/jospt.2020.9440*

KEY WORDS: *health services research, musculoskeletal, outcomes, pain, shoulder, value-based care*

[1]Division of Doctor of Physical Therapy, Department of Orthopaedic Surgery, Duke University School of Medicine, Durham, NC. [2]Department of Population Health Sciences, Duke University School of Medicine, Durham, NC. [3]Center for the Intrepid, Brooke Army Medical Center, San Antonio, TX. The project was approved by the US Army Western Regional Medical Command Institutional Review Board. All study procedures and analyses were compliant with the Health Insurance Portability and Accountability Act. The view(s) expressed herein are those of the author(s) and do not reflect the official policy or position of Brooke Army Medical Center, the US Army Medical Department, the US Army Office of the Surgeon General, the Department of the Army, the Defense Health Agency, the Department of Defense, or the US Government. 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 Derek Clewley, Duke University School of Medicine, Department of Orthopaedic Surgery, Division of Doctor of Physical Therapy, 2200 West Main Street, Suite B220, Durham, NC 27710. E-mail: derek.clewley@duke.edu

individuals receiving treatment for SAPS. Our first aim was to determine whether the amount of shoulder-related health care use (number of visits and costs) was associated with the magnitude of change in pain and disability. Our second aim was to determine the value of care, measured as the change in pain and disability per dollar spent, provided to patients receiving care for SAPS, and to compare these values between high- and low-responder groups.

METHODS

We performed a secondary analysis of a randomized clinical trial offering nonsurgical treatment approaches for SAPS,[23] which included 104 patients seeking care in a military hospital between June 2010 and March 2012. Patients were randomized to receive either a corticosteroid injection or manual therapy plus exercise. Both treatments were consistent with the treatment options most likely available and used by patients not participating in the study (standard care). Additional details regarding the design and primary results of this study have been reported previously.[23,24]

Ethics approval was provided by the US Army Western Regional Medical Command Institutional Review Board. All study procedures and analyses were compliant with the Health Insurance Portability and Accountability Act.

Development of Groups of Responders to Shoulder Interventions

We created 2 SAPS treatment responder groups, based on changes in self-reported disability on the Shoulder Pain and Disability Index (SPADI), which was the primary outcome measure in the trial (1-year change). The SPADI is a 13-item self-report questionnaire scored from 0 to 100, with higher scores indicating greater disability. It is reliable and valid for measuring pain, function, and disability in individuals with shoulder disorders.[26,27] The SPADI is responsive to change in patients with SAPS, with a minimum clinically important difference (MCID) between 8 and 13 points and a minimum detectable change (the minimum change that is sure to not be the result of measurement error) of 18 points.[19,27] We calculated change score as 1-year score minus baseline score.

Data were analyzed regardless of original treatment assignment. To establish a more robust, structured measure of responder groups, we used a fixed 2-group, 2-step cluster analysis to split the cohort into distinct groups based on SPADI change score at 1 year. The responder groups represented a complete mix of available treatments from the original study. A 2-step cluster model identifies subgroups based on selected criterion variables, allowing for inclusion of continuous and categorical variables.[11] When considering inputs for cluster analysis, the strength is determined by silhouette, which is a measure of consistency within clusters of data.[11] Silhouette values range between –1 and +1, with values closer to 1 being optimal.[11] The more closely a case is matched to its identified cluster group, and the more poorly it is matched to its neighboring group, the stronger the cluster.

Baseline Characteristics and Covariates

Baseline characteristics included age, sex, body mass index (BMI), duration of symptoms, previous health care use, and tobacco use.[4,5] We also collected baseline SPADI, numeric pain-rating scale (NPRS), and Fear-Avoidance Beliefs Questionnaire (FABQ; both physical activity and work subscales) scores. Comorbidities that were assessed included metabolic syndromes, mental health disorders, cardiovascular diseases, and sleep disorders, identified based on diagnoses rendered by medical providers and entered into electronic medical records during the prior year. The rationale for the use of specific International Classification of Diseases, ninth revision (ICD-9) codes to identify these comorbidities and their influence on orthopaedic outcomes has been previously published.[25] These baseline characteristics could serve as potential covariates that might otherwise explain higher rates of health care utilization for an individual. Nominal variables were compared using chi-square analyses and continuous variables were evaluated using *t* tests.

Health Care Use

Health care use for this cohort was abstracted from the single-payer Military Health System (MHS) data repository (MDR). Data were extracted at the person level, using diagnostic (ICD-9) and procedural (Current Procedural Terminology) codes. Procedure codes present with shoulder diagnoses were classified as shoulder-related care, and their corresponding costs and medical visit counts were marked.[25] The MDR captures any care around the world, in both military and civilian clinics, where TRICARE (MHS insurance entity) is the payer.

In military clinics, costs are assigned to procedures for business planning purposes by the US Defense Health Agency, but costs from civilian clinics are true claims costs paid by TRICARE. These costs are not equal or comparable; therefore, for our analysis, we only analyzed shoulder care taking place inside military clinics (all original study-related care and the majority of additional care took place in military clinics).

Outcomes of interest were (1) the total number of shoulder-related medical visits and (2) the total cost for all shoulder-related medical care for a period of 1 year. The differences in mean costs between groups, based on the intervention received in the original trial, were not significant (physical therapy, \$1691; 95% confidence interval [CI]: \$1184, \$2417; corticosteroid injection, \$1468; 95% CI: \$1073, \$2008; $P = .559$).

Power Analysis

We investigated the sample size needed to conduct a cluster analysis. While a cluster analysis is designed to handle large data sets, to our knowledge, there are no strict rules that establish a minimum sample size for this analysis or would preclude our use of it in this scenario. However, it is advisable to use fewer clusters if the sample size is small. Trying to identify too many similar factors within a small sample can dilute the homogeneity of groups being sought. This is one of the reasons why we chose to fix the data into a 2-group cluster instead of allowing the cluster analysis to determine the number of clusters. Furthermore, the number of inputs for clustering should be small. We used 1-year SPADI change score as the only input.

Data Analysis

Descriptive statistics were used to characterize the responder groups. Health care use is often skewed right, with most people using few resources and a small minority using many resources.[16,17,37] These models require adjustments for nonnormal distribution, which are appropriate for generalized linear models (GLMs).[16] For all cost outcomes, we used a gamma log transformation; for visit outcomes (count), we used a negative binomial regression. Outcomes were reported as estimated marginal mean and 95% CI, with a *P* value of less than .05 indicating statistical significance. All analyses were completed using SPSS Version 23.0 (IBM Corporation, Armonk, NY).

Aim 1

We compared costs and visits between the 2 responder groups. We ran 2 sets of analyses, unadjusted and adjusted models. In the adjusted model, we controlled for the presence of comorbidities. We also controlled for demographic characteristics, including age, sex, and BMI, baseline NPRS score, duration of symptoms (measured in days before index), baseline SPADI score, FABQ physical activity subscale score, FABQ work subscale score, and prior patterns of health care utilization, including both costs and visits.

Aim 2

To quantify value, we derived 2 value metrics: the amount of dollars spent for every 1-point change in the SPADI and the number of medical visits invested for each SPADI point change, from baseline to 1 year. We took total shoulder-related cost expensed during 1 year and divided it by the 1-year SPADI change score. We did the same for shoulder-related visits.

RESULTS

The original trial excluded 6 of the originally enrolled patients who never received treatment or followed up after entry into the study, which resulted in a final sample of 98 patients. The average ± SD 1-year SPADI change score across the entire sample was 23.19 ± 25.51 points, and the mean ± SD baseline SPADI score was 45.99 ± 16.81 points. The mean ± SD baseline NPRS score was 3.6 ± 2.3 points. The sample was primarily male (n = 67, 68.4%) and right-hand dominant (n = 83, 84.7%), with a mean ± SD age of 41.3 ± 12.1 years. The mean ± SD duration of shoulder symptoms was 175 ± 340 days before enrolling in the trial. The shoulder pain was on the dominant side in 52 (53.1%) patients. Tobacco use was reported by 17 (17.4%) patients. Mean ± SD BMI was 28.58 ± 4.61 kg/m^2. Comorbidities present in the prior year included sleep disorders (n = 33, 33.7%), cardiovascular disease (n = 19, 19.4%), mental health disorders (n = 19, 19.4%), and metabolic disorders (n = 10, 10.2%).

Responder Characteristics

The resultant memberships of the 2-step cluster were strong, with good cohesion and internal consistency (silhouette score, 0.7).[11] There were 38 patients (38.8% of the entire cohort) identified as high responders, with a mean ± SD 1-year SPADI change score of 46.83 ± 12.28 points. The low-responder group included 60 patients (61.2% of the entire cohort), with a mean ± SD 1-year SPADI change score of 8.21 ±19.66 points. The high-responder group had higher mean ± SD baseline SPADI scores (55.32 ± 13.56 points) compared to the low-responder group (39.27 ± 15.11 points; *P*<.01). The high-responder group was significantly older than the low-responder group, with a mean ± SD age of 44.26 ± 11.96 years compared to 38.53 ± 11.53 years (*P* = .02). No other baseline characteristics were significantly different between the 2 groups of responders (**TABLE 1**).

Health Care Use

Goodness-of-fit statistics can be used to estimate model fit for GLMs. All of the models suffered from slight overdispersion, but nothing large enough to violate assumptions. The model without any covariates had the best fit, but we reported both unadjusted and adjusted values for the association between health care use and responder level (**TABLE 2**). No significant association was found between the total number of shoulder-related medical visits and the magnitude of improvement. Mean shoulder-related visits did not differ significantly between the high (5.89 visits; 95% CI: 4.35, 7.44) and low responders (6.30 visits; 95% CI: 5.14, 7.46).

Medical costs for shoulder care were not significantly different between high ($1404.86; 95% CI: $1109.34, $1779.09) and low responders ($1679.26; 95% CI: $1391.54, $2026.48). Five patients had additional shoulder-related care in civilian hospital clinics during the 1-year follow-up (2 high responders and 3 low responders). The overall results were unchanged after adjusting for covariates (including all comorbidities) and with any of the sensitivity analyses (**TABLE 2**).

Value of Care

It cost a mean of $31.63 (95% CI: $24.21, $41.32) per SPADI 1-point change in the higher responders, compared to $120.85 (95% CI: $94.01, $155.36) in the low responders (**TABLE 3**). The mean SPADI point improvement per shoulder visit was 12.23 (95% CI: 9.62, 14.84) in the high-responder group, compared to 2.33 (95% CI: 0.25, 4.40) in the low-responder group.

Sensitivity Analyses

To ensure fidelity of the results, we performed a series of sensitivity analyses.

We conducted the same analysis with (1) grouping patients based on meeting the MCID for the SPADI (8 or more points) and (2) grouping patients based on falling above or below the median 1-year SPADI change score for the entire cohort. We performed a final sensitivity analysis using a different calculation for change score that incorporated baseline SPADI score. The overall results remained unchanged after all of these additional analyses.

DISCUSSION

In this study, we sought to answer the question, "Does more health care use result in greater magnitudes of change in pain and disability?" We found that neither the amount of visits nor the cost spent on care predicted better outcomes. In other words, on average, both high and low responders spent the same amount on shoulder-related health care. More is not necessarily better, and, while it may seem intuitive, this reinforces the notion that many factors beyond health care use alone may contribute to successful outcomes. Adoption of these metrics in clinical studies may yield a novel approach to help identify and measure value-based care.

Health Care Use

Quantity of health care can be measured using visits or cost variables.[4] However, these variables are sometimes interdependent, such that an increase in health care visits can directly impact costs. Observing visit count data is important because it provides a standardized proxy measure for health care use and potentially a behavioral component, compared to costs alone. Visit counts also provide valuable information about the patterns of health care use. Measuring patterns of health care use using visit or encounter data has been used as both an outcome measure[9] as well as a predictor of future health care use.[4,5]

Value-Based Care

Value-based health care is operationalized by the amount of medical resources used to obtain a given amount of change. Health care use and patient self-reported outcomes are important value-based measures.[13] Alone, health care use is limited in its ability to explain treatment effectiveness. By the same token, a self-reported outcome alone cannot describe what effort went into making that change. Combining self-report measures and usage metrics can provide information of greater utility for determining the value of services.[13] Some might argue that the more services provided, the better the outcomes should be. For example, underdosing or

TABLE 1 Baseline Characteristics[a]

	Full Sample (n = 98)	Low-Responder Group (n = 60)	High-Responder Group (n = 38)	*P* Value
SPADI score change: baseline to 1 y	23.19 ± 25.51	8.21 ± 19.66	46.83 ± 12.28	<.01
Baseline SPADI score	45.99 ± 16.81	39.27 ± 15.11	55.32 ± 13.56	<.01
Baseline NPRS score	3.60 ± 2.34	3.35 ± 1.78	3.89 ± 3.08	.27
Duration of symptoms (on index date), d	175 ± 340	191 ± 412	145 ± 142	.57
Age, y	41.30 ± 12.10	38.53 ± 11.53	44.26 ± 11.96	.02
Sex (male), n (%)	67 (68.4)	45 (75.0)	22 (57.9)	.08
Body mass index, kg/m^2	28.58 ± 4.61	27.90 ± 4.24	29.46 ± 4.74	.10
Tobacco smoking, n (%)	17 (17.4)	12 (20.0)	5 (13.2)	.38
Right-hand dominance, n (%)	83 (84.7)	48 (80.0)	35 (92.1)	.23
Shoulder pain on same side as hand dominance, n (%)	52 (53.1)	31 (51.7)	21 (55.3)	.75
Individuals with a comorbid condition at baseline, n (%)				
Mental health	19 (19.4)	9 (15.0)	10 (26.3)	.18
Cardiovascular	19 (19.4)	11 (18.3)	8 (21.1)	.77
Sleep disorders	33 (33.7)	21 (35.0)	12 (31.6)	.68
Metabolic disorders	10 (10.2)	4 (6.7)	6 (15.8)	.15
FABQ-work score	11.99 ± 9.86	12.53 ± 10.06	11.13 ± 9.60	.50
FABQ-physical activity score	15.15 ± 4.86	15.08 ± 4.70	15.26 ± 5.18	.86
Prior health care utilization factors				
Cost of all outpatient care	$4966.64 ± $6719.71	$4230.77 ± $5166.93	$6221.08 ± $8830.04	.17
Cost of all shoulder-related care	$596.00 ± $585.82	$548.99 ± $581.84	$712.85 ± $613.01	.19
Total outpatient medical visits	22.38 ± 18.92	20.76 ± 19.25	24.32 ± 18.35	.37
Total shoulder-related medical visits	2.23 ± 2.20	2.20 ± 2.46	2.32 ± 1.88	.81

Abbreviations: FABQ, Fear-Avoidance Beliefs Questionnaire; NPRS, numeric pain-rating scale; SPADI, Shoulder Pain and Disability Index.
[a]*Values are mean ± SD unless otherwise indicated.*

not performing a set number of exercise therapy sessions might lead to smaller effect sizes in patients with ankle and knee disorders.[38] In a study by Bergman et al,[2] additional shoulder care resulted in improved recovery, but at higher cost.

However, the opposite can also be the case: providing more services may be harmful. Examples include ordering imaging that is not warranted[8,22,30] or prescribing antibiotics when unnecessary.[10] Therefore, a greater focus on value rather than quantity is needed.

We found that similar amounts of shoulder-related visits and costs were used to achieve a higher level of treatment response. Even though both of the responder groups achieved the MCID, the high responders had significantly better SPADI change scores at 1 year. There was much greater value for the high responders, indicated by 74% less cost per SPADI 1-point change and a change of nearly 10 more SPADI points per visit than the low responders. One explanation is that some of the visits that patients received were more impactful than others, and could have been associated with less cost of delivery.

The relationship seen between magnitude of change in outcome and amount of health care use is likely influenced by the type of intervention. Treatments with known efficacy and effectiveness cannot be ignored. A little high-quality treatment may be better than copious poor-quality treatment. In our cohort, patients received guideline-adherent care with established and proven effectiveness.[6,35]

Plateau of Care

Measuring change in the context of the cost required to create that change is critical to defining value-based care, but also requires more granular surveillance to identify the optimal amount of care. For example, if 8 visits of exercise therapy for the management of SAPS are equally as effective as 12 visits, the cost incurred for the additional 4 visits could render the treatment cost-ineffective. A plateau can be reached with any effective treatment. In this regard, the same intervention can be of both high and low value, depending on when and how it is used. A treatment that was initially considered high value might become a low-value treatment when it continues to be used after a plateau has been reached. At the same time, some benefits may not manifest immediately and may require slightly longer periods (versus intersession changes) to determine their long-term effectiveness. Providing health care where no further value is obtained may lead to overuse.

There is a theoretical plateau that may have been reached in both responder groups, where additional visits added no further value. We did not capture changes at each session to determine the point of plateau. Without knowing the exact nature of each visit and the specific care rendered, it is challenging to provide definitive and granular conclusions regarding optimal treatments and dosing, which is a limitation of our study. Further research is needed to determine where this balance between too little and too much lies.

Limitations

We combined the entire cohort for analysis, even though they received 2 distinct treatments in the original trial (corticosteroid injection or physical therapy). It is possible that there are different health-seeking patterns seen in patients based on receipt of each of these interventions. It is also plausible that patterns of use for participants in a research study are different than if they were not participating in a study. However, the treatments offered to all participants are standard-of-care interventions with proven effectiveness. Patients were asked to minimize

TABLE 2 Unadjusted and Adjusted 1-Year Health Care Use[a]

	Low Responders (n = 60)	High Responders (n = 38)
Unadjusted		
Shoulder costs	$1679.26 ($1391.54, $2026.48)	$1404.86 ($1109.34, $1779.09)
Shoulder medical visits	6.30 (5.14, 7.46)	5.89 (4.35, 7.44)
Adjusted[b]		
Shoulder costs	$1750.51 ($1415.30, $2165.11)	$1445.45 ($1088.03, $1920.29)
Shoulder medical visits	6.15 (5.09, 7.21)	5.74 (4.41, 7.07)

[a]*Values are mean (95% confidence interval) per person. All analyses involved negative binomial logistic regression for count data and generalized linear models with gamma distribution for cost data.*
[b]*Adjusted analyses controlled for baseline Shoulder Pain and Disability Index score, baseline numeric pain-rating scale score, duration of symptoms, treatment group, age, sex, body mass index, mental health, cardiovascular disease, sleep disorder, metabolic disorder, baseline Fear-Avoidance Beliefs Questionnaire work and physical activity subscale scores, prior outpatient costs and visits, and prior shoulder costs and visits.*

TABLE 3 Value-Based Measures, Based on Cost and Visit Health Care Utilization and SPADI Change Score, at 1 Year[a]

	Low Responders (n = 60)	High Responders (n = 38)
Mean shoulder costs for a SPADI 1-point change	$120.85 ($94.01, $155.36)	$31.63 ($24.21, $41.32)
Mean SPADI point change for each shoulder visit	2.33 (0.25, 4.40)	12.23 (9.62, 14.84)

Abbreviation: SPADI, Shoulder Pain and Disability Index.
[a]*Values are mean (95% confidence interval) per person. All analyses involved negative binomial logistic regression for count data and generalized linear models with gamma distribution for cost data.*

additional care outside the study during the period of intervention, which might have limited health care seeking, but this message was conveyed to both treatment groups equally. In addition, shoulder-related health care costs were not significantly different between the 2 original intervention groups, indicating a similar volume of care. Finally, the fact that patients within both treatment groups were equally represented within both responder-based clusters likely increased the generalizability of these findings beyond a single treatment alone.

The mean baseline SPADI score in the high-responder group was 16 points higher than in the low-responder group. This afforded the high-responder group a greater opportunity for improvement, as there was more room for change. However, at 1 year, the high-responder group had a mean SPADI score of 8.49 (low levels of pain and disability), compared to the low-responder group mean SPADI score of 31.06 (mild to moderate levels of pain and disability). Future models should consider how to best address floor and ceiling effects of baseline scores for value-based models such as these. All else being equal, value will seem most apparent for those with higher baseline levels of pain and disability.

We conducted a series of sensitivity analyses to assess the fidelity of the results. In every analysis, the baseline score was significantly different. This information may be useful when considering potential pay-for-performance reimbursement models. Using baseline outcome scores may be necessary for improving the predictive accuracy of a high or low responder. Considerations for patients with lower disability at baseline, and therefore less room for change, will be necessary when evaluating models that pay based on outcomes achieved. More research is needed in this area to identify ideal cost/outcome measures.

Patients were seen in the MHS, a single-payer system, and results may be different in other third-party-payer systems. The results could be different using other outcome measures for the shoulder, with other types of shoulder disorders (adhesive capsulitis, instability, etc), and even in other body regions.

CONCLUSION

For patients with SAPS receiving similar interventions and dosing, the 1-year improvement was not based on the number of visits or the total costs spent on medical care for the shoulder. High responders had a significantly lower cost per SPADI 1-point change and a significantly greater SPADI point change per visit than low responders. ◉

KEY POINTS

FINDINGS: The amount of health care use was not associated with changes in pain and disability. High and low responders spent the same amount of money on shoulder-related health care and had a similar number of health care visits. The value of care was considerably better in the high-responder group compared to the low-responder group.

IMPLICATIONS: More health care may not necessarily contribute to better health care outcomes.

CAUTION: This was a secondary analysis of a randomized controlled trial whereby the entire cohort was combined and then reclassified based on responder type.

STUDY DETAILS

AUTHOR CONTRIBUTIONS: Dr Clewley contributed to the concept of the study design, analysis and interpretation of data, and writing of the manuscript, including all revisions. Dr Rhon contributed to the concept of the study design and to data collection, and assisted with analysis and interpretation of data and the revision and writing of the manuscript. Dr Horn assisted with the study design, the analysis and interpretation of data, and the revision and writing of the manuscript. Yusra Iftikhar contributed to the writing and revision of the manuscript.

DATA SHARING: Data are available on reasonable request; however, the data from the Military Health System data repository are proprietary to the US Defense Health Agency. A Data Sharing Agreement Application must be submitted to the Defense Health Agency, and the application form can be found at www.health.mil. Please contact the corresponding author for any questions about this process.

PATIENT AND PUBLIC INVOLVEMENT: This was a secondary analysis of an original trial previously published. There was no public or patient involvement in the design, conduct, interpretation, or translation of the research.

REFERENCES

1. Annaswamy TM, Kasitinon D, Royston A. Value-based care and musculoskeletal rehabilitation. *Curr Phys Med Rehabil Rep*. 2018;6:49-54. https://doi.org/10.1007/s40141-018-0176-y
2. Bergman GJ, Winter JC, van Tulder MW, Meyboom-de Jong B, Postema K, van der Heijden GJ. Manipulative therapy in addition to usual medical care accelerates recovery of shoulder complaints at higher costs: economic outcomes of a randomized trial. *BMC Musculoskelet Disord*. 2010;11:200. https://doi.org/10.1186/1471-2474-11-200
3. Bobbio M, Vernero S. Choosing Wisely, the reasons for its success. *Monaldi Arch Chest Dis*. 2019;89:117-118. https://doi.org/10.4081/monaldi.2019.1104
4. Clewley D, Rhon D, Flynn T, Koppenhaver S, Cook C. Health seeking behavior as a predictor of healthcare utilization in a population of patients with spinal pain. *PLoS One*. 2018;13:e0201348. https://doi.org/10.1371/journal.pone.0201348
5. Clewley D, Rhon DI, Flynn TW, Sissel CD, Cook CE. Does health care utilization before hip arthroscopy predict health care utilization after surgery in the US Military Health System? An investigation into health-seeking behavior. *J Orthop Sports Phys Ther*. 2018;48:878-886. https://doi.org/10.2519/jospt.2018.8259
6. Diercks R, Bron C, Dorrestijn O, et al. Guideline for diagnosis and treatment of subacromial pain syndrome: a multidisciplinary review by the Dutch Orthopaedic Association. *Acta Orthop*. 2014;85:314-322. https://doi.org/10.3109/17453674.2014.920991
7. Fenton JJ, Jerant AF, Bertakis KD, Franks P. The cost of satisfaction: a national study of patient satisfaction, health care utilization, expenditures, and mortality. *Arch Intern Med*. 2012;172:405-411. https://doi.org/10.1001/archinternmed.2011.1662
8. Ferrari R. Choosing Wisely: spine imaging [letter]. *Am J Manag Care*. 2016;22:e302-e303.
9. Horn ME, Fritz JM. Timing of physical therapy

consultation on 1-year healthcare utilization and costs in patients seeking care for neck pain: a retrospective cohort. *BMC Health Serv Res*. 2018;18:887. https://doi.org/10.1186/s12913-018-3699-0

10. Jung N, Koop H, Riessen R, Galle JC, Jany B, Märker-Hermann E. ["Choosing Wisely" in infectious diseases: overuse of antibiotics – too few vaccinations]. *Internist (Berl)*. 2016;57:527-531. https://doi.org/10.1007/s00108-016-0063-5

11. Kaufman L, Rousseeuw PJ. *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken, NJ: Wiley; 2005.

12. Kuijpers T, van Tulder MW, van der Heijden GJ, Bouter LM, van der Windt DA. Costs of shoulder pain in primary care consulters: a prospective cohort study in The Netherlands. *BMC Musculoskelet Disord*. 2006;7:83. https://doi.org/10.1186/1471-2474-7-83

13. Lee KHK, Austin JM, Pronovost PJ. Developing a measure of value in health care. *Value Health*. 2016;19:323-325. https://doi.org/10.1016/j.jval.2014.12.009

14. Luime JJ, Koes BW, Hendriksen IJ, et al. Prevalence and incidence of shoulder pain in the general population; a systematic review. *Scand J Rheumatol*. 2004;33:73-81. https://doi.org/10.1080/03009740310004667

15. Lyu H, Xu T, Brotman D, et al. Overtreatment in the United States. *PLoS One*. 2017;12:e0181970. https://doi.org/10.1371/journal.pone.0181970

16. Malehi AS, Pourmotahari F, Angali KA. Statistical models for the analysis of skewed healthcare cost data: a simulation study. *Health Econ Rev*. 2015;5:art 11. https://doi.org/10.1186/s13561-015-0045-7

17. Manning WG, Basu A, Mullahy J. Generalized modeling approaches to risk adjustment of skewed outcomes data. *J Health Econ*. 2005;24:465-488. https://doi.org/10.1016/j.jhealeco.2004.09.011

18. Östör AJ, Richards CA, Prevost AT, Speed CA, Hazleman BL. Diagnosis and relation to general health of shoulder disorders presenting to primary care. *Rheumatology (Oxford)*. 2005;44:800-805. https://doi.org/10.1093/rheumatology/keh598

19. Paul A, Lewis M, Shadforth MF, Croft PR, van der Windt DA, Hay EM. A comparison of four shoulder-specific questionnaires in primary care. *Ann Rheum Dis*. 2004;63:1293-1299. https://doi.org/10.1136/ard.2003.012088

20. Porter ME. A strategy for health care reform—toward a value-based system. *N Engl J Med*. 2009;361:109-112. https://doi.org/10.1056/NEJMp0904131

21. Porter ME. Value-based health care delivery. *Ann Surg*. 2008;248:503-509.

22. Rao VM, Levin DC. The overuse of diagnostic imaging and the Choosing Wisely initiative. *Ann Intern Med*. 2012;157:574-576. https://doi.org/10.7326/0003-4819-157-8-201210160-00535

23. Rhon DI, Boyles RB, Cleland JA. One-year outcome of subacromial corticosteroid injection compared with manual physical therapy for the management of the unilateral shoulder impingement syndrome: a pragmatic randomized trial. *Ann Intern Med*. 2014;161:161-169. https://doi.org/10.7326/M13-2199

24. Rhon DI, Boyles RE, Cleland JA, Brown DL. A manual physical therapy approach versus subacromial corticosteroid injection for treatment of shoulder impingement syndrome: a protocol for a randomised clinical trial. *BMJ Open*. 2011;1:e000137. https://doi.org/10.1136/bmjopen-2011-000137

25. Rhon DI, Clewley D, Young JL, Sissel CD, Cook CE. Leveraging healthcare utilization to explore outcomes from musculoskeletal disorders: methodology for defining relevant variables from a health services data repository. *BMC Med Inform Decis Mak*. 2018;18:10. https://doi.org/10.1186/s12911-018-0588-8

26. Roach KE, Budiman-Mak E, Songsiridej N, Lertratanakul Y. Development of a shoulder pain and disability index. *Arthritis Care Res*. 1991;4:143-149. https://doi.org/10.1002/art.1790040403

27. Roy JS, MacDermid JC, Woodhouse LJ. Measuring shoulder function: a systematic review of four questionnaires. *Arthritis Rheum*. 2009;61:623-632. https://doi.org/10.1002/art.24396

28. Santhirapala R, Fleisher LA, Grocott MPW. Choosing Wisely: just because we can, does it mean we should? *Br J Anaesth*. 2019;122:306-310. https://doi.org/10.1016/j.bja.2018.11.025

29. Schers H, Wensing M, Huijsmans Z, van Tulder M, Grol R. Implementation barriers for general practice guidelines on low back pain: a qualitative study. *Spine (Phila Pa 1976)*. 2001;26:E348-E353. https://doi.org/10.1097/00007632-200108010-00013

30. Seely JM, Appavoo S, Wadden N, Loisel Y, Scaranelo A. Choosing Wisely Canada and diagnostic imaging: what level of evidence supports the recommendations? [letter]. *Can Assoc Radiol J*. 2018;69:338-339. https://doi.org/10.1016/j.carj.2018.05.001

31. Siwek J. Choosing Wisely: top interventions to improve health and reduce harm, while lowering costs. *Am Fam Physician*. 2012;86:128-133.

32. Tan A, Zhou J, Kuo YF, Goodwin JS. Variation among primary care physicians in the use of imaging for older patients with acute low back pain. *J Gen Intern Med*. 2016;31:156-163. https://doi.org/10.1007/s11606-015-3475-3

33. Thiels CA, Anderson SS, Ubl DS, et al. Wide variation and overprescription of opioids after elective surgery. *Ann Surg*. 2017;266:564-573. https://doi.org/10.1097/SLA.0000000000002365

34. Vecchio P, Kavanagh R, Hazleman BL, King RH. Shoulder pain in a community-based rheumatology clinic. *Br J Rheumatol*. 1995;34:440-442. https://doi.org/10.1093/rheumatology/34.5.440

35. Veen EJD, Stevens M, Koorevaar CT, Diercks RL. Appropriate care for orthopedic patients: effect of implementation of the Clinical Practice Guideline for diagnosis and treatment of subacromial pain syndrome in the Netherlands. *Acta Orthop*. 2019;90:191-195. https://doi.org/10.1080/17453674.2019.1593641

36. Wofford JL, Mansfield RJ, Watkins RS. Patient characteristics and clinical management of patients with shoulder pain in U.S. primary care settings: secondary data analysis of the National Ambulatory Medical Care Survey. *BMC Musculoskelet Disord*. 2005;6:4. https://doi.org/10.1186/1471-2474-6-4

37. Yelland LN, Salter AB, Ryan P. Performance of the modified Poisson regression approach for estimating relative risks from clustered prospective data. *Am J Epidemiol*. 2011;174:984-992. https://doi.org/10.1093/aje/kwr183

38. Young JL, Rhon DI, Cleland JA, Snodgrass SJ. The influence of exercise dosing on outcomes in patients with knee disorders: a systematic review. *J Orthop Sports Phys Ther*. 2018;48:146-161. https://doi.org/10.2519/jospt.2018.7637

ARMAGHAN DABBAGH, PT, MSc[1] • JOY C. MACDERMID, PT, PhD[1-3] • JOSHUA YONG, OT, MSc[4]
LUCIANA G. MACEDO, PT, PhD[2] • TARA L. PACKHAM, OT, PhD[2]

Diagnosing Carpal Tunnel Syndrome: Diagnostic Test Accuracy of Scales, Questionnaires, and Hand Symptom Diagrams—A Systematic Review

OBJECTIVE: To summarize and evaluate research on the accuracy of clinical diagnostic scales, questionnaires, and hand symptom diagrams/maps used for diagnosis of carpal tunnel syndrome (CTS).

DESIGN: Systematic review of diagnostic test accuracy.

LITERATURE SEARCH: A comprehensive literature search of the MEDLINE, CINAHL, and Embase databases was conducted on January 20, 2020.

STUDY SELECTION CRITERIA: Studies that assessed at least 1 diagnostic accuracy property of the scales, questionnaires, and hand symptom diagrams used for the diagnosis of CTS.

DATA SYNTHESIS: The Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines were followed. Risk of bias and applicability concerns were assessed using the revised Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) tool. Diagnostic accuracy properties were summarized.

RESULTS: Out of 4052 citations after removing duplicates, 21 articles met the inclusion criteria. Twelve articles reported on the diagnostic accuracy of scales and questionnaires, including the Bland questionnaire, Kamath and Stothard questionnaire, 6-item carpal tunnel syndrome symptoms scale (CTS-6), Boston Carpal Tunnel Questionnaire, Wainner clinical prediction rule, and Lo clinical prediction rule. Positive likelihood ratios ranged from 0.94 for the Boston Carpal Tunnel Questionnaire to 10.5 for the CTS-6, and negative likelihood ratios ranged from 1.04 to 0.05 for the same diagnostic tools, respectively. Nine studies reported the diagnostic accuracy of the Katz and Stirrat hand symptom diagram. Positive and negative likelihood ratios ranged from 1.42 to 8 and from 0.78 to 0.05, respectively. Only 4 studies had high methodologic quality.

CONCLUSION: Limited evidence supports high accuracy of the CTS-6, Kamath and Stothard questionnaire, and Katz and Stirrat hand symptom diagram. Other scales have lesser and more conflicting evidence. Further high-quality studies are necessary to examine the diagnostic accuracy of these tests to assist ruling in or ruling out CTS. *J Orthop Sports Phys Ther 2020;50(11):622-631. Epub 16 Sep 2020. doi:10.2519/jospt.2020.9599*

KEY WORDS: *carpal tunnel syndrome, diagnostic accuracy, diagnostic scales and questionnaires, hand symptom diagram/map, systematic review*

Carpal tunnel syndrome (CTS) is caused by compression of the median nerve in the carpal canal and is the most prevalent type of compression neuropathy of the upper extremity.[3] An important barrier to treating CTS is the lack of a diagnostic gold standard.[1] In clinical decision making, an ongoing process of gathering enough information to decide on the optimal plan of care,[21] diagnosis is a central feature. Clinical examination tests are quick, inexpensive, and give an immediate answer, which makes them appealing for diagnosing CTS.

Carpal tunnel syndrome can be diagnosed with a variety of clinical examination tests and by the patient's history.[1] However, the final confirmation is often made based on neurophysiological tests assessing median nerve conduction velocity.[13] The most recent CTS management guideline of the American Academy of Orthopaedic Surgeons (AAOS) concludes that only limited evidence supports the use of handheld nerve conduction studies (NCS), ultrasound, and magnetic resonance imaging in CTS diagnosis.[1] Advanced diagnostic testing can be expensive and painful in some cases (eg, NCS). Electrodiagnostic studies report higher false-positive and false-negative results compared to studies of clinical examination tests.[1]

According to a previous systematic review[29] and the AAOS guideline,[1] clinical examination tests for diagnosing CTS can be categorized into 4 major groups: (1) provocative maneuvers (eg, Phalen test, Tinel sign), (2) sensory and motor

[1]School of Physical Therapy, Faculty of Health Sciences, Western University, London, Canada. [2]School of Rehabilitation Science, McMaster University, Hamilton, Canada. [3]Roth McFarlane Hand and Upper Limb Centre, St Joseph's Hospital, London, Canada. [4]Sengkang General Hospital, Singapore. This review was registered with PROSPERO (CRD42018109031). Dr MacDermid was supported by a Canadian Institutes of Health Research Chair in Gender, Work and Health and by the James Roth Research Chair in Musculoskeletal Measurement and Knowledge Translation. 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 Armaghan Dabbagh, School of Physical Therapy, Faculty of Health Sciences, Elborn College, Room 1588, Western University, London, ON N6A 1H1 Canada. E-mail: adabbag@uwo.ca

tests (eg, 2-point discrimination, thenar weakness test), (3) diagnostic scales (eg, the 6-item carpal tunnel syndrome symptoms scale [CTS-6]) and questionnaires (eg, the Kamath and Stothard questionnaire), and (4) hand symptom diagrams/maps (the Katz and Stirrat hand symptom diagram). In this systematic review, we focused on diagnostic scales, questionnaires, and symptom diagrams/maps. Systematic reviews addressing the 2 other categories (provocative and sensory/motor tests) will be presented in 2 separate systematic reviews.[11]

The last systematic review on diagnostic test accuracy for CTS[29] is outdated. The purpose of our systematic review was to appraise and synthesize the evidence on the diagnostic accuracy of diagnostic scales, questionnaires, and hand symptom diagrams for diagnosing CTS.

METHODS

We registered this study with the International Prospective Register of Systematic Reviews (PROSPERO) on December 20, 2018 (CRD42018109031). The Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) and Cochrane Collaboration guidelines were followed.[12,27]

Information Sources

We conducted a literature search in 3 electronic databases—MEDLINE (through Ovid, from 1946), Embase, and CINAHL—from their inception to January 20, 2020. The search strategy was designed to identify studies of diagnostic test accuracy of at least 1 clinical diagnostic test for the diagnosis of CTS. We reported the results for diagnostic scales, questionnaires, and hand symptom diagrams. We developed the search strategy (**APPENDIX A**, available at www.jospt.org) in 2 consecutive meetings with a librarian who specializes in health science research methodology at McMaster University, by combining vocabulary and key words related to the diagnostic accuracy of the clinical examination tests for the diagnosis of CTS.

To identify the names of the clinical diagnostic tests for CTS to be included in the search strategy, we searched previous reviews on this topic and the AAOS guidelines. The terms used in the search were also reviewed by a physical therapist and a hand therapist (A.D. and J.M.) to ensure that all known physical examination tests for CTS were included. We hand searched the reference lists of included articles.

Study Selection

Two reviewers (A.D. and J.Y.) performed the study selection independently in 2 phases. In the first phase, titles and abstracts were reviewed against predetermined inclusion and exclusion criteria. The agreement of the reviewers in this phase was calculated using kappa statistics.[25] Kappa values less than 0.20 indicated poor agreement, and values greater than 0.80 indicated almost perfect agreement in rating.[25] All statistical analyses were conducted using Stata 15 (StataCorp LLC, College Station, TX). In the second phase, full-text articles were retrieved and reviewed. To resolve any disagreement during the first or second phase of the study selection process, a third reviewer (J.M.) moderated a consensus through discussion.

Eligibility Criteria

There were no restrictions on study selection based on sample size, language, or sex. Studies were included in this systematic review when the below criteria were met.

Design Systematic reviews and case-control, cross-sectional, or cohort studies that collected either prospective or retrospective data in a full-report format were included.

Participants Studies that included adults (18 years old or older) diagnosed with or suspected to have CTS and that had a control group of participants with any diagnosis of neurological, musculoskeletal, or vascular conditions of the upper extremity (eg, cervical radiculopathy or tennis elbow) were included. Studies with healthy control groups were excluded.

Diagnostic Test Studies that assessed at least 1 diagnostic accuracy property of the physical examination tests for the diagnosis of CTS (restricted to diagnostic scales, questionnaires, and hand symptom diagrams/maps) were included.

Comparison Because there is no known gold standard for the diagnosis of CTS, we accepted any physical examination test used as a reference standard (eg, NCS, surgical decompression of the carpal canal, other clinical examination diagnostic tests, or a combination of reference standard tests) made by a physician or expert clinician.

Outcome Articles reporting diagnostic accuracy properties, such as sensitivity and specificity or likelihood ratio (LR), or that provided enough data to (re)construct 2-by-2 contingency tables, were included.

Time All time frames reporting the diagnostic accuracy of clinical examination for the diagnosis of CTS were accepted.

We excluded (1) reviews, letters, conference abstracts, editorials, and case reports; (2) studies not using diagnostic scales, questionnaires, or hand symptom diagrams as an index test for the diagnosis of CTS; (3) studies on median nerve conditions other than CTS; and (4) studies not reporting sensitivity, specificity, or other diagnostic accuracy properties or not providing sufficient data to calculate the statistics.

Data Extraction

Two authors (A.D. and J.Y.) independently extracted information from 3 included articles, and the agreement was discussed with a third author (J.M.). Because the agreement was very high, the first author completed data extraction alone, using a predetermined, self-developed data-extraction form. In the case of any uncertainty in data extraction, the other 2 reviewers were contacted and a consensus was acquired through discussion. We extracted author identification, publication year, country of study, study design, participant characteristics (age, sex, CTS severity and duration), sample size, inclusion and exclusion criteria, the participant selection process, clinical examination test, reference standard, and all of the available information regarding diagnostic accuracy measures. In the case of any values missing from the articles, the study authors were contacted by e-mail.

Data Synthesis and Analysis

Where possible, we extracted sensitivities, specificities, positive and negative predictive values, as well as positive likelihood ratios (+LRs) and negative likelihood ratios (–LRs). When data were not provided, we tried to calculate values using the information reported about true positives, false positives, false negatives, and true negatives.[19] We then created 2-by-2 contingency tables and calculated the sensitivity, specificity, +LR, and –LR, including the 95% confidence interval for each physical examination test if possible.[19] The sensitivity of a diagnostic test is the ability of the test to truly label people (ie, true positive) with a given medical condition, and specificity of a diagnostic test is defined as the identification of those without the disease or disorder (true negative).[19]

Likelihood ratios are diagnostic accuracy properties that are independent of the prevalence of the disease.[33] We calculated +LR [sensitivity/(1 – specificity)] and –LR [(1 – sensitivity)/specificity].[33] Positive likelihood ratio values of greater than 10 and –LRs less than 0.1 comprise one of the most useful measures in diagnostic decision making.[33] Values between 5 and 10 (+LR) and between 0.1 and 0.2 (–LR) suggest that the test has a moderate ability to change the probability of having a condition.[1] Last, +LR and –LR values of less than 5 or more than 0.5, respectively, suggest that the test has a small ability to change the probability of a diagnosis.[1] Data heterogeneity (eg, different index tests, different sample characteristics, different cutoff points for positive test results) precluded meta-analysis.

Assessment of Risk of Bias and Applicability Concerns

Two authors (A.D. and J.Y.) independently rated the risk of bias and applicability concerns using the revised Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2[37]) tool. To resolve discrepancies, we reached a consensus through discussion with a third author (J.M.). The QUADAS-2 tool rates the risk of bias of articles in 4 domains: participant selection, index test, reference standard, and flow and timing.[37] The applicability concerns regarding articles are rated for all of the domains in the QUADAS-2 tool except for the flow and timing of the participants.[37] Each domain has a set of signaling questions that can be answered as "yes," "no," or "unclear."[37]

If the answers to all of the signaling questions were yes, then that domain was considered to have a low risk of bias or applicability concerns. If the answer to any of the signaling questions of a domain was no or unclear, then the risk of bias or applicability concerns of that domain were rated as high or unclear. To generate an overall rating of the risk of bias or applicability concerns of an article, studies rated as low on all of the domains were defined as "low risk of overall bias" or "low applicability concerns." Ratings of high or unclear on any of the domains resulted in the overall judgment of the articles as "at risk of bias" or "concerns regarding applicability."

RESULTS

After removing duplicates and evaluating 4052 records, 161 references were assessed in full-text review. Twenty-one articles met the inclusion criteria (**FIGURE 1**): 9 studies assessed the diagnostic accuracy of diagnostic hand symptom diagrams/maps[2,7,10,15,22-24,32,34] and 12 articles reported on the diagnostic accuracy of diagnostic scales and questionnaires.[4-6,8,9,14,18,20,28,30,31,36] The studies were conducted in 6 countries: Austria, Canada, Greece, Spain, the United Kingdom, and the United States. Conflicts of interest of the included studies are available in **APPENDIX B** (available at www.jospt.org). The kappa value of the agreement for title and abstract screening was 0.70 (95% confidence interval: 0.66, 0.74; SE, 0.02). The methodological assessments of all included articles are presented in **FIGURE 2**. Overall, 4 studies had a low risk of bias[18,22,24,28] and 4 had an unclear rating only in 1 domain.[6,7,14,20] Regarding applicability, 12 studies had no concerns (**FIGURE 3**).[5-8,14,18,20,22,24,28,30,31]

The detailed characteristics of the participants in the included studies, as well as the clinical diagnostic tests and the reference standards used in each study, are presented in **APPENDICES C** and **D** (available at www.jospt.org). Eighteen studies had a prospective cross-sectional study design, and the remaining 3 articles[6,14,23] had retrospective study designs. All but 3 studies[7,10,15] recruited participants from persons with suspected CTS referred to orthopaedic, rheumatology, or hand clinics (or similar clinical settings) or nerve conduction labs. Only 3 studies reported the pretest probability of having CTS in their study sample.[14,15,34] High variability among studies (eg, a variety of index tests, criteria for positive test results, and population characteristics) precluded meta-analysis.

Diagnostic Accuracy of the Diagnostic Scales and Questionnaires for CTS Diagnosis

Six different diagnostic tools were used and assessed across all included studies. Of the 12 studies on the diagnostic accuracy of scales and questionnaires for the diagnosis of CTS, the following diagnostic tools were assessed: (1) the Bland questionnaire,[4,5,18] (2) the Kamath and Stothard questionnaire,[6,9,20,36] (3) the CTS-6,[14,30,36] (4) the Boston Carpal Tunnel Questionnaire,[8,31] (5) the Lo clinical prediction rule,[28,36] and (6) the Wainner clinical prediction rule.[36] Thorough descriptions of the CTS diagnostic scales, questionnaires, and hand symptom diagrams, as well as their methods of administration and positive result thresholds, are presented in **TABLE 1**. The overall sample size of these studies was 17 768 wrists with suspected CTS; 7488 wrists were diagnosed with true-positive CTS (positive results confirmed by both the index and the reference standard tests). Positive likelihood ratios to diagnose or rule in CTS ranged from 0.94 for the Boston Carpal Tunnel Questionnaire[31] to 10.5 for the CTS-6,[14] and –LRs to exclude or rule out CTS ranged from 1.04 for the Boston Carpal Tunnel Questionnaire[31] to 0.05 for the CTS-6[14] (**TABLE 2**). Only 1 study combined tests, which resulted in high sensitivity (95.9%) and moderate specificity (50%).[6]

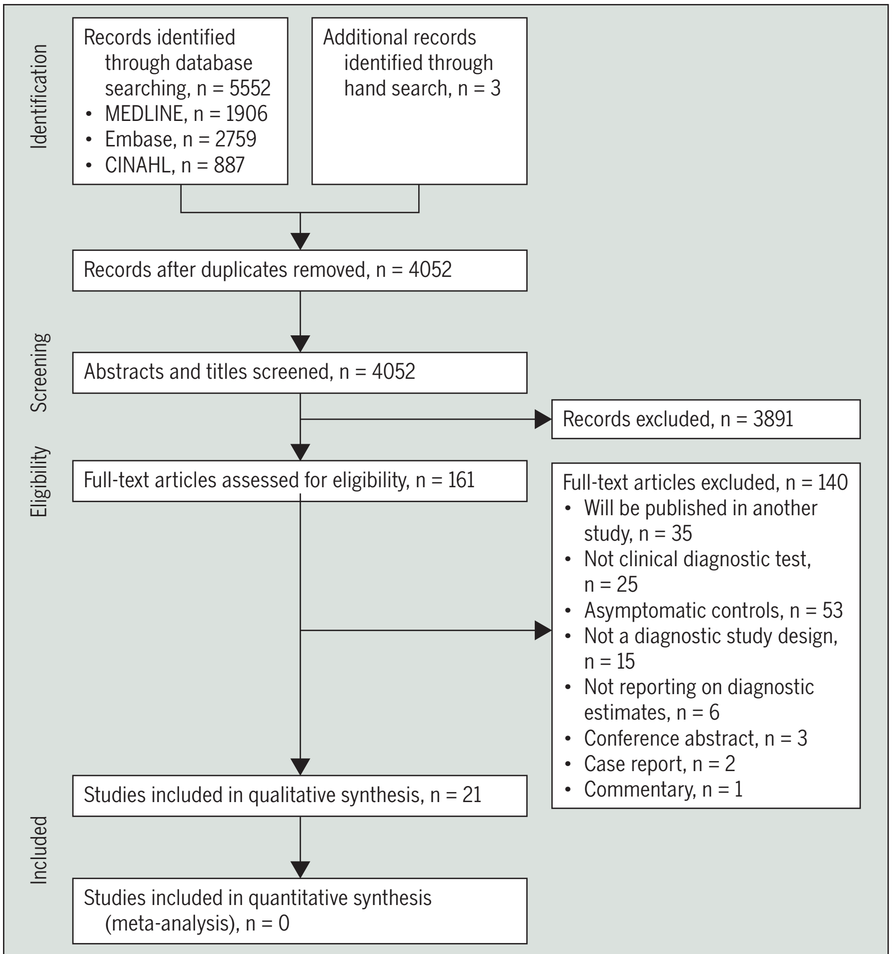


FIGURE 1. PRISMA flow diagram.

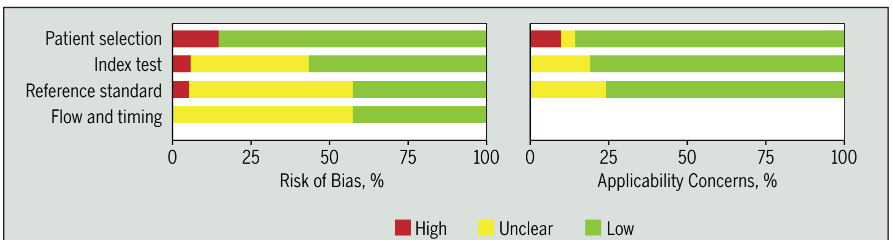


FIGURE 2. The proportion of included studies with low, high, or unclear risk of bias and concerns regarding applicability, using the revised Quality Assessment of Diagnostic Accuracy Studies tool.

Diagnostic Accuracy of Hand Diagrams/Maps for CTS Diagnosis

Nine studies evaluated the diagnostic accuracy of the Katz and Stirrat[23] hand symptom diagram.[2,7,10,15,22-24,32,34] The sample size was 1796 wrists with suspected CTS and 930 true-positive CTS wrists. Positive likelihood ratios to diagnose or rule in CTS ranged from 1.42[7] to 8,[23] and –LRs to exclude or rule out CTS ranged from 0.78[15] to 0.05[24] (**TABLE 3**).

Reference Standards for CTS Diagnosis

Seventeen studies[2,4-10,15,18,22,24,28,30-32,34] used electrodiagnosis as the reference standard, although the methodology and criteria for positive test results varied between the studies. Two studies performed clinical diagnosis[23,36] and 1 study[20] used carpal tunnel release surgery as the reference standard. One study did not have a reference standard and compared the results of the CTS-6 to NCS and diagnostic ultrasound findings, using a statistical method called latent class analysis.[14]

DISCUSSION

The purpose of our systematic review was to summarize and assess the quality of the diagnostic scales, questionnaires, and hand symptom diagrams/maps for the diagnosis of CTS. We found 12 clinical studies reporting on 6 different diagnostic scales and questionnaires, and 9 studies reporting on the Katz and Stirrat[23] hand symptom diagram. Among these tests, the CTS-6, the Kamath and Stothard[20] questionnaire, and the Katz and Stirrat[23] hand symptom diagram (when used with the classical categorization) had the greatest accuracy in deciding whether to rule in or rule out CTS.

Accurate diagnosis is the key to establishing appropriate treatment plans and prognosis. Given the high prevalence of CTS, the clinical diagnosis tends to be an important concern for clinicians. No other condition seen by hand therapists seems to have this variety of available clinical diagnostic tests.[29] The consider-

	Risk of Bias				Applicability Concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Ammer et al[2]	−	?	?	?	−	+	?
Bland[4]	+	?	?	+	+	+	?
Bland et al[6]	+	+	+	?	+	+	+
Bland et al[5]	+	?	?	+	+	+	+
Bonauto et al[7]	+	+	+	?	+	+	+
Bougea et al[8]	+	?	?	+	+	+	+
Bridges et al[9]	+	−	−	+	+	?	?
Calfee et al[10]	+	?	?	?	+	?	+
Fowler et al[14]	+	+	+	?	+	+	+
Franzblau et al[15]	+	+	?	?	+	+	?
Hems et al[18]	+	+	+	+	+	+	+
Kamath and Stothard[20]	+	+	+	?	+	+	+
Katz and Stirrat[23]	−	?	+	+	?	?	+
Katz et al[24]	+	+	+	+	+	+	+
Katz et al[22]	+	+	+	+	+	+	+
Lo et al[28]	+	+	+	+	+	+	+
Makanji et al[30]	+	?	?	?	+	+	+
Naranjo et al[31]	+	+	?	?	+	+	+
O'Gradaigh and Merry[32]	+	+	?	?	+	+	?
Szabo et al[34]	−	+	?	?	−	+	+
Wang et al[36]	+	?	?	?	+	?	+

− High ? Unclear + Low

FIGURE 3. Risk of bias and applicability concerns of the included studies, using the revised Quality Assessment of Diagnostic Accuracy Studies tool.

able number of available studies on the clinical diagnosis of CTS also reflects the prominent emphasis on diagnosing CTS.

The lack of a gold standard for CTS diagnosis contributes to heterogeneity of diagnostic methods. The diagnostic approaches used for people with suspected CTS vary in different settings, determined by the clinical background of the treating clinician.[16] A classic approach is to first gather information from patient history and physical examination tests and create a list of possible diagnoses, then to determine further ancillary testing to confirm one of these diagnoses.[16] In some clinical settings, electrodiagnostic tests are almost always performed for CTS diagnosis. In other settings, these tests are rarely administered.[16]

Only 3 studies reported the prevalence of CTS in their sample.[14,15,34] Settings with a higher prevalence of CTS (eg, hand clinics or electrodiagnosis labs) have higher pretest probability of CTS. It is important to consider the setting in which the study is being conducted. Although the results from the studies done in a clinical setting tend to be closer to what a clinician might encounter, higher pretest probability of CTS in these settings leads to inflated estimates of the diagnostic accuracy properties of the tools.[16,36] Only 3 studies recruited their sample from a nonclinical population, where the probability of having CTS was still high, because the inclusion criteria consisted of workers with current hand symptoms.[7,10,15] To eliminate the effect of pretest probability of CTS in the sample population on diagnostic accuracy measures, we extracted (or calculated) +LRs and –LRs, as these values are independent of the prevalence of the condition.

Clinicians are often presented with clients who have similar upper extremity signs and symptoms but different diagnoses. Therefore, a healthy control group does not reflect the clinical setting and might decrease the applicability of this systematic review. A case-control design, where the controls are healthy individuals, risks erroneous estimates of (inflated) specificity and negative predictive values.[19] To avoid bias, we excluded studies with solely asymptomatic (ie, healthy) control groups.

Due to the lack of a diagnostic gold standard for CTS,[1] different reference standards were used in the included studies. We included studies regardless of their choice of reference standard.[11] The most common reference standard test used in the included studies was electrodiagnosis; however, this comparison is flawed because electrodiagnosis has false-positive and false-negative results.[1] Only 1 study used latent class analysis,[14] which is a statistical technique that can be used when there is no established gold standard.

Classifying CTS Diagnostic Tools

The available clinical examination tests for diagnosing CTS can be categorized into 4 main groups, each test having limited capability of being used alone as

TABLE 1 — Description of Scales, Questionnaires, and Hand Symptom Diagrams for CTS

Diagnostic Test	Method	Positive Result Threshold
6-item CTS symptoms scale[17]	Six criteria are assessed and scored: 1. Numbness in the median nerve distribution (3.5 points) 2. Nocturnal numbness (4 points) 3. Thenar musculature weakness/atrophy (5 points) 4. Positive Tinel sign (4 points) 5. Positive Phalen test (5 points) 6. Loss of 2-point discrimination (4.5 points)	A score of 12 points (46%)[9,14,20,30,36] A score of 18 points[36]
BCTQ[26]	It comprises 2 subscales: the symptom severity scale (11 questions) and the functional status scale (8 questions of hand function during daily activities)	Scores of 1.95 or greater[8] Scores of 3 or greater[31]
Bland questionnaire[4]	It has 2 sections: 1. Background information, including age, occupation, hand dominance, and diabetes, is recorded. There is an open question regarding the type of symptoms experienced by the patient 2. Questions 6-12 cover details of symptoms, including the location of paresthesia in the hand, nocturnal pain, relief of paresthesia by shaking the hand, relief by the use of a wrist splint, impairment of manual dexterity, and duration of symptoms	A score of 7 or greater[18] A cutoff probability of 0.5[4] A score of 40% or greater[5]
Lo clinical prediction rule[28]	Nine clinical variables are assessed and scored: 1. Sex 2. Duration of symptoms 3. Presence of wrist pain (negative predictor) 4. Presence of neck pain (negative predictor) 5. Nocturnal symptoms 6. Presence of thenar atrophy 7. Abductor pollicis brevis weakness 8. Median sensory symptoms 9. Pinprick sensation examination	A score of 10 or greater[36]
Wainner clinical prediction rule[35]	Five items are assessed and scored: 1. Shaking hand for symptom relief 2. Wrist ratio index 3. BCTQ symptom severity scale 4. Reduced median sensory field of digit 1 5. Age greater than 45 y	A score of 3 or greater[36]
Katz and Stirrat hand symptom diagram[23]	A self-administered hand symptom diagram that depicts both hands with dorsal and palmar views. Patients are asked to mark areas on the diagram corresponding to the location of their symptoms and to indicate the quality of their discomfort	Classic CTS: tingling, numbness, or decreased sensation with or without pain in at least 2 digits (1, 2, or 3); symptoms in the dorsa and palms of the hands excluded; fifth-finger symptoms, wrist pain, or radiation proximal to the wrist allowed Probable CTS: same as classic, except palmar symptoms are allowed unless confined solely to the ulnar aspect Possible CTS: tingling, numbness, and/or decreased sensation in at least 1 digit (1, 2, or 3) Unlikely CTS: no symptoms in digits 1, 2, or 3[23]
Kamath and Stothard questionnaire[20]	It has 9 questions about signs and symptoms of CTS, with "yes," "no," and "not applicable" response options	Scores greater than 6 and below 3[9] Score greater than 5[9,20,36]

Abbreviations: BCTQ, Boston Carpal Tunnel Questionnaire; CTS, carpal tunnel syndrome.

the diagnostic criterion to rule in or rule out CTS. In practice, diagnosis is often a triangulation of a representative test from several of the 4 main categories, that is, provocative tests, sensorimotor tests, and the self-reported questionnaires of symptoms or clinician-based evaluations. Following is a discussion of the available scales, questionnaires, and hand symptoms/maps.

Scales and Questionnaires

The 2 tests most frequently studied were the CTS-6 and the Kamath and Stothard[20] diagnostic tests. The CTS-6 test comprises 6 criteria, as ranked by a Delphi consensus of expert clinicians.[17] Based on the results of 1 study with unclear risk of bias[14] (in only 1 domain), the CTS-6 has a strong ability to change the pretest probability of having CTS (+LR = 10.50 and –LR = 0.05). These findings are opposed by the findings of 2 other papers, which indicate that the CTS-6 has a small ability to change the pretest probability of having CTS.[30,36] We believe that more studies are needed to confirm the diagnostic accuracy of the CTS-6 test.

TABLE 2 DIAGNOSTIC ACCURACY OF SCALES AND QUESTIONNAIRES FOR CTS DIAGNOSIS

Study/Examination Tool	Sensitivity, %[a]	Specificity, %[a]	PPV, %[a]	NPV, %[a]	+LR	–LR
Bland[4]						
Bland questionnaire	79.1	55.6	69[b]	67[b]	2.66[b]	0.56[b]
Bland et al[6]						
Combined KSQ and CTS-7[c]	95.9	50	NR	NR	1.92[b]	0.08[b]
Bland et al[5]						
Bland web-based questionnaire	78	68	NR	NR	2.43[b]	0.32[b]
Bougea et al[8]						
Greek version of the BCTQ	75.5	68.3	NR	NR	2.38[b]	0.35[b]
Bridges et al[9]						
KSQ	87 (80, 94)	87 (80, 93)	NR	NR	6.70[b]	0.15[b]
Fowler et al[14]						
CTS-6	95 (86, 99)	91 (74, 99)	NR	NR	10.50	0.05
Hems et al[18]						
Bland questionnaire	82 (72, 90)	67 (41, 87)	91	48	2.48	0.26
Kamath and Stothard[20]						
KSQ	85	NR	90	NR	NR	NR
Lo et al[28]						
Lo clinical prediction rule	76	68	NR	NR	2.37[b]	0.35[b]
Makanji et al[30]						
CTS-6	87	60	89	55	2.17[b]	0.27[b]
Naranjo et al[31]						
BCTQ functional status scale	35.1	62.5	NR	NR	0.94	1.04
BCTQ symptom severity scale	48.6	60	NR	NR	1.22	0.86
Wang et al[36]						
CTS-6	56 (50, 62)	71 (62, 79)	83 (76, 88)	40 (33, 47)	1.93[b]	0.62[b]
KSQ	74 (68, 79)	64 (54, 72)	83 (78, 87)	50 (42, 58)	2.05[b]	0.41[b]
Lo clinical prediction rule	66 (59, 71)	56 (47, 65)	78 (72, 83)	40 (33, 48)	1.50[b]	0.78[b]
Wainner clinical prediction rule	70 (64, 75)	64 (54, 72)	82 (77, 87)	47 (39, 55)	1.94[b]	0.47[b]

Abbreviations: BCTQ, Boston Carpal Tunnel Questionnaire; CTS, carpal tunnel syndrome; CTS-6, 6-item carpal tunnel syndrome symptoms scale; CTS-7, 7-item carpal tunnel syndrome symptoms scale; KSQ, Kamath and Stothard questionnaire; –LR, negative likelihood ratio; +LR, positive likelihood ratio; NPV, negative predictive value; NR, not reported; PPV, positive predictive value.
[a]*Values in parentheses are confidence interval.*
[b]*Values calculated by the authors of this study.*
[c]*Only questionnaire segments were included for the analysis in the original study. The authors called this test the CTS-7; however, the rationale behind this naming is unclear, and they referenced the CTS-6 paper.*[17]

Kamath and Stothard[20] stated that they developed a questionnaire based on the previous work of Levine et al[26]; however, a clear description of this process is lacking.[20] Despite this shortcoming in validation, 1 low-quality paper[9] and 2 papers[6,20] with unclear ratings in only 1 domain have assessed the diagnostic accuracy properties of this tool. Originally, Kamath and Stothard[20] only included persons with a definite diagnosis of CTS (determined by carpal tunnel release surgery). Therefore, specificity, negative predictive values, and +LRs and –LRs could not be calculated.[20] Good diagnostic accuracy values have been reported in 2 studies[9,20] (sensitivity, 85%-87%; specificity, 87%). The +LR of 6.70 and –LR of 0.15 of the Kamath and Stothard[20] questionnaire from 1 study with high risk of bias[9] indicate that the Kamath and Stothard[20] questionnaire has moderate to good utility to change the pretest probability of having CTS. Further validation of the Kamath and Stothard[20] questionnaire might improve the diagnostic ability of this test.

The Boston Carpal Tunnel Questionnaire, a tool most frequently used as an outcome measure for CTS treatment, was assessed in 2 studies.[8,31] According to the LRs of these 2 studies, the Boston Carpal Tunnel Questionnaire had a small value in deciding whether to rule in or rule out CTS in the clinical setting. Clinicians should also be aware that the Boston Carpal Tunnel Questionnaire has different names across studies (eg, Levine's questionnaire[20]); however, all of these names refer to the same diagnostic test.[26]

The Bland questionnaire was evaluated in 3 studies, 2 with large samples[4,5] and 1 high-quality paper.[18] Compared to NCS as the reference standard, the Bland questionnaire had moderate sensitivity and strong positive predictive values (82% and 91%, respectively)[18]; therefore, we suggest that it might be a good tool to rule in CTS, but that it is not very specific in ruling it out. The Bland questionnaire had +LRs ranging from 2.43 to 2.66 and –LRs ranging from 0.56 to 0.26. These findings indicate that the Bland questionnaire has a small ability to predict the pretest probability of having CTS.

Two clinical prediction rules (the Lo and Wainner rules), reported in the included studies by Lo et al[28] and Wang et al,[36] were used in the included articles. The reference standard of Wang et al[36] was not well defined, and the extracted information is at unclear risk of bias.[36] In addition, the Lo clinical prediction rule was assessed in a high-quality study, showing that only moderate sensitivity (76%) and specificity (68%) are predictable when electrodiagnostic studies serve as the reference standard.[28] According to the results of this systematic review, we do not recommend clinicians to use the Lo clinical prediction rule or the Wainner clinical prediction rule until future high-quality studies with larger sample sizes establish their diagnostic accuracy.

Hand Symptom Diagrams/Maps

Three different criteria for positive test interpretation were identified in the included studies. Six studies categorized people with suspected CTS into 4 groups of classic, probable, possible, and unlikely (**TABLE 1**), which is the original categorization method suggested by Katz and Stirrat in 1990.[23] This categorization resulted in the highest +LR and lowest –LR among all of the included studies.[23,24] Two studies interpreted people in classic or probable categories as having a positive, and those in possible or unlikely categories as having a negative, CTS diagnosis.[10,32] One study interpreted those with classic, probable, and possible results according to the diagram as having CTS, and those in the unlikely category as not having CTS; this categorization led to the small ability of this test to indicate CTS diagnosis.[7] Our results suggest that the Katz and Stirrat[23] hand symptom diagram is a valuable clinical tool for diagnosing CTS. We recommend clinicians to use the classic categorization method (classic, probable, possible, and unlikely CTS) for the most accurate results.

Clinical Implications

The CTS-6 and the Kamath and Stothard[20] questionnaire had the highest accuracy in diagnosing CTS. The Boston Carpal Tunnel Questionnaire and the Bland questionnaire had a small ability to change the pretest probability of having CTS. The evidence on diagnostic accuracy of the Lo clinical prediction rule and the Wainner clinical prediction rule was inadequate and inconclusive. The Katz and Stirrat[23] hand symptom diagram had the highest accuracy in diagnosing CTS when interpreted by the classical method of categorization: classic, probable, possible, and unlikely CTS.

Limitations

Studies had different interpretation criteria, samples, and reference standards, which made comparisons difficult and resulted in a heterogeneity of data that precluded meta-analysis. Three of the included diagnostic tests (the Boston Carpal Tunnel Questionnaire and the Lo and Wainner clinical prediction rules) were only examined in 1 or 2 studies, which limited our conclusions regarding the diagnostic accuracy of these tests. We might have missed studies, due to variations in the terminology of diagnostic tests. Although our search strategy was developed in consultation with a professional health sciences librarian, we cannot be certain that all of the eligible studies were included. There is a risk for potential publication bias, because we only included published literature.

Although some of the diagnostic scales and questionnaires seem promising, they are not supported by high-quality evidence. We recommend that future studies strictly adhere to established guidelines to produce results with low risk of bias and of high quality. There is also a need to compare the clinical triangulation process to electrodiagnosis or NCS.

TABLE 3 Diagnostic Accuracy of the Hand Symptom Diagrams/Maps for CTS Diagnosis

Study[a]	Sensitivity, %[b]	Specificity, %[b]	PPV, %[b]	NPV, %[b]	+LR	–LR
Ammer et al[2]	92.6	50	62.5	88.2	1.85[c]	0.15[c]
Bonauto et al[7]	61	58	52	67	1.42[c]	0.67[c]
Calfee et al[10]	38 (28, 50)	81 (73, 87)	54 (41, 67)	69 (61, 76)	1.63[c]	0.76[c]
Franzblau et al[15]	34	84	27	88	2.12[c]	0.78[c]
Katz and Stirrat[23]	80	90	NR	NR	8[c]	0.22[c]
Katz et al[24]	96	73	58	91	3.55[c]	0.05[c]
Katz et al[22]	61	71	59 (48, 68)	73 (66, 80)	2.10[c]	0.54[c]
O'Gradaigh and Merry[32]	92	40	92	14	1.53[c]	0.2[c]
Szabo et al[34]	76 (62, 89)	76 (52, 77)	36	95	3.17[c]	0.32[c]

Abbreviations: CTS, carpal tunnel syndrome; –LR, negative likelihood ratio; +LR, positive likelihood ratio; NPV, negative predictive value; NR, not reported; PPV, positive predictive value.
[a]All studies used the Katz and Stirrat[23] hand symptom diagram.
[b]Values in parentheses are confidence interval.
[c]Values calculated by the authors of this study.

CONCLUSION

The CTS-6 and the Kamath and Stothard[20] questionnaire had promising diagnostic accuracies for diagnosing CTS. More high-quality papers are necessary to confirm these findings. The Katz and Stirrat[23] hand symptom diagram yielded the most accurate results in predicting the pretest probability of having CTS, when used with the classic categorization method (classic, probable, possible, and unlikely CTS). More invasive diagnostic tools for CTS (ie, NCS) might only be necessary when there is concern regarding the certainty of clinical diagnoses. ◉

KEY POINTS

FINDINGS: The 6-item carpal tunnel syndrome symptoms scale (CTS-6), Kamath and Stothard questionnaire, and Katz and Stirrat hand symptom diagram were most able to change the pretest probability of having carpal tunnel syndrome (CTS). We recommend a multifaceted strategy that combines several diagnostic tests (eg, the CTS-6 and the Katz and Stirrat hand symptom diagram) to confirm CTS diagnosis.

IMPLICATIONS: Although the evidence for the diagnostic accuracy of some of the scales and questionnaires is still inconclusive, we recommend that costly and invasive tests for CTS may only be needed when diagnostic scales, questionnaires, and hand symptom diagrams, as the first line of diagnosis, lack conclusiveness.

CAUTION: Our conclusions are mainly based on studies with moderate to high risk of bias and moderate concerns regarding applicability.

STUDY DETAILS

AUTHOR CONTRIBUTIONS: All authors made a substantial contribution to (1) the conception and design of the review, development of the search strategy, and analysis and interpretation of the data; and (2) drafting the article or revising it critically for important intellectual content. All authors assume public responsibility for the work.

DATA SHARING: All data relevant to the study are included in the article or within the online appendices.

PATIENT AND PUBLIC INVOLVEMENT: There was no patient or public involvement in this research.

REFERENCES

1. American Academy of Orthopaedic Surgeons. Management of Carpal Tunnel Syndrome: Evidence-Based Clinical Practice Guideline. Rosemont, IL: American Academy of Orthopaedic Surgeons; 2016.
2. Ammer K, Mayr H, Thür H. Self-administered diagram for diagnosing carpal tunnel syndrome. *Eur J Phys Med Rehabil*. 1993;3:42-45.
3. Atroshi I, Gummesson C, Johnsson R, Ornstein E, Ranstam J, Rosén I. Prevalence of carpal tunnel syndrome in a general population. *JAMA*. 1999;282:153-158. https://doi.org/10.1001/jama.282.2.153
4. Bland JDP. The value of the history in the diagnosis of carpal tunnel syndrome. *J Hand Surg Br*. 2000;25:445-450. https://doi.org/10.1016/S0266-7681(00)80009-1
5. Bland JDP, Rudolfer S, Weller P. Prospective analysis of the accuracy of diagnosis of carpal tunnel syndrome using a web-based questionnaire. *BMJ Open*. 2014;4:e005141. https://doi.org/10.1136/bmjopen-2014-005141
6. Bland JDP, Weller P, Rudolfer S. Questionnaire tools for the diagnosis of carpal tunnel syndrome from the patient history. *Muscle Nerve*. 2011;44:757-762. https://doi.org/10.1002/mus.22158
7. Bonauto DK, Silverstein BA, Fan ZJ, Smith CK, Wilcox DN. Evaluation of a symptom diagram for identifying carpal tunnel syndrome. *Occup Med (Lond)*. 2008;58:561-566. https://doi.org/10.1093/occmed/kqn123
8. Bougea A, Zambelis T, Voskou P, et al. Reliability and validation of the Greek version of the Boston Carpal Tunnel Questionnaire. *Hand (N Y)*. 2018;13:593-599. https://doi.org/10.1177/1558944717725379
9. Bridges MJ, Robertson DC, Chuck AJ. Predicting the result of nerve conduction tests in carpal tunnel syndrome using a questionnaire. *Hand Surg*. 2011;16:39-42. https://doi.org/10.1142/S0218810411005047
10. Calfee RP, Dale AM, Ryan D, Descatha A, Franzblau A, Evanoff B. Performance of simplified scoring systems for hand diagrams in carpal tunnel syndrome screening. *J Hand Surg Am*. 2012;37:10-17. https://doi.org/10.1016/j.jhsa.2011.08.016
11. Dabbagh A, MacDermid J, Yong J, Macedo L, Packham T. The accuracy of clinical examination tests for the diagnosis of carpal tunnel syndrome compared to the reference standard in adults diagnosed with carpal tunnel syndrome: a systematic review of diagnostic test accuracy [protocol]. Available at: https://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42018109031&ID=CRD42018109031. Accessed December 31, 2018.
12. Deeks JJ, Wisniewski S, Davenport C. Guide to the contents of a Cochrane Diagnostic Test Accuracy protocol. In: Deeks JJ, Bossuyt PM, Gatsonis C, eds. *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy*. London, UK: The Cochrane Collaboration; 2013:ch 4.
13. de Jesus Filho AG, do Nascimento BF, de Carvalho Amorim M, Naus RA, de Araújo Loures E, Moratelli L. Comparative study between physical examination, electroneuromyography and ultrasonography in diagnosing carpal tunnel syndrome. *Rev Bras Ortop*. 2014;49:446-451. https://doi.org/10.1016/j.rboe.2014.09.002
14. Fowler JR, Cipolli W, Hanson T. A comparison of three diagnostic tests for carpal tunnel syndrome using latent class analysis. *J Bone Joint Surg Am*. 2015;97:1958-1961. https://doi.org/10.2106/JBJS.O.00476
15. Franzblau A, Werner RA, Albers JW, Grant CL, Olinski D, Johnston E. Workplace surveillance for carpal tunnel syndrome using hand diagrams. *J Occup Rehabil*. 1994;4:185-198. https://doi.org/10.1007/BF02331615
16. Graham B. The value added by electrodiagnostic testing in the diagnosis of carpal tunnel syndrome. *J Bone Joint Surg Am*. 2008;90:2587-2593. https://doi.org/10.2106/JBJS.G.01362
17. Graham B, Regehr G, Naglie G, Wright JG. Development and validation of diagnostic criteria for carpal tunnel syndrome. *J Hand Surg Am*. 2006;31:919-924. https://doi.org/10.1016/j.jhsa.2006.03.005
18. Hems TEJ, Miller R, Massraf A, Green J. Assessment of a diagnostic questionnaire and protocol for management of carpal tunnel syndrome. *J Hand Surg Eur Vol*. 2009;34:665-670. https://doi.org/10.1177/1753193409105566
19. Jaeschke R, Guyatt GH, Sackett DL, et al. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? *JAMA*. 1994;271:703-707. https://doi.org/10.1001/jama.1994.03510330081039
20. Kamath V, Stothard J. A clinical questionnaire for the diagnosis of carpal tunnel syndrome. *J Hand Surg Br*. 2003;28:455-459. https://doi.org/10.1016/s0266-7681(03)00151-7
21. Kassirer JP. Our stubborn quest for diagnostic certainty. *N Engl J Med*. 1989;320:1489-1491. https://doi.org/10.1056/NEJM198906013202211
22. Katz JN, Larson MG, Sabra A, et al. The carpal tunnel syndrome: diagnostic utility of the history and physical examination findings. *Ann Intern Med*. 1990;112:321-327. https://doi.org/10.7326/0003-4819-112-5-321
23. Katz JN, Stirrat CR. A self-administered hand diagram for the diagnosis of carpal tunnel syndrome. *J Hand Surg Am*. 1990;15:360-363. https://doi.org/10.1016/0363-5023(90)90124-a
24. Katz JN, Stirrat CR, Larson MG, Fossel AH, Eaton HM, Liang MH. A self-administered hand symptom diagram for the diagnosis and epidemiologic study of carpal tunnel syndrome. *J Rheumatol*. 1990;17:1495-1498.
25. Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers.

Biometrics. 1977;33:363-374. https://doi.org/10.2307/2529786
26. Levine DW, Simmons BP, Koris MJ, et al. A self-administered questionnaire for the assessment of severity of symptoms and functional status in carpal tunnel syndrome. *J Bone Joint Surg Am*. 1993;75:1585-1592. https://doi.org/10.2106/00004623-199311000-00002
27. 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
28. Lo JK, Finestone HM, Gilbert K. Prospective evaluation of the clinical prediction of electrodiagnostic results in carpal tunnel syndrome. *PM R*. 2009;1:612-619. https://doi.org/10.1016/j.pmrj.2009.05.004
29. MacDermid JC, Wessel J. Clinical diagnosis of carpal tunnel syndrome: a systematic review. *J Hand Ther*. 2004;17:309-319. https://doi.org/10.1197/j.jht.2004.02.015
30. Makanji HS, Zhao M, Mudgal CS, Jupiter JB, Ring D. Correspondence between clinical presentation and electrophysiological testing for potential carpal tunnel syndrome. *J Hand Surg Eur Vol*. 2013;38:489-495. https://doi.org/10.1177/1753193412461860
31. Naranjo A, Ojeda S, Mendoza D, Francisco F, Quevedo JC, Erausquin C. What is the diagnostic value of ultrasonography compared to physical evaluation in patients with idiopathic carpal tunnel syndrome? *Clin Exp Rheumatol*. 2007;25:853-859.
32. O'Gradaigh D, Merry P. A diagnostic algorithm for carpal tunnel syndrome based on Bayes's theorem. *Rheumatology (Oxford)*. 2000;39:1040-1041. https://doi.org/10.1093/rheumatology/39.9.1040
33. Sedighi I. Interpretation of diagnostic tests: likelihood ratio vs. predictive value [letter]. *Iran J Pediatr*. 2013;23:717.
34. Szabo RM, Slater RR, Jr., Farver TB, Stanton DB, Sharman WK. The value of diagnostic testing in carpal tunnel syndrome. *J Hand Surg Am*. 1999;24:704-714. https://doi.org/10.1053/jhsu.1999.0704
35. Wainner RS, Fritz JM, Irrgang JJ, Delitto A, Allison S, Boninger ML. Development of a clinical prediction rule for the diagnosis of carpal tunnel syndrome. *Arch Phys Med Rehabil*. 2005;86:609-618. https://doi.org/10.1016/j.apmr.2004.11.008
36. Wang WL, Buterbaugh K, Kadow TR, Goitz RJ, Fowler JR. A prospective comparison of diagnostic tools for the diagnosis of carpal tunnel syndrome. *J Hand Surg Am*. 2018;43:833-836.e2. https://doi.org/10.1016/j.jhsa.2018.05.022
37. 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

APPENDIX A

SEARCH STRATEGIES

Ovid (MEDLINE)

1. Carpal Tunnel Syndrome/
2. Carpal Tunnel Syndrome.mp.
3. Carpal Tunnel Syndrome/ or Nerve Compression Syndromes/ or Median Neuropathy/
4. Carpal Tunnel Syndrome/di [Diagnosis]
5. Median Neuropathy/di [Diagnosis]
6. median nerve entrapment*.mp. [mp = title, abstract, original title, name of substance word, subject heading word, floatingsubheading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
7. compression neuropathy.mp. [mp = title, abstract, original title, name of substance word, subject heading word, floating subheading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
8. Nerve Compression Syndromes/
9. cts.mp. [mp = title, abstract, original title, name of substance word, subject heading word, floating subheading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
10. syndrome, carpal tunnel.mp. [mp = title, abstract, original title, name of substance word, subject heading word, floating subheading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
11. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10
12. diagnostic test*.mp. [mp = title, abstract, original title, name of substance word, subject heading word, floating subheading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
13. clinical test*.mp.
14. diagnostic accuracy.mp. [mp = title, abstract, original title, name of substance word, subject heading word, floating subheading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
15. “Sensitivity and Specificity”/
16. sensitivity.mp. [mp = title, abstract, original title, name of substance word, subject heading word, floating subheading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
17. specificity.mp. [mp = title, abstract, original title, name of substance word, subject heading word, floating subheading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
18. roc curve.mp. [mp = title, abstract, original title, name of substance word, subject heading word, floating subheading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
19. 12 or 13 or 14 or 15 or 16 or 17 or 18
20. 11 and 19
21. (“Symptom diagram” or “hand diagram” or “Flick sign” or “Provocative Test*” or “Phalen’s test” or “phalen test” or “wrist flexion test” or “wrist extension test” or “reverse Phalen test” or “carpal compression test” or “Durkan’s test” or “Tinel’s sign” or “Tourniquet test” or “Gilliat test” or “Sensory test*” or “Motor Test*” or “Touchor vibration threshold” or “Current perception threshold” or “Two-point discrimination Semmes-Weinstein Monofilament Test” or “Thenar weakness” or “Thumb Abduction Weakness” or “thenar atrophy” or “Abductor Pollicis Brevis Manual Muscle Testing” or “CTS-Relief Maneuver” or “CTS-RM” or “Pin Prick Sensory Deficit” or “ULNT Criterion C” or “upper limb neurodynamic test Tethered median nerve stress test” or “Luthy’s sign” or “luthy sign” or “scratch collapse test” or “Pinwheel” or “CTS-6 evaluation tool” or “The Alderson-McGall hand function questionnaire” or “Hand elevation test” or “Katz and Stirrat hand diagram” or “katz hand diagram” or “Purdue Pegboard Test” or “Levine’s Self-Assessment Questionnaire” or “Dellon-modified Moberg pick-up test” or “Self-administered diagram” or “web-based questionnaire” “Kamath and Stothard questionnaire” or “Lo Carpal Tunnel Questionnaire” or “scratch-collapse test” or “hyperextension test” or “Flinn Performance Screening Tool” or “FPST”).mp. [mp = title, abstract, original title, name of substance word, subject heading word, floating subheading word, keyword heading word, protocol supplementary concept word, rare disease supplementary concept word, unique identifier, synonyms]
22. 20 and 21

Ovid (Embase)

1. carpal tunnel syndrome.mp. [mp = title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]
2. median neuropath*.mp. [mp = title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]
3. median nerve entrapment*.mp. [mp = title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]
4. compression neuropath*.mp. [mp = title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]

APPENDIX A

5. entrapment neuropath*.mp. [mp = title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]
6. carpal canal syndrome.mp. [mp = title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]
7. carpal tunnel compression*.mp. [mp = title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]
8. "neuropathy, median".mp. [mp = title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]
9. "syndrome,carpal tunnel".mp. [mp = title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]
10. carpal tunnel syndrome/
11. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10
12. clinical test*.mp.
13. "sensitivity and specificity"/
14. receiver operating characteristic/
15. differential diagnosis.mp. [mp = title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]
16. "diagnostic test*".mp. [mp = title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]
17. ("sensitivity" or "specificity").mp. [mp = title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]
18. "ROC curve".mp. [mp = title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]
19. diagnostic accuracy/ or diagnostic test accuracy study/ or differential diagnosis/ or physical examination/
20. 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19
21. 11 and 20
22. ("Symptom diagram" or "hand diagram" or "Flick sign" or "Provocative Test*" or "Phalen's test" or "phalen test" or "wrist flexion test" or "wrist extension test" or "reverse Phalen test" or "carpal compression test" or "Durkan's test" or "Tinel's sign" or "Tourniquet test" or "Gilliat test" or "Sensory test*" or "Motor Test*" or "Touchor vibration threshold" or "Current perception threshold" or "Two-point discrimination Semmes-Weinstein Monofilament Test" or "Thenar weakness" or "Thumb Abduction Weakness" or "thenar atrophy" or "Abductor Pollicis Brevis Manual Muscle Testing" or "CTS-Relief Maneuver" or "CTS-RM" or "Pin Prick Sensory Deficit" or "ULNT Criterion C" or "upper limb neurodynamic test Tethered median nerve stress test" or "Luthy's sign" or "luthy sign" or "scratch collapse test" or "Pinwheel" or "CTS-6 evaluation tool" or "The Alderson-McGall hand function questionnaire" or "Hand elevation test" or "Katz and Stirrat hand diagram" or "katz hand diagram" or "Purdue Pegboard Test" or "Levine's Self-Assessment Questionnaire" or "Dellon-modified Moberg pick-up test" or "Self-administered diagram" or "web-based questionnaire" or "scratch-collapse test" or "hyperextension test" or "Kamath and Stothard questionnaire" or "Lo Carpal Tunnel Questionnaire" or "Flinn Performance Screening Tool" or "FPST").mp. [mp = title, abstract, heading word, drug trade name, original title, device manufacturer, drug manufacturer, device trade name, keyword, floating subheading word, candidate term word]
23. 21 and 22

CINAHL

S1. (MH "Carpal Tunnel Syndrome")
S2. "median neuropath*"
S3. "median nerve entrapment*"
S4. "compression neuropath*"
S5. "entrapment neuropath*"
S6. S1 OR S2 OR S3 OR S4 OR S5
S7. "diagnosis or assessment"
S8. "diagnosis"
S9. "diagnostic"
S10. (MH "Diagnosis") OR (MH "Diagnosis, Neurologic") OR (MH "Diagnosis, Musculoskeletal") OR (MH "Exercise Test") OR (MH "Functional Assessment") OR (MH "Patient Assessment") OR (MH "Patient History Taking") OR (MH "Physical Examination") OR (MH "Sensitivity and Specificity")
S11. (MH "Diagnosis, Musculoskeletal") OR (MH "Diagnosis, Neurologic") OR (MH "Functional Assessment") OR (MH "Patient Assessment") OR (MH "Patient History Taking") OR (MH "Physical Examination") OR (MH "Sensitivity and Specificity") OR (MH "Skin Tests")
S12. (MH "Sensitivity and Specificity") OR "sensitivity and specificity" OR (MH "ROC Curve")

APPENDIX A

S13. "sensitivity"
S14. "specificity"
S15. S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13 OR S14
S16. S6 AND S15
S17. "Symptom diagram" or "hand diagram" or "Flick sign" or "Provocative Test*" or "Phalen's test" or "phalen test" or "wrist flexion test" or "wrist extension test" or "reverse Phalen test" or "carpal compression test" or "Durkan's test" or "Tinel's sign" or "Tourniquet test" or "Gilliat test" or "Sensory test*" or "Motor Test*" or "Touch"or "vibration threshold" or "Current perception threshold" or "Two-point discrimination" "Semmes-Weinstein Monofilament Test" or "Thenar weakness" or "Thumb Abduction Weakness" or "thenar atrophy" or "Abductor Pollicis Brevis Manual Muscle Testing" or "CTS-Relief Maneuver" or "CTS-RM" or "Pin Prick Sensory Deficit" or "ULNT Criterion C" or "upper limb neurodynamic test" "Tethered median nerve stress test" or "Luthy's sign" or "luthy sign" or "scratch collapse test" or "Pinwheel" or "CTS-6 evaluation tool" or "The Alderson-McGall hand function questionnaire" or "Hand elevation test" or "Katz and Stirrat hand diagram" or "katz hand diagram" or "Purdue Pegboard Test" or "Levine's Self-Assessment Questionnaire" or "Dellon-modified Moberg pick-up test" or "Self-administered diagram" or "web-based questionnaire" or "Kamath and Stothard questionnaire" or "Lo Carpal Tunnel Questionnaire" or "scratch-collapse test" or "hyperextension test" or "Flinn Performance Screening Tool" or "FPST"
S18. S16 AND S17

APPENDIX B

CONFLICTS OF INTEREST FOR INCLUDED STUDIES

Study	Conflict of Interest
Scales and questionnaires	
Bland[4]	No conflict of interest statement
Bland et al[6]	No conflict of interest statement
Bland et al[5]	No conflicts of interest
Bougea et al[8]	No conflicts of interest
Bridges et al[9]	No conflict of interest statement
Fowler et al[14]	There was no outside funding for this study
Hems et al[18]	No conflict of interest statement
Kamath and Stothard[20]	No conflicts of interest
Lo et al[28]	No conflict of interest statement
Makanji et al[30]	No conflicts of interest
Naranjo et al[31]	No conflict of interest statement
Wang et al[36]	No benefits in any form have been received or will be received related directly or indirectly to the subject of this article
Hand symptom diagrams/maps	
Ammer et al[2]	No conflict of interest statement
Bonauto et al[7]	None declared
Calfee et al[10]	No benefits in any form have been received or will be received related directly or indirectly to the subject of this article
Franzblau et al[15]	No conflict of interest statement
Katz and Stirrat[23]	No conflict of interest statement
Katz et al[24]	Grant support: National Institutes of Health grants AR36308 and AR07530 and the Kellogg Program for Training in Research in Clinical Effectiveness
Katz et al[22]	No conflict of interest statement
O'Gradaigh and Merry[32]	No conflict of interest statement
Szabo et al[34]	No benefits in any form have been received or will be received from a commercial party related directly or indirectly to the subject of this article

APPENDIX C

CHARACTERISTICS OF THE STUDIES ASSESSING DIAGNOSTIC SCALES AND QUESTIONNAIRES FOR THE DIAGNOSIS OF CTS, AND THEIR REFERENCE STANDARDS

Study, Design, Country	Diagnostic Tool	Population Characteristics	Participant Inclusion and Exclusion Criteria	Index Test	Reference Standard Test	Risk of Bias
Bland[4] Prospective cross-sectional United Kingdom	Bland questionnaire	n = 7768 (5392 female) consecutive subjects; CTS cases, n = 3710 TP; age, 10-98 y; symptom duration: 0-3 mo, 5.6%; 3-6 mo, 17.8%; 6-12 mo, 18.0%; >12 mo, 46.1%	All patients with suspected CTS were referred for NCS No exclusion criteria were reported	A small selection of questions, all of which were arranged in multiple choice/checkbox, performed by the participants A cutoff probability of 0.5	NCS of median and ulnar orthodromic sensory conduction from finger to wrist and measures of the motor terminal latency to the APB recorded on both hands, supplemented by either a sensory potential recorded at the wrist on ring finger stimulation, performed by a neurologist Normal values were defined as those within 2.5 SDs of the mean	Unclear (2 domains)
Bland et al[6] Retrospective cross-sectional United Kingdom	Combined KSQ and CTS-7[a]	n = 5860 consecutive subjects; no more detail	Patients who sought medical attention with suspected CTS for the first time Excluded those with previous surgery to either side or recurrence after successful conservative treatment. Did not exclude patients with concomitant pathologies, such as diabetic polyneuropathy or ulnar neuropathy	The CTS-7 includes examination findings (Tinel's and/or Phalen's signs), and the authors aimed to study data that could be collected from the patient without medical intervention	NCS were carried out on both hands of all patients according to AANEM standards	Unclear (1 domain)
Bland et al[5] Prospective cross-sectional United Kingdom	Bland web-based questionnaire	n = 2655 consecutive subjects (67% female); age, 54.2 y; CTS cases, n = 1430 TP	Primary care physicians' referrals of suspected CTS patients Excluded those who already had known CTS prior to visiting the website, those having tests for follow-up purposes, or those who had already had treatment for one hand and were returning for management of the other. No exclusions were made on the grounds of age, sex, or coincident pathology	Patients were asked to visit http://www.carpal-tunnel.net prior to their appointment (takes 20-30 min) Cutoff point of website score of 40% was used to diagnose CTS	NCS according to guidelines published by the AANEM The NCS results were graded using the Canterbury severity scale for CTS, which represents the changes in sensory and motor NCVs and amplitudes as a numeric scale, increasing in severity from 0 (no abnormality) to 6 (extremely severe CTS)	Unclear (2 domains)
Bougea et al[8] Prospective cross-sectional Greece	Greek version of the BCTQ	n = 90 consecutive subjects (75% female); age, 57.3 ± 13.8 y; CTS severity: grade 1, 18.9%; grade 2, 6.7%; grade 3, 42.2%; grades 4-6, 12.2%	Patients referred to the electrophysiology laboratory with symptoms consistent with CTS Included: age, ≥18 y; first-time visitors not previously diagnosed by the investigators; absence of severe intellectual disability or cognitive impairment Excluded: polyneuropathy, systemic diseases potentially associated with polyneuropathy (diabetes mellitus, renal failure, hypothyroidism, or amyloidosis), other diseases that cause hand symptoms (eg, cervical radiculopathy or thoracic outlet syndrome), pregnancy	The overall FSS and SSS scores from the BCTQ were calculated Cutoff point: scores of ≥1.95	EMG based on the AANEM guidelines Used the Canterbury severity scale for CTS, which expresses the modifications of sensory and motor NCVs and amplitudes as a numeric scale for the EMG grading of severity, from 0 (no abnormality) to 6 (extremely severe CTS)	Unclear (2 domains)

Table continues on page D6.

APPENDIX C

Study, Design, Country	Diagnostic Tool	Population Characteristics	Participant Inclusion and Exclusion Criteria	Index Test	Reference Standard Test	Risk of Bias
Bridges et al[9] Prospective cross-sectional United Kingdom	KSQ	n = 211 consecutive subjects (57% female); age, 52.7 ± 14.0 y	Patients who had been referred for electrophysiological testing with symptoms suggestive of CTS Excluded patients with diabetes	All patients attending a hand clinic routinely fill out the KSQ, which consists of 9 questions relating to possible symptoms of CTS, performed by a rheumatologist Cutoff point: scores of >6 and <3	NCS, performed by a single trained doctor who was also responsible for administering the questionnaire Positive test if: onset motor latency to the APB of >4.2 ms, peak sensory latency to index finger of >4.0 ms, a difference in onset motor latency between the APB and ipsilateral ADM of >1.5 ms, a difference in motor latencies between both APBs of >1.0 ms, or a reduction of median sensory amplitude of >50% of either the ipsilateral ulnar sensory latency or the contralateral median nerve	High
Fowler et al[14] Retrospective cross-sectional United Kingdom	CTS-6	n = 85 consecutive subjects; age, 55 y (range, 28-87 y); pre-exam CTS probability, 6%; CTS cases, n = 55 TP	A data set of patients referred to EDS from an orthopaedic hand surgery practice with a higher prevalence of CTS than that in the general population	The CTS-6 score was calculated by a blinded examiner who was not involved in US or NCS A score of 12 points was considered a positive CTS-6 score	The authors used latent class analysis (Bayesian methods) as their reference standard and compared the scores obtained from the CTS-6 to NCS and US. NCS was conducted according to AANEM guidelines. A DML of 4.2 ms and a DSL of 3.2 ms were used as the cutoffs for a positive diagnosis of CTS. The cross-sectional area of the median nerve was measured at the inlet to the carpal tunnel, using a 13- to 6-MHz linear array transducer, by a blinded hand surgeon. The a priori cutoff of 10 mm^2 was used as the cutoff for a positive US examination	Unclear (1 domain)
Hems et al[18] Prospective cross-sectional United Kingdom	Bland questionnaire	n = 152 consecutive subjects (108 female); symptom duration: >12 mo, n = 125; 6-12 mo, n = 20; 3-6 mo, n = 7	All patients referred to the hand clinic with suspected CTS during the period of the study were asked to consent to participation in the study and to complete the questionnaire	A questionnaire that has 2 parts, filled out by both the participants and clinicians Cutoff point: score of ≥7	NCS: they measured the latency of sensory conduction (positive if thumb to median nerve was 40.5 ms greater than thumb to radial/motor latency, or positive if median nerve to APB was >4.1 ms)	Low
Kamath and Stothard[20] Prospective cross-sectional United Kingdom	KSQ	n = 58 consecutive subjects with definite CTS diagnosis (67 female in the original population before exclusion criteria were applied)	Patients referred with a diagnosis of CTS to a hand clinic Included: definite diagnosis of CTS by a physician Excluded: a possible generalized neuropathy (eg, those with diabetes mellitus), renal transplant patients, pregnant patients	A questionnaire based on the BCTQ, filled out by a hand surgeon Cutoff point: score of >5 on the KSQ	CTR Positive criterion: symptom relief at 2 wk after surgery	Unclear (1 domain)

Table continues on page D7.

APPENDIX C

Study, Design, Country	Diagnostic Tool	Population Characteristics	Participant Inclusion and Exclusion Criteria	Index Test	Reference Standard Test	Risk of Bias
Lo et al[28] Prospective cross-sectional Canada	Lo CPR	n = 278 consecutive subjects (58.8% female); age, 50 ± 12.7 y; CTS cases, n = 149 TP	Subjects referred to the electrodiagnostic laboratory over a 1-y period with a clinical suspicion of CTS	The subject's point score was determined by a physiatrist, based on the information and clinical findings obtained during history and physical examination	NCS by a blinded electrodiagnostic technologist; AANEM references Positive criteria: a combination of a median-to-ulnar transcarpal latency difference of 0.4 ms and median transcarpal latency of 2.2 ms	Low
Makanji et al[30] Prospective cross-sectional United States	CTS-6	n = 78 consecutive subjects (62% female); age, 55 ± 15 y; CTS severity: mild, n = 16; moderate, n = 46; severe, n = 16	Adult patients in the practice of 3 hand surgeons were prescribed electrophysiological testing and invited to participate Included: patients suspected to have CTS Excluded: prior CTR, injury to the wrist or hand, previous electrophysiological testing of the median nerve, and pregnancy	The instrument assigns varying weights to 6 symptoms and clinical maneuvers and determines the probability of having CTS using a logistic regression equation Cutoff point: score of 50%	NCS and EMG. The median nerve was stimulated at the wrist, and antidromic sensory action potentials were recorded 13 cm distally at the index finger for DSL studies. The median nerve motor action potential was recorded at the APB muscle and stimulated at the wrist 7 cm proximal to the electrodes for DML studies The presence of 1 or both of the following: DSL of ≥3.6 ms and/or DML of ≥4.4 ms based on AANEM	Unclear (3 domains)
Naranjo et al[31] Prospective cross-sectional Spain	BCTQ	n = 68 consecutive subjects (56 female), 105 wrists (54% bilateral); age, 47 ± 11 y; CTS cases, 80 TP wrists; CTS severity: mild, n = 13; moderate, n = 30; severe, n = 37; symptom duration, 21 mo (IQR, 8-36 mo)	Adult patients with suspected CTS referred to the outpatient rheumatology clinic Included: burning pain or numbness aggravated by sustained positions and relief by shaking or moving the hands, sleep disruption by symptoms, and daily complaints over at least a 3-mo period Excluded: surgery or traumatic injuries at the target wrist, hypothyroidism, acromegaly, polyneuropathy or radiculopathy, pregnancy, fibromyalgia, rheumatoid arthritis or crystal arthritis, had received injections or presented ganglions, tenosynovitis, or arthritis	The BCTQ has 2 components: a hand function scale and hand sensitivity (sensory). Filled out by a rheumatologist. Two different diagnostic accuracy measures were calculated for each component Cutoff point: score of >3	NCS, AANEM referenced Performed by 2 neurologists	Unclear (2 domains)
Wang et al[36] Prospective cross-sectional United States	CTS-6, KSQ, Lo CPR, Wainner CPR	n = 408 consecutive wrists of 250 subjects (181 female); wrists with definite CTS, n = 255; age, 52 ± 14 y	Patients were identified and recruited through an orthopaedic hand surgery clinic Included: patients who returned to the office after being previously referred for electrodiagnostic testing for the assessment of CTS Excluded: patients <18 y of age and unable to comprehend English or give consent	Questionnaires were filled out by a hand fellowship–trained surgeon Cutoff points of 18 on the CTS-6, 5 on the KSQ, 10 on the Lo CPR, and 3 on the Wainner CPR	Clinical diagnosis (no further explanations)	Unclear (3 domains)

Abbreviations: AANEM, American Association of Neuromuscular and Electrodiagnostic Medicine; ADM, abductor digitorum minimi; APB, abductor pollicis brevis; BCTQ, Boston Carpal Tunnel Questionnaire; CPR, clinical prediction rule; CTR, carpal tunnel release; CTS, carpal tunnel syndrome; CTS-6, 6-item carpal tunnel syndrome symptoms scale; CTS-7, 7-item carpal tunnel syndrome symptoms scale; DML, distal motor latency; DSL, distal sensory latency; EDS, electrodiagnostic studies; EMG, electromyography; FSS, functional status scale; IQR, interquartile range; KSQ, Kamath and Stothard questionnaire; NCS, nerve conduction studies; NCV, nerve conduction velocity; SSS, symptom severity scale; TP, true positive; US, ultrasound.

[a]*Only questionnaire segments were used in this study*[6]*; the authors called this test "CTS-7"; however, the rationale behind this naming is unclear, and they have referenced the CTS-6 paper.*[17]

APPENDIX D

CHARACTERISTICS OF THE STUDIES ASSESSING SYMPTOM DIAGRAMS/MAPS FOR THE DIAGNOSIS OF CTS, AND THEIR REFERENCE STANDARDS

Study,[a] Design, Country	Population Characteristics	Participant Inclusion and Exclusion Criteria	Index Test	Reference Standard Test	Risk of Bias
Ammer et al[2] Prospective cross-sectional Austria	n = 101 consecutive subjects (68 female; 147 wrists); age, 57.7 ± 15.8 y; wrists with classic, probable, or possible CTS, n = 120	Patients suspected to have CTS Asymptomatic hands and patients with normal NCV were excluded from the analysis	Patients were asked to mark pain, tingling, and numbness in the diagram	NCS. All tests were performed with an EMG system Normal values were: distal latency of motor fibers at a distance of 5.5 cm = 2.994 ms + 0.004 × age (SD, 0.392) and antidromic conduction velocity of sensory fibers: Vs = 71.99 m/s – 0.3 × age (SD, 4.86)	High
Bonauto et al[7] Prospective cross-sectional United States	n = 253 subjects with current hand symptoms and n = 179 subjects with numbness, tingling, or pain (48% female); age, 39.5 ± 10.9 y	Workers from 12 work sites in the manufacturing (electronics, automotive parts, windows, cabinets, medical and fitness equipment) and health care (hospitals, excluding direct patient care and health research) sectors Excluded: sudden shoulder injury, part-time workers, temporary workers, workers in a mobile job such as a forklift driver, or with more than 4 job tasks	Workers were asked to complete a body map describing the distribution of pain or discomfort in the neck, shoulder, elbow/forearm, and hand/wrist if they had problems in the past year that either lasted a week or more or had occurred at least 3 times. A classic/probable/possible diagram rating was considered "positive"	NCS, AANEM referenced, performed by nerve conduction technicians. Positive if at least 1 of the following findings present: median motor latency of 4.0 ms and/or median sensory latency of 3.7 ms	Unclear (1 domain)
Calfee et al[10] Prospective cross-sectional United States	n = 221 consecutive subjects (71% male; 216 with DSL analysis); age, 31.8 ± 10.6 y; positive CTS according to Katz scores, n = 59	CTS suspects: workers with hand symptoms from 11 companies or organizations Included: symptoms of burning, pain, tingling, or numbness Excluded: a history of CTS, peripheral neuropathy, current pregnancy, or inability to have nerve conduction testing	The instructions asked subjects to shade in the area of the problem but not to try to represent the type of their symptoms on the diagram Scoring was performed according to the recommendations of Katz and Stirrat,[23] with modification: scores were dichotomized as positive ("classic" or "probable") or negative ("possible" or "unlikely") The scoring of the diagrams was done by 2 physicians and 1 occupational therapist	NCS with an automated device. Positive if a DSL of >3.5 ms, a DML of >4.5 ms, or paired transcarpal median-ulnar sensory difference of >0.5 ms. Transcarpal DSL measurements were recorded in the long finger	Unclear (3 domains)
Franzblau et al[15] Prospective cross-sectional United States	n = 411 consecutive subjects (41.6% male); age, 35.7 ± 10.5 y; pre-exam CTS probability, 15%	At-risk workers from 4 unrelated companies Included: certain jobs were selected on the basis of the frequency of repetitive hand movements ("low," "medium," and "high"), and all workers with at least 6 mo of tenure in those jobs were invited to participate	Similar to the diagram and instructions used by Katz et al[22] Patients were instructed to shade in the distribution of numbness, tingling, burning, or pain in the wrists, hands, or fingers on the hand diagram	NCS performed by physicians certified in EDS medicine and median and ulnar sensory conduction studies in the wrists, using surface electrodes and fixed distances (14 cm, antidromic stimulation) Positive if a difference of at least 0.5 ms between median and ulnar sensory peak latencies in the same wrist	Unclear (2 domains)

Table continues on page D9.

APPENDIX D

Study,[a] Design, Country	Population Characteristics	Participant Inclusion and Exclusion Criteria	Index Test	Reference Standard Test	Risk of Bias
Katz and Stirrat[23] Retrospective United States	n = 149 random subjects (73% female); wrists with definite CTS, n = 85; age, 45.6 ± 14 y	Patients with UE paresthesia Included: CTS diagnosis based on NCS, unequivocal response to corticosteroid injection in the carpal tunnel, and improvement after CTR Excluded: patients with presumptive diagnoses that were not confirmed by these criteria, diabetes, heavy ethanol use, hypothyroidism, rheumatoid arthritis, renal disease, ulnar entrapment, cartilaginous lesions, dorsal cutaneous nerve injury, C6-7 radiculopathy, symptomatic hamate fracture	Patients were asked to shade in the area of their discomfort on the diagram and indicate their quality of symptoms. Patients with CTS were categorized into 4 categories: classic, probable, possible, unlikely	Clinical diagnosis Diagnoses, comorbid conditions, and demographic data were abstracted from patients' charts without knowledge of their diagram ratings	High
Katz et al[24] Prospective cross-sectional United States	n = 110 consecutive subjects (145 wrists); age, 45.6 ± 14.4 y; CTS severity: classic, n = 18; probable, n = 16; possible, n = 17; unlikely, n = 2	Patients >18 y of age referred to a nerve conduction lab for evaluation of UE discomfort Control group diagnosis: cervical radiculopathy, ulnar neuropathy, brachial plexopathy, polyneuropathy	Patients were asked to complete a diagram before the NCS. Patients with CTS were categorized into 4 categories: classic, probable, possible, unlikely	NCS performed on an EMG apparatus, with skin temperature maintained at 34°C-37°C. Positive if at least 1 of the following findings present: motor latency >4.0 ms, sensory latency >3.7 ms, sensory velocity <50 m/s	Low
Katz et al[22] Prospective cross-sectional United States	n = 110 consecutive subjects (66.4% female; 165 wrists); wrists with definite CTS, n = 44; age, 45.6 ± 14.4 y	Patients >18 y of age suspected to have CTS were referred to a nerve conduction lab for evaluation of UE discomfort Control group diagnosis: cervical radiculopathy, ulnar neuropathy	Patients completed a self-administered hand pain diagram that depicted both hands with dorsal and palmar views. Patients were asked to mark areas on the diagram corresponding to the location of their symptoms and to indicate the quality of their discomfort. Patients with CTS were categorized into 4 categories: classic, probable, possible, unlikely	NCS. The protocol included bilateral median and ulnar sensory and motor testing and EMG recording from the APB on the most symptomatic hand. Testing was done with standard techniques on an EMG apparatus, with skin temperature maintained at 34°C-37°C Positive if patients had median motor latency >4.0 ms, sensory latency >3.7 ms, or sensory velocity <50 m/s. Performed by neurologist	Low
O'Gradaigh and Merry[32] Prospective cross-sectional United Kingdom	n = 105 consecutive subjects	Suspicion of CTS by the refereeing clinician on any grounds Excluded: previously treated for CTS or with recognized associated conditions	Patients outlined their symptomatic areas on the diagram Those with classic or probable distributions were considered positive	NCS. Sensory amplitude <10 μV or motor latency >3.7 ms	Unclear (2 domains)
Szabo et al[34] Prospective cross-sectional United States	n = 100 consecutive subjects; subjects with definite CTS, n = 50 (38 female; 87 wrists); subjects with other diagnoses of the UE, n = 50 (40 female; 90 hands); age (CTS), 20-73 y; age (non-CTS), 28-70 y; symptom duration, 2 mo to 20 y; pre-exam CTS probability, 15%	Patients who were evaluated and treated at an institution for hand, wrist, and forearm problems Included: a clinical history of numbness and tingling in the median nerve distribution and/or night pain relieved by shaking of the hand; results of physical examination, including sensibility and provocative examinations, consistent with CTS; and relief of symptoms after CTR Control group diagnosis: epicondylitis, De Quervain's tenosynovitis and other tendinosis, radiculopathy, and hand pain of unknown etiology	Subjects completed the hand diagram themselves; it was then scored blindly by one of the authors as classic, probable, possible, or unlikely for CTS, according to the criteria described by Katz and Stirrat[23]	NCS. Bilateral median and ulnar motor and sensory nerve conduction tests were the electrodiagnostic parameters considered. Testing was done with standard techniques, with the skin temperature maintained at 34°C-37°C	High

Abbreviations: AANEM, American Association of Neuromuscular and Electrodiagnostic Medicine; APB, abductor pollicis brevis; CTR, carpal tunnel release; CTS, carpal tunnel syndrome; DML, distal motor latency; DSL, distal sensory latency; EDS, electrodiagnostic studies; EMG, electromyography; NCS, nerve conduction studies; NCV, ; UE, upper extremity nerve conduction velocity.
[a]*All studies used the Katz and Stirrat[23] hand symptom diagram.*

NORMAN E. D'HONDT, PT, MSc[1,2] • JAN J.M. POOL, PT, PhD, MT[2] • HENRI KIERS, PT, PhD, MT[2]
CAROLINE B. TERWEE, PhD[3] • (DIRKJAN) H.E.J. VEEGER, PhD[1,4]

Validity of Clinical Measurement Instruments Assessing Scapular Function: Insufficient Evidence to Recommend Any Instrument for Assessing Scapular Posture, Movement, and Dysfunction—A Systematic Review

OBJECTIVE: To determine the construct validity, criterion validity, and responsiveness of measurement instruments evaluating scapular function.

DESIGN: Systematic review of measurement properties.

LITERATURE SEARCH: The MEDLINE, Embase, CINAHL, and SPORTDiscus databases were systematically searched from inception until March 2019.

STUDY SELECTION CRITERIA: Studies published in Dutch, English, or German were included when they evaluated at least 1 of the measurement properties of interest. No restrictions were made regarding participants' health status.

DATA SYNTHESIS: Two reviewers independently evaluated study quality using the COSMIN checklist and extracted and analyzed data. Quality of evidence was graded by measurement property for each distinctive type of measurement.

RESULTS: Thirty-one measurement instruments in 14 studies were categorized into instruments to measure scapular posture and movement, and to assess scapular dyskinesis. Quality of evidence was at most moderate for 4 instruments with respect to criterion validity. Of these, criterion validity for instruments measuring scapular protraction/retraction posture and rotation angles up to 120° of thoracohumeral elevation was sufficient. Criterion validity for instruments measuring asymmetrical scapular posture, range of motion, and the lateral scapular slide test was insufficient. Quality of evidence for measurement properties of all other instruments was graded lower.

CONCLUSION: There is currently insufficient evidence to recommend any instrument for the clinical examination of scapular function. Measurement instruments to assess scapular dyskinesis are prone to misinterpretation and should therefore not be used as such. *J Orthop Sports Phys Ther 2020;50(11):632-641. doi:10.2519/jospt.2020.9265*

KEY WORDS: *movement, physical examination, review, shoulder pain*

The etiology of shoulder complaints is multifactorial.[56] Scapular dyskinesis is one factor frequently suggested to mechanically contribute to the onset and persistence of shoulder complaints.[3,16] Although debated, scapular dyskinesis generally encompasses a concept of scapular dysfunction that is reflected by aberrant scapular posture and movement patterns, such as scapular winging and asymmetry.[22] Such scapular movement and posture are believed to evoke circumstances where soft tissue structures in and around the shoulder joint are exposed to detrimental mechanical stress.[26,53] Assessing scapular movement and posture is therefore often part of the clinical-reasoning process of clinicians who manage shoulder pain.

Assessments of scapular upward rotation and visual evaluation systems to assess scapular dyskinesis have been advocated as reliable and valid for clinical use.[12,24] However, we recently studied the reliability of these measures and found that the evidence was insufficient to recommend their use.[8] There is also growing evidence that common treatment programs that target scapular dyskinesis may have limited clinical

[1]Neuromechanics Section, Department of Human Movement Sciences, Faculty of Behavioral and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam Movement Sciences, Amsterdam, the Netherlands. [2]Lifestyle and Health Research Group, Institute of Human Movement Studies, Hogeschool Utrecht University of Applied Sciences Utrecht, Utrecht, the Netherlands. [3]Department of Epidemiology and Biostatistics, Vrije Universiteit Medical Center, Amsterdam Public Health Research Institute, Amsterdam, the Netherlands. [4]Department of Biomechanical Engineering, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Delft, the Netherlands. This review was partially funded by the Dutch Research Council (Nederlandse Organisatie voor Wetenschappelijk Onderzoek) with a PhD grant. 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 Norman Eduard D'hondt, Faculty of Behavioral and Movement Sciences, Vrije Universiteit Amsterdam, Van der Boechorststraat 9, 1081 BT, Amsterdam, the Netherlands. E-mail: norman.dhondt@gmail.com

value.[5,7] Improvements in pain and function following such treatment programs were similar to nonscapula-focused treatments and were often not accompanied by changes in scapular kinematics or clinically established dyskinesis.[4,43] Scapular movement and posture that are clinically classified as dyskinesis might also represent adaptive strategies to avoid further tissue damage or threat, or they might be random expressions of natural variations in movement and be clinically irrelevant.[17,18] These findings raise questions about contemporary ideas of scapular dyskinesis and how measurement instruments are used to assess scapular function. Critically evaluating the underlying constructs and validity of measurement instruments will help determine the extent to which such measurement instruments truly reflect scapular function and dyskinesis, and whether they are suitable to assess scapular movement under either clinical or research conditions. Therefore, in this systematic review, we evaluated the current evidence on construct validity, criterion validity, and responsiveness of measurement instruments used to measure and assess scapular movement and posture.

METHODS

This systematic review adhered to the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) taxonomy,[31] the COSMIN Protocol for Systematic Reviews of Measurement Properties (November 2011), which is a precursor to the recently published COSMIN guideline,[41] and the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines.[28]

The review protocol was not registered in PROSPERO.[37] Following the recommendations of the Ann Arbor Clinical Heterogeneity Consensus Group,[13] we assembled a diverse group of reviewers, consisting of clinical epidemiologists (J.P., H.K., C.T.), clinicians (J.P., N.D., H.K.), and a biomechanist (H.V.).

Search Strategy

A comprehensive search for eligible studies was conducted in the electronic databases MEDLINE, Embase, CINAHL, and SPORTDiscus, from inception until March 2019. Search strings (**APPENDIX A**, available at www.jospt.org) were constructed in collaboration with a research librarian (Liedeke van Schoot),[44] using validated search filters for studies on measurement properties.[52] Additional searches were performed in each database, including the naming of the instruments found in the initial search. Reference lists of the obtained articles were browsed and experts in our network were contacted.

Eligibility Criteria for Selecting Studies

Studies were included if (1) the aim of the measurement instrument was to clinically measure or to assess scapular posture and/or motion; (2) at least 1 of the measurement's properties included hypotheses testing for construct validity, criterion validity, or responsiveness; (3) the design was original validation; and (4) they were published in Dutch, English, or German. As instruments used to measure only scapular posture and scapular movement do not require a symptomatic population to study measurement properties, no restrictions were made on the basis of health status of the study population. Studies were excluded if they (1) only determined diagnostic test accuracy, (2) involved the evaluation of a measurement instrument to measure or assess thoracohumeral function, without reporting separate data on scapular posture or motion, (3) evaluated measurement instruments to measure or assess scapular muscular length or strength, or (4) evaluated measurement instruments requiring an extensive calibration procedure (ie, when deployed in a laboratory setting). An expert panel (H.K., N.D., H.V.) judged eligibility when the clinical applicability of a measurement instrument was unclear.

Study Selection Process

Two reviewers (N.D., J.P.) independently performed the study selection. Eligibility was determined by applying the criteria to the title, abstract, and full text. Full-text articles were obtained when the abstract clearly confirmed the study's eligibility or when the abstract provided insufficient information to definitively reject the study. Discordance in study selection was resolved by discussion rounds.

Data Extraction and Management

Data and study characteristics were independently extracted and analyzed using forms by 2 reviewers (N.D., J.P.). Results of the studies were then classified as evidence for hypotheses testing for construct validity, criterion validity, and responsiveness. In the COSMIN taxonomy, hypotheses testing for construct validity is "the degree to which the scores of a test are consistent with hypotheses based on the assumption that an instrument validly measures the construct to be measured."[30] Criterion validity is "the degree to which the scores of a test are an adequate reflection of a gold standard of the construct to be measured."[30] Responsiveness is "the ability of a test to detect change over time in the construct to be measured" and can therefore be seen as longitudinal validity.[30] Authors were not contacted to retrieve missing data.[46] We grouped the measurement instruments into 3 main categories based on whether (1) scapular posture or (2) scapular movement was measured, or whether (3) scapular dyskinesis was assessed. To counteract difficulties due to heterogeneous descriptions and measurement protocols of measurement instruments with the same name across studies,[8] we only considered what was factually measured for further analysis, and ignored the names of the instruments.

Assessment of Methodological Quality

Two reviewers (N.D., J.P.) independently assessed the methodological quality of each study on a 4-point rating scale version of the COSMIN checklist.[2] Although

TABLE 1 Criteria for Measurement Properties, Based on Terwee et al[50]

Measurement Property	Sufficient	Insufficient	Indeterminate
Hypotheses testing for construct validity	Correlation with an instrument measuring the same construct of ≥0.50, or at least 75% of the results are in accordance with the hypotheses and the correlation with related constructs is higher than with unrelated constructs	Correlation with an instrument measuring the same construct of <0.50, or <75% of the results are in accordance with the hypotheses or the correlation with related constructs is lower than with unrelated constructs	Correlations solely determined with unrelated constructs
Criterion validity	Correlation with gold standard of ≥0.70 or AUC≥0.70	Correlation with gold standard of <0.70 or AUC<0.70, despite adequate design and method	Doubtful design or method[a]
Responsiveness	Correlation with an instrument measuring the same construct of ≥0.50, or at least 75% of the results are in accordance with the hypotheses, or an AUC≥0.70 and the correlation with related constructs is higher than with unrelated constructs	<75% of the results are in accordance with the hypotheses, or an AUC<0.70, or the correlation with related constructs is lower than with unrelated constructs	Correlations solely determined with unrelated constructs

Abbreviation: AUC, area under the receiver operating characteristic curve.
[a]*Lacking a clear description of the design or methods of the study, or any important methodological flaw in the design or execution of the study.*

TABLE 2 Quality of Evidence for the Overall Quality of the Measurement Property, Based on Schellingerhout et al[45]

Quality of Evidence	Criteria[a]
High	Multiple studies with methodological quality rated “good” or above or 1 study with methodological quality rated “excellent” and consistent findings (sufficient or insufficient) and total sample size ≥100
Moderate	Multiple studies with methodological quality rated “fair” or 1 study with methodological quality rated “good” and consistent findings (sufficient or insufficient) and total sample size ≥50
Low	1 study with methodological quality rated “fair” and consistent findings (sufficient or insufficient)
Conflicting	Multiple studies and conflicting findings
No evidence	Only studies with methodological quality rated “poor,” or only studies with indeterminate findings

[a]*For definitions of sufficient, insufficient, and indeterminate, see* **TABLE 1**.

the checklist was primarily developed for evaluating health-related patient-reported outcome instruments, it can also be used to evaluate measurement properties of other types of evaluative measurement instruments.[2,10,51]

We only used boxes F (hypotheses testing for construct validity), H (criterion validity), and I (responsiveness) of the checklist in our systematic review. Both assessors were experienced in using the assessment tool[8] and were not blinded to study identifiers during the assessment procedure. The assessment procedure involved a “worst score counts” method. Each item was scored on a 4-point ordinal scale as excellent, good, moderate, or poor. The lowest scored item in the box determined the final score. If the deployed reference standard in a study on criterion validity could not be considered as a reasonable gold standard (box H, item 4), then the study was defined as a study on hypotheses testing for construct validity and further assessed accordingly (box F).

Because several small, high-quality studies can together provide sufficient evidence for a measurement property, assessment of sample-size requirements was omitted from the scoring procedure and considered in the procedure of grading evidence instead.[41] Disagreements on item assessments were resolved by consensus rounds.

Evaluation of Measurement Properties

We used a precursor[50] to the most recent COSMIN criteria[41] to evaluate the measurement properties (**TABLE 1**). Each measurement property was rated sufficient, insufficient, or indeterminate.

Grading Quality of Evidence

Descriptive analyses were undertaken, and quality of evidence was graded by measurement property for each distinctive type of measurement[45] (**TABLE 2**). We added sample-size requirements to the criteria for quality of evidence.[2,8,10]

RESULTS

The primary electronic search yielded 10 057 records. Of these, 200 records were considered potentially eligible. Based on abstract screening, 30 studies remained for full-text review. Ultimately, 14 studies including 31 different measurement instruments were included (**FIGURE**).

Measurement instruments for scapular posture and scapular movement (categories 1 and 2, respectively) generally involved inclinometers and methods measuring distances between several scapular and thoracic bony landmarks. Measurement instruments for scapular dyskinesis (category 3) consisted mainly of subjective methods based on the assumption that observed deviation from either a predefined "normal" scapular posture or an ideal scapular movement pattern was dyskinesis.

Methodological Quality of Selected Studies

Two studies[34,54] that reported on a measurement instrument's construct validity (box F) had fair methodological quality, and 6 studies[20,32,35,38,49,55] had poor methodological quality. Four studies[20,23,33,48] that reported on a measurement instrument's criterion validity (box H) had good methodological quality, and 2 studies[15,39] had poor methodological quality. The methodological quality of 1 study[6] that reported on responsiveness (box I) was poor (**TABLE 3**). Poor methodological ratings for studies evaluating criterion validity were mainly due to deployed reference standards that could not be considered as an adequate gold standard. Five studies[32,35,38,49,55] that intended to determine a measurement instrument's criterion validity were therefore considered as studies on hypotheses testing. In 5 studies on hypotheses testing for construct validity,[20,32,35,38,49] hypothesis formulation, including the magnitude and direction of the expected relations between the instrument under study and comparator instruments, was unclear or not reported. Four studies[6,38,48,49] had poor data reporting, hindering the calculation and interpretation of relevant outcomes. Two studies[34,35] did not adequately report the measurement properties of the comparator instruments. In 1 study on responsiveness,[6] no comparator instrument was deployed to reflect the construct to be measured.

Grading quality of evidence for the overall quality of the instruments' measurement properties yielded moderate-quality evidence for sufficient and insufficient criterion validity, at most (**TABLE 4**). Results of all quality-of-evidence gradings are presented in **APPENDIX B TABLE 1** (available at www.jospt.org).

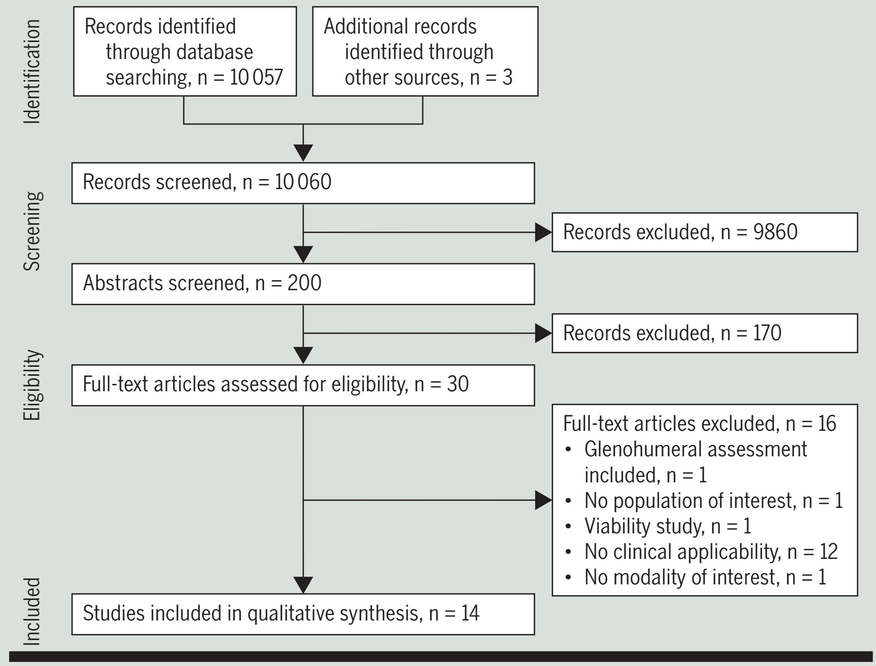


FIGURE. PRISMA flow diagram of the literature search and study selection process, based on Moher et al.[28] The list of excluded studies can be found in **APPENDIX C** (available at www.jospt.org).

TABLE 3 METHODOLOGICAL QUALITY RATINGS BY MEASUREMENT PROPERTY, USING THE COSMIN CHECKLIST

	Quality Rating[a]		
Measurement Property/ Instrument Category[b]	**Good**	**Fair**	**Poor**
Box F: hypotheses testing for construct validity			
1		Tucker and Ingram[54]	Johnson et al[20] O'Shea et al[35c]
3		Nijs et al[34]	Morrissey et al[32c] Park et al[38c] Tate et al[49c] Uhl et al[55c]
Box H: criterion validity			
1	Johnson et al[20] Sobush et al[48]		Greenfield et al[15] Peterson et al[39]
2	Nadeau et al[33]		
3	Koslow et al[23]		
Box I: responsiveness			
3			Crotty and Smith[6]

Abbreviation: COSMIN, COnsensus-based Standards for the selection of health Measurement INstruments.
[a]*Details on study characteristics and methodological quality ratings of included studies by measurement property can be found in* **APPENDIX B TABLE 1**. *No study was rated as "excellent."*
[b]*1, Instruments to measure scapular posture; 2, Instruments to measure scapular movement; 3, Instruments to assess scapular dysfunction.*
[c]*The study intended to determine criterion validity but received a poor score on item 4 ("the criterion used cannot be considered an [adequate] 'gold standard'") of the COSMIN checklist, box H. The study was therefore considered as a study on hypotheses testing for construct validity.*

Instruments to Measure Scapular Posture

Twenty different measurement instruments were identified in 6 studies.[15,20,35,39,48,54] The instruments generally involved measurements of distances between palpated scapular and thoracic bony landmarks. Their validity was established by referencing the instruments against (1) X-ray images,[15,39,48] (2) positional data derived from either 2-D camera footage[35] or 3-D motion-analysis systems,[20] and (3) simple measurement devices (eg, inclinometers)[54] (**APPENDIX B TABLE 2**). There was moderate-quality evidence for sufficient criterion validity of instruments to determine scapular protraction/retraction posture using a scoliometer,[48] and scapular rotation angles in various thoracohumeral elevation angles using a digital inclinometer.[20] There was moderate-quality evidence for insufficient criterion validity of instruments to determine scapular asymmetry.[48] There was low-quality evidence for sufficient hypotheses testing for construct validity of measurement instruments to determine scapular rotation angles in various thoracohumeral elevation angles using a 2-D inclinometer.[54]

Instruments to Measure Scapular Movement

Active scapular elevation and retraction range of motion measured with a goniometer and a tape measure were referenced against positional data derived from a 3-D motion-analysis system.[33] There was moderate-quality evidence for

TABLE 4 MEASUREMENT INSTRUMENTS WITH MODERATE-QUALITY EVIDENCE FOR SUFFICIENT AND INSUFFICIENT CRITERION VALIDITY BY CATEGORY[a]

Sufficient: Measurement/Study/Method and/or Material	Sufficient: Reference Standard	Insufficient: Measurement/Study/Method and/or Material	Insufficient: Reference Standard
Category 1: Measurement of Scapular Posture			
Protraction/retraction posture			
Sobush et al[48]			
Thoracic spine-to-inferior angle of scapula distance (Lennie test) with Scoliometer (Red Bank, NJ)	Measurement of distances between identical landmarks displayed on anteroposterior-view X-rays		
Thoracic spine-to-root of scapula distance (Lennie test) with Scoliometer (Red Rank, NJ)			
Scapular upward rotation with arms in rest position			
Johnson et al[20]		Sobush et al[48]	
Digital protractor (inclinometer)	3-D kinematic data derived from an electromagnetic tracking device (Fastrak 3; Polhemus, Colchester, VT)	Distance between thoracic midline and superior angle of scapula and distance between thoracic midline and inferior angle of scapula, and length of medial border of scapula (Lennie test)	Measurement of distances between identical landmarks displayed on anteroposterior-view X-rays
Scapular upward rotation at 60° thoracohumeral elevation			
Johnson et al[20]			
Pro 360 digital protractor (inclinometer)	3-D positional data derived from an electromagnetic tracking device (Fastrak 3, Polhemus, Colchester, VT)		
Scapular upward rotation at 90° thoracohumeral elevation			
Johnson et al[20]			
Pro 360 digital protractor (inclinometer)	3-D positional data derived from an electromagnetic tracking device (Fastrak 3, Polhemus, Colchester, VT)		
Scapular upward rotation at 120° thoracohumeral elevation			
Johnson et al[20]			
Pro 360 digital protractor (inclinometer)	3-D positional data derived from an electromagnetic tracking device (Fastrak 3, Polhemus, Colchester, VT)		

Table continues on page 637.

insufficient criterion validity of all 3 instruments (**APPENDIX B TABLE 3**).

Instruments to Assess Scapular Dyskinesis

Seven studies[6,23,32,34,38,49,55] established the validity of 8 different types of measurement instruments used to assess scapular movement and positioning as dyskinesis (**APPENDIX B TABLE 4**). The scapular dyskinesis test[49] and the 2-type and 4-type scapular dyskinesis observational methods[55] were referenced against kinematic data obtained by a 3-D electromagnetic tracking system. In addition, the scapular dyskinesis test was referenced against the pain subscale of the Penn Shoulder Score,[25] whereas a 3-D wing computed tomography scan method[38] to analyze 5 specific scapular angles was referenced against the 4-type dyskinesis observational method. There was no evidence for either criterion validity or hypotheses testing for construct validity of all 3 instruments.

The lateral scapular slide test[23,34] was referenced against an asymptomatic population, using the Shoulder Disability Questionnaire (SDQ)[57] and a visual analog scale (VAS).[14] There was moderate-quality evidence for insufficient criterion validity and low-quality evidence for insufficient hypotheses testing for construct validity. The medial rotation test[32] was referenced against 3-D ultrasound measurements. There was no evidence for hypotheses testing for construct validity of the medial rotation test. Measurements of acromion-to-wall and scapula-to-spine distances to determine scapular malpo-

TABLE 4 MEASUREMENT INSTRUMENTS WITH MODERATE-QUALITY EVIDENCE FOR SUFFICIENT AND INSUFFICIENT CRITERION VALIDITY BY CATEGORY[a] (CONTINUED)

Sufficient: Measurement/Study/Method and/or Material	Sufficient: Reference Standard	Insufficient: Measurement/Study/Method and/or Material	Insufficient: Reference Standard
Category 2: Measurement of Scapular Movement			
Active range of motion scapular elevation			
		Nadeau et al[33]	
		Linear distance from neutral position to end range with (1) tape measure and (2) goniometer	3-D kinematic data derived from a motion-analysis system (Optotrak 3020; Northern Digital Inc, Waterloo, Canada)
Active range of motion scapular retraction			
		Nadeau et al[33]	
		Linear distance from neutral position to end range with tape measure	3-D kinematic data derived from a motion-analysis system (Optotrak 3020; Northern Digital Inc., Waterloo, Canada)
Category 3: Assessment of Scapular Dyskinesis			
Lateral scapular slide test			
		Koslow et al[23]	
		Measurement of side-to-side difference in linear distance between inferior angle of scapula and closest spinous process, with participant's arms dangling relaxed at the sides (position 1), with tape measure	Symmetrical scapular positioning (<1.5-cm side-to-side difference) in asymptomatic athletes[21]
		Measurement of side-to-side difference in linear distance between inferior angle of scapula and closest spinous process, with participant's hands placed on the hips, fingers directed anteriorly, and thumbs directed posteriorly (position 2), with tape measure	
		Measurement of side-to-side difference in linear distance between inferior angle of scapula and closest spinous process, with participant's arms positioned in 90° of elevation and maximal internal rotation of the glenohumeral joint (position 3), with tape measure	
		Multitest regimen of positions 1, 2, and 3	

[a]*Results of all quality-of-evidence gradings are presented in* **APPENDIX B TABLES 2** *through* **5**.

sitioning[34] were referenced against the SDQ and a VAS for pain. There was low-quality evidence for insufficient hypotheses testing for construct validity for both measurement instruments.

One study[6] established the responsiveness of the DiVeta technique[9] and the Kibler technique[21] to assess scapular function based on changes in scapular resting posture. There was no evidence for the responsiveness of both instruments (**APPENDIX B TABLE 5**).

DISCUSSION

The aim of this systematic review was to evaluate the validity of measurement instruments to clinically assess scapular function. The quality of evidence for the validity of 31 instruments, grouped in the 3 main categories derived from 14 studies, was mainly low. Methodological limitations (ie, poor data reporting; poor or lacking formulation of hypotheses; inaccurate use and description of constructs, reference standards, and comparator instruments) hampered the synthesis and interpretation of the available evidence. Moderate-quality evidence for sufficient criterion validity suggests clinical potential for instruments to measure scapular positioning[48] and rotation angles[20] (category 1). The potential of instruments to assess scapular dyskinesis (category 3) is highly questionable.

Interpretation of Scapular Movement and Posture

Determining whether scapular movement or posture contributes to the persistence of shoulder complaints appears to be challenging with current clinical measurement instruments. Although there is moderate evidence for sufficient criterion validity for some instruments measuring scapular positioning,[20,48] the implication of their outcomes is insignificant if it is not clear how to interpret scapular positioning in terms of its relation to symptoms. The same applies to instruments measuring scapular movement (category 2). Measurement instruments to assess scapular dyskinesis (category 3), on the other hand, involve an interpretation of scapular movement or posture characteristics, assuming that the observed characteristics are directly related to the presence of symptoms. However, the concept of scapular dyskinesis is under debate. Several authors suggest that it is likely to determine a movement or position as dysfunctional only when it is shown that the movement or position itself provokes symptoms and negatively affects successful attainment of a goal, within a context that is relevant to the individual.[27,58] In 6 studies,[6,23,34,38,49,55] scapular movement and posture were observed in symptomatic and in asymptomatic individuals during a standardized set of arm movements, and were considered dyskinesis when they were asymmetrical or when they deviated from a predefined ideal posture or movement pattern. None of the measurement instruments, however, involved some sort of symptom alteration or context. In 2 studies,[49,55] measurement instruments used to assess scapular dyskinesis were referenced against 3-D kinematic data. We question the reference standards used to validate the measurement instruments and argue that the concept of scapular movement causing shoulder complaints is inaccurately operationalized. First, kinematic data objectively state a measure of movement or position; they do not reflect the construct of dyskinesis. Whether or not a movement is dysfunctional is a matter of interpretation.[17,18,58] Second, there is considerable variation in bilateral scapular resting positions as well as in scapular movement, in both healthy and symptomatic individuals.[19,26,36,40,47] There also seems to be no consistent scapular posture or movement that correlates with shoulder symptoms.[42] Third, a standardized generic movement task that is performed within a research or clinic environment may not evoke the same movement patterns or perceptions of pain that occur in daily life.[1] Thus, it is difficult, if not impossible, to indicate what "normal" and "aberrant" scapular movements or postures are, let alone to clinically assess whether an aberrant movement or posture is dyskinesis.

Comparison With Other Reviews

Two previous systematic reviews[12,24] evaluated the validity of clinical measurement instruments to assess scapular function and concluded that both observational evaluation systems and assessments of scapular upward rotation are suitably evidence based to be recommended for clinical determination of scapular dyskinesis.

In contrast, our systematic review questions the use of the previously advocated measurement instruments, and shows that the quality of current evidence is insufficient to recommend any specific measurement instrument. The discrepancy between reviews is most likely the result of the different methods we used to analyze the validity of measurement instruments. We have counteracted the difficulties of heterogeneous descriptions and inconsistent protocols of measurement instruments with the same name (eg, lateral scapular slide test) across studies by analyzing what was measured, ignoring the names of the measurement instruments. This provides clinicians with a more accurate insight and useful overview of the available evidence.

Limitations

Because studies with positive results are more likely to be published than negative or null results,[11] our systematic review is at risk of publication bias. However, as the outcome of our review indicates a lack of validity, additional articles with negative results would likely have strengthened our conclusions. We refrained from contacting authors to obtain additional data, as the impact of this proportionally extensive effort on the outcome is relatively small.[46]

Recently, a new COSMIN guideline for systematic reviews was published.[41] As this guideline was still under development when our systematic review started, we used its precursor. Different from the new guideline, the precursor does not explicitly deploy a modified Grading of Recommendations Assessment, Development and Evaluation approach to rate the quality of evidence. Nevertheless, it does consider imprecision and inconsistency when rating the quality of evidence (**TABLE 2**) and recommends a risk-of-bias assessment using the COSMIN checklist to rate a study's methodological quality. The new guideline, however, uses a newly developed risk-of-bias checklist[29] containing fewer items, and refers in particular to design requirements and preferred statistical methods of studies on measurement properties. Using the new COSMIN checklist would have affected the methodological rating of 3 studies,[20,38,49] resulting in a less negative score for risk of bias. This would not have affected our conclusion.

The criteria used to assess the adequacy of a measurement property (**TABLE 1**) are similar to those presented in the new COSMIN guideline. However, these criteria are arbitrary to a certain extent. For example, if boundaries of a confidence interval are not presented in a study, or exceed those of the assessment criteria, then the correlation remains uncertain.

Clinical Implications

Current measurement instruments aimed at assessing scapular dyskinesis should not be used, as they are prone to misinterpretation. Future research to establish a robust theoretical framework on how to interpret scapular movement in individuals with shoulder complaints is warranted.[17,27]

CONCLUSION

Clinical examination of scapular function is based on questionable evidence. Given the low quality of evidence for instruments measuring scapular posture and scapular movement and assessing scapular dyskinesis, it is not possible to recommend any specific measurement instrument. ◉

KEY POINTS

FINDINGS: There is insufficient evidence to recommend any specific measurement instrument for clinical examination of scapular function. The concept of scapular movement causing shoulder complaints is inadequately operationalized in measurement instruments to assess scapular dyskinesis.

IMPLICATIONS: Measurement instruments to assess scapular dyskinesis are prone to misinterpretation of scapular movement and should not be used as such. Future research to establish a robust theoretical framework on how to interpret scapular movement in individuals with shoulder complaints is warranted.

CAUTION: Due to the limited number of validity studies conducted, our review does not cover all types of measurement instruments to assess scapular function used in clinical practice. For instance, symptom alteration tests are not included in this review, as studies on their validity are lacking.

STUDY DETAILS

AUTHOR CONTRIBUTIONS: Norman E. D'hondt and Drs Terwee and Veeger contributed to the conception or design of the work. All authors contributed to the analysis and interpretation of data. Norman E. D'hondt drafted the manuscript. All authors substantially contributed to revising the manuscript critically for important intellectual content and gave final approval of the version to be published. All authors agreed to be accountable for all aspects of the work to ensure that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

DATA SHARING: All data relevant to the study are included in the article or within the online appendices.

PATIENT AND PUBLIC INVOLVEMENT: There was no patient or public involvement in this research.

ACKNOWLEDGMENTS: *We thank Liedeke van Schoot, research librarian, for her help with the construction and execution of the search strings. We also thank Sijmen Hacquebord for his contribution to the study selection process. Finally, we thank Dr Kylie Tucker for the language editing of the manuscript.*

REFERENCES

1. Babič J, Oztop E, Kawato M. Human motor adaptation in whole body motion. *Sci Rep*. 2016;6:32868. https://doi.org/10.1038/srep32868
2. Bartels B, de Groot JF, Terwee CB. The six-minute walk test in chronic pediatric conditions: a systematic review of measurement properties. *Phys Ther*. 2013;93:529-541. https://doi.org/10.2522/ptj.20120210
3. Burn MB, McCulloch PC, Lintner DM, Liberman SR, Harris JD. Prevalence of scapular dyskinesis in overhead and nonoverhead athletes: a systematic review. *Orthop J Sports Med*. 2016;4:2325967115627608. https://doi.org/10.1177/2325967115627608
4. Bury J, West M, Chamorro-Moriana G, Littlewood C. Effectiveness of scapula-focused approaches in patients with rotator cuff related shoulder pain: a systematic review and meta-analysis. *Man Ther*. 2016;25:35-42. https://doi.org/10.1016/j.math.2016.05.337
5. Cools AM, Struyf F, De Mey K, Maenhout A, Castelein B, Cagnie B. Rehabilitation of scapular dyskinesis: from the office worker to the elite overhead athlete. *Br J Sports Med*. 2014;48:692-697. https://doi.org/10.1136/bjsports-2013-092148
6. Crotty NM, Smith J. Alterations in scapular position with fatigue: a study in swimmers. *Clin J Sport Med*. 2000;10:251-258. https://doi.org/10.1097/00042752-200010000-00005
7. De Mey K, Danneels LA, Cagnie B, Huyghe L, Seyns E, Cools AM. Conscious correction of scapular orientation in overhead athletes performing selected shoulder rehabilitation exercises: the effect on trapezius muscle activation measured by surface electromyography. *J Orthop Sports Phys Ther*. 2013;43:3-10. https://doi.org/10.2519/jospt.2013.4283
8. D'hondt NE, Kiers H, Pool JJM, Hacquebord ST, Terwee CB, Veeger D. Reliability of performance-based clinical measurements to assess shoulder girdle kinematics and positioning: systematic review. *Phys Ther*. 2017;97:124-144. https://doi.org/10.2522/ptj.20160088
9. DiVeta J, Walker ML, Skibinski B. Relationship between performance of selected scapular muscles and scapular abduction in standing subjects. *Phys Ther*. 1990;70:470-476; discussion 476-479. https://doi.org/10.1093/ptj/70.8.470
10. Dobson F, Hinman RS, Hall M, Terwee CB, Roos EM, Bennell KL. Measurement properties of performance-based measures to assess physical function in hip and knee osteoarthritis: a

systematic review. *Osteoarthritis Cartilage*. 2012;20:1548-1562. https://doi.org/10.1016/j.joca.2012.08.015

11. Dwan K, Altman DG, Arnaiz JA, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS One*. 2008;3:e3081. https://doi.org/10.1371/journal.pone.0003081

12. Fisher MI, Levangie PK. Oncology Section Task Force on Breast Cancer Outcomes: scapular assessment. *Rehabil Oncol*. 2013;31:11-18. https://doi.org/10.1097/01893697-201331010-00004

13. Gagnier JJ, Morgenstern H, Altman DG, et al. Consensus-based recommendations for investigating clinical heterogeneity in systematic reviews. *BMC Med Res Methodol*. 2013;13:106. https://doi.org/10.1186/1471-2288-13-106

14. Good M, Stiller C, Zauszniewski JA, Anderson GC, Stanton-Hicks M, Grass JA. Sensation and distress of pain scales: reliability, validity, and sensitivity. *J Nurs Meas*. 2001;9:219-238. https://doi.org/10.1891/1061-3749.9.3.219

15. Greenfield B, Catlin PA, Coats PW, Green E, McDonald JJ, North C. Posture in patients with shoulder overuse injuries and healthy individuals. *J Orthop Sports Phys Ther*. 1995;21:287-295. https://doi.org/10.2519/jospt.1995.21.5.287

16. Hickey D, Solvig V, Cavalheri V, Harrold M, McKenna L. Scapular dyskinesis increases the risk of future shoulder pain by 43% in asymptomatic athletes: a systematic review and meta-analysis. *Br J Sports Med*. 2018;52:102-110. https://doi.org/10.1136/bjsports-2017-097559

17. Hodges PW, Smeets RJ. Interaction between pain, movement, and physical activity: short-term benefits, long-term consequences, and targets for treatment. *Clin J Pain*. 2015;31:97-107. https://doi.org/10.1097/AJP.0000000000000098

18. Hodges PW, Tucker K. Moving differently in pain: a new theory to explain the adaptation to pain. *Pain*. 2011;152:S90-S98. https://doi.org/10.1016/j.pain.2010.10.020

19. Hosseinimehr SH, Anbarian M, Norasteh AA, Fardmal J, Khosravi MT. The comparison of scapular upward rotation and scapulohumeral rhythm between dominant and non-dominant shoulder in male overhead athletes and non-athletes. *Man Ther*. 2015;20:758-762. https://doi.org/10.1016/j.math.2015.02.010

20. Johnson MP, McClure PW, Karduna AR. New method to assess scapular upward rotation in subjects with shoulder pathology. *J Orthop Sports Phys Ther*. 2001;31:81-89. https://doi.org/10.2519/jospt.2001.31.2.81

21. Kibler WB. The role of the scapula in athletic shoulder function. *Am J Sports Med*. 1998;26:325-337. https://doi.org/10.1177/03635465980260022801

22. Kibler WB, Ludewig PM, McClure PW, Michener LA, Bak K, Sciascia AD. Clinical implications of scapular dyskinesis in shoulder injury: the 2013 consensus statement from the 'Scapular Summit'. *Br J Sports Med*. 2013;47:877-885. https://doi.org/10.1136/bjsports-2013-092425

23. Koslow PA, Prosser LA, Strony GA, Suchecki SL, Mattingly GE. Specificity of the lateral scapular slide test in asymptomatic competitive athletes. *J Orthop Sports Phys Ther*. 2003;33:331-336. https://doi.org/10.2519/jospt.2003.33.6.331

24. Larsen CM, Juul-Kristensen B, Lund H, Søgaard K. Measurement properties of existing clinical assessment methods evaluating scapular positioning and function. A systematic review. *Physiother Theory Pract*. 2014;30:453-482. https://doi.org/10.3109/09593985.2014.899414

25. Leggin BG, Michener LA, Shaffer MA, Brenneman SK, Iannotti JP, Williams GR, Jr. The Penn Shoulder Score: reliability and validity. *J Orthop Sports Phys Ther*. 2006;36:138-151. https://doi.org/10.2519/jospt.2006.36.3.138

26. Ludewig PM, Reynolds JF. The association of scapular kinematics and glenohumeral joint pathologies. *J Orthop Sports Phys Ther*. 2009;39:90-104. https://doi.org/10.2519/jospt.2009.2808

27. McQuade KJ, Borstad J, de Oliveira AS. Critical and theoretical perspective on scapular stabilization: what does it really mean, and are we on the right track? *Phys Ther*. 2016;96:1162-1169. https://doi.org/10.2522/ptj.20140230

28. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol*. 2009;62:1006-1012. https://doi.org/10.1016/j.jclinepi.2009.06.005

29. Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27:1171-1179. https://doi.org/10.1007/s11136-017-1765-4

30. 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

31. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol*. 2010;63:737-745. https://doi.org/10.1016/j.jclinepi.2010.02.006

32. Morrissey D, Morrissey MC, Driver W, King JB, Woledge RC. Manual landmark identification and tracking during the medial rotation test of the shoulder: an accuracy study using three-dimensional ultrasound and motion analysis measures. *Man Ther*. 2008;13:529-535. https://doi.org/10.1016/j.math.2007.07.009

33. Nadeau S, Kovacs S, Gravel D, et al. Active movement measurements of the shoulder girdle in healthy subjects with goniometer and tape measure techniques: a study on reliability and validity. *Physiother Theory Pract*. 2007;23:179-187. https://doi.org/10.1080/09593980701209246

34. Nijs J, Roussel N, Vermeulen K, Souvereyns G. Scapular positioning in patients with shoulder pain: a study examining the reliability and clinical importance of 3 clinical tests. *Arch Phys Med Rehabil*. 2005;86:1349-1355. https://doi.org/10.1016/j.apmr.2005.03.021

35. O'Shea A, Kelly R, Williams S, McKenna L. Reliability and validity of the measurement of scapular position using the protractor method. *Phys Ther*. 2016;96:502-510. https://doi.org/10.2522/ptj.20150144

36. Oyama S, Myers JB, Wassinger CA, Ricci RD, Lephart SM. Asymmetric resting scapular posture in healthy overhead athletes. *J Athl Train*. 2008;43:565-570. https://doi.org/10.4085/1062-6050-43.6.565

37. Page MJ, Shamseer L, Tricco AC. Registration of systematic reviews in PROSPERO: 30,000 records and counting. *Syst Rev*. 2018;7:32. https://doi.org/10.1186/s13643-018-0699-4

38. Park JY, Hwang JT, Kim KM, Makkar D, Moon SG, Han KJ. How to assess scapular dyskinesis precisely: 3-dimensional wing computer tomography-a new diagnostic modality. *J Shoulder Elbow Surg*. 2013;22:1084-1091. https://doi.org/10.1016/j.jse.2012.10.046

39. Peterson DE, Blankenship KR, Robb JB, et al. Investigation of the validity and reliability of four objective techniques for measuring forward shoulder posture. *J Orthop Sports Phys Ther*. 1997;25:34-42. https://doi.org/10.2519/jospt.1997.25.1.34

40. Plummer HA, Sum JC, Pozzi F, Varghese R, Michener LA. Observational scapular dyskinesis: known-groups validity in patients with and without shoulder pain. *J Orthop Sports Phys Ther*. 2017;47:530-537. https://doi.org/10.2519/jospt.2017.7268

41. 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

42. Ratcliffe E, Pickering S, McLean S, Lewis J. Is there a relationship between subacromial impingement syndrome and scapular orientation? A systematic review. *Br J Sports Med*. 2014;48:1251-1256. https://doi.org/10.1136/bjsports-2013-092389

43. Reijneveld EA, Noten S, Michener LA, Cools A, Struyf F. Clinical outcomes of a scapular-focused treatment in patients with subacromial pain syndrome: a systematic review. *Br J Sports Med*. 2017;51:436-441. https://doi.org/10.1136/bjsports-2015-095460

44. 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

45. 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

46. Selph SS, Ginsburg AD, Chou R. Impact of contacting study authors to obtain additional data for systematic reviews: diagnostic accuracy studies for hepatic fibrosis. *Syst Rev*. 2014;3:107. https://doi.org/10.1186/2046-4053-3-107
47. Shimpi AP, Bhakti S, Roshni K, Rairikar SA, Shyam A, Sancheti PK. Scapular resting position and gleno-humeral movement dysfunction in asymptomatic racquet players: a case-control study. *Asian J Sports Med*. 2015;6:e24053. https://doi.org/10.5812/asjsm.24053
48. Sobush DC, Simoneau GG, Dietz KE, Levene JA, Grossman RE, Smith WB. The Lennie test for measuring scapular position in healthy young adult females: a reliability and validity study. *J Orthop Sports Phys Ther*. 1996;23:39-50. https://doi.org/10.2519/jospt.1996.23.1.39
49. Tate AR, McClure P, Kareha S, Irwin D, Barbe MF. A clinical method for identifying scapular dyskinesis, part 2: validity. *J Athl Train*. 2009;44:165-173. https://doi.org/10.4085/1062-6050-44.2.165
50. 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
51. Terwee CB, Bouwmeester W, van Elsland SL, de Vet HC, Dekker J. Instruments to assess physical activity in patients with osteoarthritis of the hip or knee: a systematic review of measurement properties. *Osteoarthritis Cartilage*. 2011;19:620-633. https://doi.org/10.1016/j.joca.2011.01.002
52. Terwee CB, Jansma EP, Riphagen II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res*. 2009;18:1115-1123. https://doi.org/10.1007/s11136-009-9528-5
53. Timmons MK, Thigpen CA, Seitz AL, Karduna AR, Arnold BL, Michener LA. Scapular kinematics and subacromial-impingement syndrome: a meta-analysis. *J Sport Rehabil*. 2012;21:354-370. https://doi.org/10.1123/jsr.21.4.354
54. Tucker WS, Ingram RL. Reliability and validity of measuring scapular upward rotation using an electrical inclinometer. *J Electromyogr Kinesiol*. 2012;22:419-423. https://doi.org/10.1016/j.jelekin.2012.02.013
55. Uhl TL, Kibler WB, Gecewich B, Tripp BL. Evaluation of clinical assessment methods for scapular dyskinesis. *Arthroscopy*. 2009;25:1240-1248. https://doi.org/10.1016/j.arthro.2009.06.007
56. van der Windt DA, Thomas E, Pope DP, et al. Occupational risk factors for shoulder pain: a systematic review. *Occup Environ Med*. 2000;57:433-442. https://doi.org/10.1136/oem.57.7.433
57. van der Windt DA, van der Heijden GJ, de Winter AF, Koes BW, Deville W, Bouter LM. The responsiveness of the Shoulder Disability Questionnaire. *Ann Rheum Dis*. 1998;57:82-87. https://doi.org/10.1136/ard.57.2.82
58. Willmore EG, Smith MJ. Scapular dyskinesia: evolution towards a systems-based approach. *Shoulder Elbow*. 2016;8:61-70. https://doi.org/10.1177/1758573215618857

APPENDIX A

SEARCH STRINGS

PubMed (March 30, 2019)

	Search
#1. Construct of interest	((((((((((((((((((((((((((((motion[Title/Abstract]) OR motions[Title/Abstract]) OR "Articular range of motion"[Title/Abstract]) OR "joint range of motion"[Title/Abstract]) OR "joint flexibility"[Title/Abstract]) OR "range of motion"[Title/Abstract]) OR "passive range of motion"[Title/Abstract]) OR "range of movement"[Title/Abstract]) OR "articular arthrometry"[Title/Abstract]) OR "articular goniometry"[Title/Abstract]) OR movement[Title/Abstract]) OR movements[Title/Abstract]) OR kinesis[Title/Abstract]) OR kineses[Title/Abstract]) OR kinematics[Title/Abstract]) OR "scapular kinematics"[Title/Abstract]) OR "scapulothoracic kinematics"[Title/Abstract]) OR dyskinesis[Title/Abstract]) OR dyskinesis[Title/Abstract]) OR "abnormal movement"[Title/Abstract]) OR "abnormal movements"[Title/Abstract]) OR dyskinesis[Title/Abstract]) OR "psychomotor performance"[Title/Abstract]) OR "psychomotor performances"[Title/Abstract]) OR "visual motor coordination"[Title/Abstract]) OR "visual motor coordinations"[Title/Abstract]) OR "perceptual motor performance"[Title/Abstract]) OR "perceptual motor performances"[Title/Abstract]) OR "sensory motor performance"[Title/Abstract]) OR "sensory motor performances"[Title/Abstract])) OR (((((("motion"[MeSH Terms] OR "range of motion, articular"[MeSH Terms]) OR "arthrometry, articular"[MeSH Terms]) OR "movement"[MeSH Terms]) OR "kinesis"[MeSH Terms]) OR "dyskinesis"[MeSH Terms]) OR "psychomotor performance"[MeSH Terms])))
#2. Target population	((((("shoulder"[MeSH] OR "shoulder joint"[MeSH] OR "scapula"[MeSH] OR "glenoid cavity"[MeSH] OR "clavicle"[MeSH] OR "sternoclavicular joint"[MeSH] OR "thorax"[MeSH] OR "cervical vertebrae"[MeSH] OR "shoulder"[Tiab] OR "shoulders"[Tiab] OR "shoulder girdle"[Tiab] OR "shoulder joint"[Tiab] OR "shoulder joints"[Tiab] OR "glenohumeral joint"[Tiab] OR "glenohumeral joints"[Tiab] OR "scapulothoracic"[Tiab] OR "scapulohumeral"[Tiab] OR "scapula"[Tiab] OR "glenoid cavity"[Tiab] OR "glenoid cavities"[Tiab] OR "glenoid fossa of the scapula"[Tiab] OR "glenohumeral"[Tiab] OR "clavicle"[Tiab] OR "sternoclavicular joint"[Tiab] OR "sternoclavicular joints"[Tiab] OR "thorax"[Tiab] OR "thoraces"[Tiab] OR "chest"[Tiab] OR "chests"[Tiab] OR "rotator cuff"[Tiab] OR "rotator cuffs"[Tiab] OR "biceps"[Tiab] OR "cervical vertebrae"[Tiab] OR "cervicothoracic"[Tiab] OR "cervicothoracic spine"[Tiab]))) AND (("shoulder pain"[MeSH] OR "shoulder impingement syndrome"[MeSH] OR "shoulder fractures"[MeSH] OR "shoulder dislocation"[MeSH] OR "brachial plexus neuritis"[MeSH] OR "bursitis"[MeSH] OR "tendinopathy"[MeSH] OR "shoulder pain"[Tiab] OR "shoulder pains"[Tiab] OR "shoulder impingement syndrome"[Tiab] OR "shoulder impingement syndromes"[Tiab] OR "subacromial impingement syndrome"[Tiab] OR "subacromial impingement syndromes"[Tiab] OR "shoulder fracture"[Tiab] OR "shoulder fractures"[Tiab] OR "proximal humeral fracture"[Tiab] OR "proximal humeral fractures"[Tiab] OR "shoulder dislocation"[Tiab] OR "shoulder dislocations"[Tiab] OR "glenohumeral dislocation"[Tiab] OR "glenohumeral dislocations"[Tiab] OR "glenohumeral subluxation"[Tiab] OR "glenohumeral subluxations"[Tiab] OR "glenohumeral instability"[Tiab] OR "brachial plexus neuritis"[Tiab] OR "brachial plexus neuritides"[Tiab] OR "shoulder-girdle neuropathy"[Tiab] OR "brachial neuritis"[Tiab] OR "brachial neuritides"[Tiab] OR "parsonage-aldren-turner syndrome"[Tiab] OR "parsonage-turner syndrome"[Tiab] OR "parsonage-aldren-turner syndromes"[Tiab] OR "parsonage-turner syndromes"[Tiab] OR "cervicobrachial neuralgia"[Tiab] OR "cervicobrachial neuralgias"[Tiab] OR "brachial neuralgia"[Tiab] OR "brachial neuralgias"[Tiab] OR "bursitides adhesive"[Tiab] OR "capsulitis adhesive"[Tiab] OR "capsulitides frozen shoulder"[Tiab] OR "capsulitides frozen shoulders"[Tiab] OR "sprengel deformity"[Tiab] OR "sprengel deformities"[Tiab] OR "maladie de sprengel"[Tiab] OR "familiale sprengel's shoulder"[Tiab] OR "familiale sprengel's shoulders"[Tiab] OR "high scapula"[Tiab] OR "bursitis"[Tiab] OR "scapular dyskinesis"[Tiab] OR "pseudoarthrosis clavicle"[Tiab] OR "congenital clavicle"[Tiab] OR "tendinopathy"[Tiab] OR "tendinopathies"[Tiab] OR "tendinitis"[Tiab] OR "tendonitis"[Tiab] OR "tendonitides"[Tiab] OR "tendinosis"[Tiab] OR "tendinoses"[Tiab] OR "Shoulder injuries"[Tiab] OR "Shoulder injury"[Tiab])))) OR (("shoulder"[MeSH] OR "shoulder joint"[MeSH] OR "scapula"[MeSH] OR "glenoid cavity"[MeSH] OR "clavicle"[MeSH] OR "sternoclavicular joint"[MeSH] OR "thorax"[MeSH] OR "cervical vertebrae"[MeSH] OR "shoulder"[Tiab] OR "shoulders"[Tiab] OR "shoulder girdle"[Tiab] OR "shoulder joint"[Tiab] OR "shoulder joints"[Tiab] OR "glenohumeral joint"[Tiab] OR "glenohumeral joints"[Tiab] OR "scapulothoracic"[Tiab] OR "scapulohumeral"[Tiab] OR "scapula"[Tiab] OR "glenoid cavity"[Tiab] OR "glenoid cavities"[Tiab] OR "glenoid fossa of the scapula"[Tiab] OR "glenohumeral"[Tiab] OR "clavicle"[Tiab] OR "sternoclavicular joint"[Tiab] OR "sternoclavicular joints"[Tiab] OR "thorax"[Tiab] OR "thoraces"[Tiab] OR "chest"[Tiab] OR "chests"[Tiab] OR "rotator cuff"[Tiab] OR "rotator cuffs"[Tiab] OR "biceps"[Tiab] OR "cervical vertebrae"[Tiab] OR "cervicothoracic"[Tiab] OR "cervicothoracic spine"[Tiab]))
#3. Type of instrument	((((((((("physical examination"[MeSH Terms]) OR "diagnostic tests, routine"[MeSH Terms]) OR "anatomic landmarks"[MeSH Terms]) OR "delayed diagnosis"[MeSH Terms]) OR "diagnosis, differential"[MeSH Terms]) OR "diagnostic imaging"[MeSH Terms]) OR "diagnostic errors"[MeSH Terms])) OR (((((((((((((((((((("physical examination"[Title/Abstract]) OR "physical examinations"[Title/Abstract]) OR ("physical examinations and diagnoses"[Title/Abstract])) OR "routine diagnostic tests"[Title/Abstract]) OR "routine diagnostic test"[Title/Abstract]) OR "diagnostic test"[Title/Abstract]) OR "diagnostic tests"[Title/Abstract]) OR "anatomic landmarks"[Title/Abstract]) OR "anatomic landmark"[Title/Abstract]) OR "delayed diagnosis"[Title/Abstract]) OR "delayed diagnoses"[Title/Abstract]) OR "late diagnosis"[Title/Abstract]) OR "late diagnoses"[Title/Abstract]) OR "differential diagnosis"[Title/Abstract]) OR "differential diagnoses"[Title/Abstract]) OR "diagnostic errors"[Title/Abstract]) OR "diagnostic error"[Title/Abstract]) OR misdiagnosis[Title/Abstract]) OR misdiagnoses[Title/Abstract]) OR "diagnostic imaging"[Title/Abstract]) OR "medical imaging"[Title/Abstract])))))

Table continues on page E2.

APPENDIX A

	Search
#4. Measurement properties	instrumentation[sh] OR methods[sh] OR "Validation Studies"[pt] OR "Comparative Study"[pt] OR "psychometrics"[MeSH] OR psychometr*[tiab] OR clinimetr*[tw] OR clinometr*[tw] OR "outcome assessment (health care)"[MeSH] OR "outcome assessment"[tiab] OR "outcome measure*"[tw] OR "observer variation"[MeSH] OR "observer variation"[tiab] OR "Health Status Indicators"[Mesh] OR "reproducibility of results"[MeSH] OR reproducib*[tiab] OR "discriminant analysis"[MeSH] OR reliab*[tiab] OR unreliab*[tiab] OR valid*[tiab] OR "coefficient of variation"[tiab] OR coefficient[tiab] OR homogeneity[tiab] OR homogeneous[tiab] OR "internal consistency"[tiab] OR (cronbach*[tiab] AND (alpha[tiab] OR alphas[tiab])) OR (item[tiab] AND (correlation*[tiab] OR selection*[tiab] OR reduction*[tiab])) OR agreement[tw] OR precision[tw] OR imprecision[tw] OR "precise values"[tw] OR test-retest[tiab] OR (test[tiab] AND retest[tiab]) OR (reliab*[tiab] AND (test[tiab] OR retest[tiab])) OR stability[tiab] OR interrater[tiab] OR interrater[tiab] OR intrarater[tiab] OR intrarater[tiab] OR intertester[tiab] OR intertester[tiab] OR intratester[tiab] OR within-tester[tiab] OR interobserver[tiab] OR interobserver[tiab] OR intraobserver[tiab] OR intraobserver[tiab] OR intertechnician[tiab] OR intertechnician[tiab] OR intratechnician[tiab] OR intratechnician[tiab] OR interexaminer[tiab] OR interexaminer[tiab] OR intraexaminer[tiab] OR intraexaminer[tiab] OR interassay[tiab] OR interassay[tiab] OR intraassay[tiab] OR intraassay[tiab] OR interindividual[tiab] OR interindividual[tiab] OR intraindividual[tiab] OR intraindividual[tiab] OR interparticipant[tiab] OR interparticipant[tiab] OR intraparticipant[tiab] OR intraparticipant[tiab] OR kappa[tiab] OR kappa's[tiab] OR kappas[tiab] OR repeatab*[tw] OR ((replicab*[tw] OR repeated[tw]) AND (measure[tw] OR measures[tw] OR findings[tw] OR result[tw] OR results[tw] OR test[tw] OR tests[tw])) OR generaliza*[tiab] OR generalisa*[tiab] OR concordance[tiab] OR (intraclass[tiab] AND correlation*[tiab]) OR discriminative[tiab] OR "known group"[tiab] OR "factor analysis"[tiab] OR "factor analyses"[tiab] OR "factor structure"[tiab] OR "factor structures"[tiab] OR dimension*[tiab] OR subscale*[tiab] OR (multitrait[tiab] AND scaling[tiab] AND (analysis[tiab] OR analyses[tiab])) OR "item discriminant"[tiab] OR "interscale correlation*"[tiab] OR error[tiab] OR errors[tiab] OR "individual variability"[tiab] OR "interval variability"[tiab] OR "rate variability"[tiab] OR (variability[tiab] AND (analysis[tiab] OR values[tiab])) OR (uncertainty[tiab] AND (measurement[tiab] OR measuring[tiab])) OR "standard error of measurement"[tiab] OR sensitiv*[tiab] OR responsive*[tiab] OR (limit[tiab] AND detection[tiab]) OR "minimal detectable concentration"[tiab] OR interpretab*[tiab] OR ((minimal[tiab] OR minimally[tiab] OR clinical[tiab] OR clinically[tiab]) AND (important[tiab] OR significant[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR (small*[tiab] AND (real[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR "meaningful change"[tiab] OR "ceiling effect"[tiab] OR "floor effect"[tiab] OR "Item response model"[tiab] OR IRT[tiab] OR Rasch[tiab] OR "Differential item functioning"[tiab] OR DIF[tiab] OR "computer adaptive testing"[tiab] OR "item bank"[tiab] OR "cross-cultural equivalence"[tiab]
#5. Exclusion filter	("addresses"[Publication Type] OR "biography"[Publication Type] OR "case reports"[Publication Type] OR "comment"[Publication Type] OR "directory"[Publication Type] OR "editorial"[Publication Type] OR "festschrift"[Publication Type] OR "interview"[Publication Type] OR "lectures"[Publication Type] OR "legal cases"[Publication Type] OR "legislation"[Publication Type] OR "letter"[Publication Type] OR "news"[Publication Type] OR "newspaper article"[Publication Type] OR "patient education handout"[Publication Type] OR "popular works"[Publication Type] OR "congresses"[Publication Type] OR "consensus development conference"[Publication Type] OR "consensus development conference, nih"[Publication Type] OR "practice guideline"[Publication Type]) NOT ("animals"[MeSH Terms] NOT "humans"[MeSH Terms])
#6. Combining elements	#1 AND #2 AND #3 AND #4 NOT #5

APPENDIX A

CINAHL (EBSCOhost) (March 30, 2019)

	Search
#1. Construct of interest	TI "Motion" OR TI "motions" OR TI "articular range of motion" OR TI "joint range of motion" OR TI "joint flexibility" OR TI "range of motion" OR TI "range of movement" OR TI "passive range of motion" OR TI "articular arthrometry" OR TI "articular goniometry" OR TI "Movement" OR TI "Movements" OR TI "kinesis" OR TI "kineses" OR TI "scapular kinematics" OR TI "scapulothoracic kinematics" OR TI "Dyskinesis" OR TI "Dyskinesis" OR TI "Dyskinesis" OR TI "abnormal movement" OR TI "abnormal movements" OR TI "psychomotor performance" OR TI "psychomotor performances" OR TI "visual motor coordination" OR TI "visual motor coordinations" OR TI "perceptual motor performance" OR TI "perceptual motor performances" OR TI "sensory motor performance" OR TI "sensory motor performances" OR MH "Arthrometry" OR MH "Dyskinesis" OR MH "Kinesthesis" OR MH "Locomotion" OR MH "Motion" OR MH "Movement" OR MH "Psychomotor Performance" OR MH "Range of Motion" OR MH "Kinematics" OR TI "Motion" OR TI "motions" OR TI "articular range of motion" OR TI "joint range of motion" OR TI "joint flexibility" OR TI "range of motion" OR TI "range of movement" OR TI "passive range of motion" OR TI "articular arthrometry" OR TI "articular goniometry" OR TI "Movement" OR TI "Movements" OR TI "kinesis" OR TI "kineses" OR TI "scapular kinematics" OR TI "scapulothoracic kinematics" OR TI "Dyskinesis" OR TI "Dyskinesis" OR TI "Dyskinesis" OR TI "abnormal movement" OR TI "abnormal movements" OR TI "psychomotor performance" OR TI "psychomotor performances" OR TI "visual motor coordination" OR TI "visual motor coordinations" OR TI "perceptual motor performance" OR TI "perceptual motor performances" OR TI "sensory motor performance" OR TI "sensory motor performances" OR AB motion OR AB motions OR AB "articular range of motion" OR AB "joint range of motion" OR AB "joint flexibility" OR AB "range of motion" OR AB "range of movement" OR AB "passive range of motion" OR AB "articular arthrometry" OR AB "articular goniometry" OR AB Movement OR AB Movements OR AB kinesis OR AB kineses OR AB "scapular kinematics" OR AB "scapulothoracic kinematics" OR AB Dyskinesis OR AB Dyskinesis OR AB Dyskinesis OR AB "abnormal movement" OR AB "abnormal movements" OR AB "psychomotor performance" OR AB "psychomotor performances" OR AB "visual motor coordination" OR AB "visual motor coordinations" OR AB kineses OR AB "perceptual motor performance" OR AB "scapulothoracic kinematics" OR AB "perceptual motor performances" OR AB "sensory motor performance" OR AB "sensory motor performances"
#2. Target population	(MH shoulder+ OR MH scapula+ OR MH clavicle+ OR MH "sternoclavicular joint+" OR MH "shoulder joint+" OR MH "cervical vertebrae+" OR MH "biceps brachii muscles+" OR MH "rotator cuff+" OR TI "Glenoid cavity" OR TI Shoulder OR TI Shoulders OR TI "shoulder joint" OR TI "shoulder joint+" OR TI "shoulder joints" OR TI "glenohumeral joint" OR TI "glenohumeral joints" OR TI "glenoid cavities" OR TI "glenoid fossa of the scapula" OR TI "sternoclavicular joint" OR TI "sternoclavicular joints" OR TI Biceps OR TI "rotator cuff" OR TI "rotator cuffs" OR TI "cervical vertebrae" OR TI "cervicothoracic spine" OR TI Cervicothoracic OR AB "Glenoid cavity" OR AB Shoulder OR AB Shoulders OR AB "shoulder joint" OR AB "shoulder joints" OR AB "glenohumeral joint" OR AB "glenohumeral joints" OR AB "glenoid cavities" OR AB "glenoid fossa of the scapula" OR AB "sternoclavicular joint" OR AB "sternoclavicular joints" OR AB Biceps OR AB "rotator cuff" OR AB "rotator cuffs" OR AB "cervical vertebrae" OR AB "cervicothoracic spine" OR AB Cervicothoracic) OR ((MH shoulder+ OR MH scapula+ OR MH clavicle+ OR MH "sternoclavicular joint+" OR MH "shoulder joint+" OR MH "cervical vertebrae+" OR MH "biceps brachii muscles+" OR MH "rotator cuff+" OR TI "Glenoid cavity" OR TI Shoulder OR TI Shoulders OR TI "shoulder joint" OR TI "shoulder joint+" OR TI "shoulder joints" OR TI "glenohumeral joint" OR TI "glenohumeral joints" OR TI "glenoid cavities" OR TI "glenoid fossa of the scapula" OR TI "sternoclavicular joint" OR TI "sternoclavicular joints" OR TI Biceps OR TI "rotator cuff" OR TI "rotator cuffs" OR TI "cervical vertebrae" OR TI "cervicothoracic spine" OR TI Cervicothoracic OR AB "Glenoid cavity" OR AB Shoulder OR AB Shoulders OR AB "shoulder joint" OR AB "shoulder joints" OR AB "glenohumeral joint" OR AB "glenohumeral joints" OR AB "glenoid cavities" OR AB "glenoid fossa of the scapula" OR AB "sternoclavicular joint" OR AB "sternoclavicular joints" OR AB Biceps OR AB "rotator cuff" OR AB "rotator cuffs" OR AB "cervical vertebrae" OR AB "cervicothoracic spine" OR AB Cervicothoracic) AND(MH "Shoulder pain+" OR MH "Shoulder fractures+" OR MH "Shoulder impingement syndrome+" OR MH "Shoulder injuries+" OR MH "Shoulder dislocation+" OR MH "Brachial plexus neuritis+" OR MH "Bursitis+" OR MH "Tendinopathy+" OR TI "shoulder pain" OR TI "shoulder impingement syndrome" OR TI "shoulder fractures" OR TI "shoulder fracture" OR TI "shoulder dislocation" OR TI "shoulder injuries" OR TI "shoulder injury" OR TI Impingement OR TI "proximal humeral fracture" OR TI "proximal humeral fractures" OR TI "glenohumeral dislocation" OR TI "glenohumeral dislocations" OR TI "glenohumeral subluxation" OR TI "glenohumeral subluxations" OR TI "glenohumeral instability" OR TI "brachial plexus neuritis" OR TI "brachial plexus neuritides" OR TI "parsonage-aldren-turner-syndrome" OR TI "parsonage-aldren-turner-syndromes" OR TI "parsonage-turner-syndrome" OR TI "parsonage-turner-syndromes" OR TI "cervicobrachial neuralgia" OR TI "cervicobrachial neuralgias" OR TI "brachial neuralgia" OR TI "brachial neuralgias" OR TI "bursitides adhesive" OR TI "capsulitis adhesive" OR TI Sprengel OR TI Bursitis OR TI "scapular dyskinesis" OR TI Tendinopathy OR TI tendinitis OR TI Tendonitis OR TI tendonitides OR TI Tendinosis OR AB "shoulder pain" OR AB "shoulder impingement syndrome" OR AB "shoulder fractures" OR AB "shoulder fracture" OR AB "shoulder dislocation" OR AB "shoulder injuries" OR AB "shoulder injury" OR AB Impingement OR AB "proximal humeral fracture" OR AB "proximal humeral fractures" OR AB "glenohumeral dislocation" OR AB "glenohumeral dislocations" OR AB "glenohumeral subluxation" OR AB "glenohumeral subluxations" OR AB "glenohumeral instability" OR AB "brachial plexus neuritis" OR AB "brachial plexus neuritides" OR AB "parsonage-aldren-turner-syndrome" OR AB "parsonage-aldren-turner-syndromes" OR AB "parsonage-turner-syndrome" OR AB "parsonage-turner-syndromes" OR AB "cervicobrachial neuralgia" OR AB "cervicobrachial neuralgias" OR AB "brachial neuralgia" OR AB "brachial neuralgias" OR AB "bursitides adhesive" OR AB Sprengel OR AB Bursitis OR AB "scapular dyskinesis" OR AB Tendinopathy OR AB tendinitis OR AB Tendonitis OR AB tendonitides OR AB Tendinosis))

Table continues on page E4.

APPENDIX A

	Search
#3. Type of instrument	MH "physical examination+" OR MH "diagnostic tests, routine+" OR MH "diagnosis, delayed+" OR MH "diagnosis, differential+" OR MH "diagnostic errors+" OR MH "diagnostic imaging+" OR TI "anatomic landmarks" OR TI "anatomic landmark" OR TI "physical examination" OR TI "physical examinations" OR TI ("physical examinations and diagnoses") OR TI "routine diagnostic tests" OR TI "routine diagnostic test" OR TI "diagnostic test" OR TI "diagnostic tests" OR TI "delayed diagnosis" OR TI "delayed diagnoses" OR TI "late diagnosis" OR TI "late diagnoses" OR TI "differential diagnosis" OR TI "differential diagnoses" OR TI "diagnostic error" OR TI "diagnostic errors" OR TI Misdiagnosis OR TI Misdiagnoses OR TI "diagnostic imaging" OR TI "medical imaging" OR AB "anatomic landmarks" OR AB "anatomic landmark" OR AB "physical examination" OR AB "physical examinations" OR AB ("physical examinations and diagnoses") OR AB "routine diagnostic tests" OR AB "routine diagnostic test" OR AB "diagnostic test" OR AB "diagnostic tests" OR AB "delayed diagnosis" OR AB "delayed diagnoses" OR AB "late diagnosis" OR AB "late diagnoses" OR AB "differential diagnosis" OR AB "differential diagnoses" OR AB "diagnostic error" OR AB "diagnostic errors" OR AB Misdiagnosis OR AB Misdiagnoses OR AB "diagnostic imaging" OR AB "medical imaging"
#4. Measurement properties	MH "Research Measurement+" OR MH "Outcome Assessment" OR MH "Outcomes Research"
#5. Combining elements	#1 AND #2 AND #3 AND #4

SPORTDiscus (EBSCOhost) (March 30, 2019)

	Search
#1. Construct of interest	DE "MOTION" OR DE "JOINTS -- Range of motion" OR DE "DYSKINESISS" OR DE "PSYCHOLOGY of movement" OR TI motion OR TI motions OR TI "articular range of motion" OR TI "joint range of motion" OR TI "joint flexibility" OR TI "range of motion" OR TI "range of movement" OR TI "passive range of motion" OR TI "articular arthrometry" OR TI "articular goniometry" OR TI Movement OR TI Movements OR TI kinesis OR TI kineses OR TI "scapular kinematics" OR TI "scapulothoracic kinematics" OR TI Dyskinesis OR TI Dyskinesis OR TI Dyskinesis OR TI "abnormal movement" OR TI "abnormal movements" OR TI "psychomotor performance" OR TI "psychomotor performances" OR TI "visual motor coordination" OR TI "visual motor coordinations" OR TI "perceptual motor performance" OR TI "perceptual motor performances" OR TI "sensory motor performance" OR TI "sensory motor performances" OR AB Motion OR AB motions OR AB "articular range of motion" OR AB "joint range of motion" OR AB "joint flexibility" OR AB "range of motion" OR AB "range of movement" OR AB "passive range of motion" OR AB "articular arthrometry" OR AB "articular goniometry" OR AB Movement OR AB Movements OR AB kinesis OR AB kineses OR AB "scapular kinematics" OR AB "scapulothoracic kinematics" OR AB Dyskinesis OR AB Dyskinesis OR AB Dyskinesis OR AB "abnormal movement" OR AB "abnormal movements" OR AB "psychomotor performance" OR AB "psychomotor performances" OR AB "visual motor coordination" OR AB "visual motor coordinations" OR AB "perceptual motor performance" OR AB "perceptual motor performances" OR AB "sensory motor performance" OR AB "sensory motor performances"
#2. Target population	(DE "SHOULDER" OR DE "CLAVICLE" OR DE "SHOULDER joint" OR DE "SCAPULA" OR DE "CERVICAL vertebrae" OR DE "SHOULDER joint -- Rotator cuff" OR TI "Glenoid cavity" OR TI Shoulder OR TI Shoulders OR TI "shoulder joint" OR TI "shoulder joints" OR TI "glenohumeral joint" OR TI "glenohumeral joints" OR TI "glenoid cavities" OR TI "glenoid fossa of the scapula" OR TI "sternoclavicular joint" OR TI "sternoclavicular joints" OR TI Biceps OR TI "rotator cuff" OR TI "rotator cuffs" OR TI "cervical vertebrae" OR TI "cervicothoracic spine" OR TI cervicothoracic) OR ((DE "SHOULDER" OR DE "CLAVICLE" OR DE "SHOULDER joint" OR DE "SCAPULA" OR DE "CERVICAL vertebrae" OR DE "SHOULDER joint -- Rotator cuff" OR TI "Glenoid cavity" OR TI Shoulder OR TI Shoulders OR TI "shoulder joint" OR TI "shoulder joints" OR TI "glenohumeral joint" OR TI "glenohumeral joints" OR TI "glenoid cavities" OR TI "glenoid fossa of the scapula" OR TI "sternoclavicular joint" OR TI "sternoclavicular joints" OR TI Biceps OR TI "rotator cuff" OR TI "rotator cuffs" OR TI "cervical vertebrae" OR TI "cervicothoracic spine" OR TI cervicothoracic) AND (DE "SHOULDER pain" OR DE "SHOULDER joint -- Wounds & injuries" OR DE "SHOULDER joint -- Rotator cuff -- Wounds & injuries" OR DE "BRACHIAL plexus -- Wounds & injuries" OR DE "BURSITIS" OR DE "ADHESIVE capsulitis" OR DE "TENDINOSIS" OR DE "TENDINITIS" OR TI "shoulder pain" OR TI "shoulder impingement syndrome" OR TI "shoulder fractures" OR TI "shoulder fracture" OR TI "shoulder dislocation" OR TI "shoulder injuries" OR TI "shoulder injury" OR TI Impingement OR TI "proximal humeral fracture" OR TI "proximal humeral fractures" OR TI "glenohumeral dislocation" OR TI "glenohumeral dislocations" OR TI "glenohumeral subluxation" OR TI "glenohumeral subluxations" OR TI "glenohumeral instability" OR TI "brachial plexus neuritis" OR TI "brachial plexus neuritides" OR TI "parsonage-aldren-turner-syndrome" OR TI "parsonage-aldren-turner-syndromes" OR TI "parsonage-turner-syndrome" OR TI "parsonage-turner-syndromes" OR TI "cervicobrachial neuralgia" OR TI "cervicobrachial neuralgias" OR TI "brachial neuralgia" OR TI "brachial neuralgias" OR TI "bursitides adhesive" OR TI "capsulitis adhesive" OR TI Sprengel OR TI Bursitis OR TI "scapular dyskinesis" OR TI Tendinopathy OR TI tendinitis OR TI Tendonitis OR TI tendonitides OR TI Tendinosis OR AB "shoulder pain" OR AB "shoulder impingement syndrome" OR AB "shoulder fractures" OR AB "shoulder fracture" OR AB "shoulder dislocation" OR AB "shoulder injuries" OR AB "shoulder injury" OR AB Impingement OR AB "proximal humeral fracture" OR AB "proximal humeral fractures" OR AB "glenohumeral dislocation" OR AB "glenohumeral dislocations" OR AB "glenohumeral subluxation" OR AB "glenohumeral subluxations" OR AB "glenohumeral instability" OR AB "brachial plexus neuritis" OR AB "brachial plexus neuritides" OR AB "parsonage-aldren-turner-syndrome" OR AB "parsonage-aldren-turner-syndromes" OR AB "parsonage-turner-syndrome" OR AB "parsonage-turner-syndromes" OR AB "cervicobrachial neuralgia" OR AB "cervicobrachial neuralgias" OR AB "brachial neuralgia" OR AB "brachial neuralgias" OR AB "bursitides adhesive" OR AB "capsulitis adhesive" OR AB Sprengel OR AB Bursitis OR AB "scapular dyskinesis" OR AB Tendinopathy OR AB tendinitis OR AB Tendonitis OR AB tendonitides OR AB Tendinosis))

Table continues on page E5.

APPENDIX A

	Search
#3. Type of instrument	DE "PERIODIC health examinations" OR DE "DIAGNOSIS" OR DE "DIAGNOSTIC imaging" OR TI "anatomic landmarks" OR TI "anatomic landmark" OR TI "physical examination" OR TI "physical examinations" OR TI ("physical examinations and diagnoses") OR TI "routine diagnostic tests" OR TI "routine diagnostic test" OR TI "diagnostic test" OR TI "diagnostic tests" OR TI "delayed diagnosis" OR TI "delayed diagnoses" OR TI "late diagnosis" OR TI "late diagnoses" OR TI "differential diagnosis" OR TI "differential diagnoses" OR TI "diagnostic error" OR TI "diagnostic errors" OR TI Misdiagnosis OR TI Misdiagnoses OR TI "diagnostic imaging" OR TI "medical imaging" OR AB "anatomic landmarks" OR AB "anatomic landmark" OR AB "physical examination" OR AB "physical examinations" OR AB ("physical examinations and diagnoses") OR AB "routine diagnostic tests" OR AB "routine diagnostic test" OR AB "diagnostic test" OR AB "diagnostic tests" OR AB "delayed diagnosis" OR AB "delayed diagnoses" OR AB "late diagnosis" OR AB "late diagnoses" OR AB "differential diagnosis" OR AB "differential diagnoses" OR AB "diagnostic error" OR AB "diagnostic errors" OR AB Misdiagnosis OR AB Misdiagnoses OR AB "diagnostic imaging" OR AB "medical imaging"
#4. Combining elements	#1 AND #2 AND #3

Embase (March 30, 2019)

	Search
#1. Construct of interest	'psychomotor performance' AND [embase]/lim OR 'dyskinesis'/de AND [embase]/lim OR 'orientation' AND [embase]/lim OR 'movement (physiology)' AND [embase]/lim OR 'arthrometry' AND [embase]/lim OR 'motion' AND [embase]/lim OR 'range of motion' AND [embase]/lim AND 'human'/de OR motion:ab,ti OR motions:ab,ti OR 'articular range of motion':ab,ti OR 'joint range of motion':ab,ti OR 'joint flexibility':ab,ti OR 'range of motion':ab,ti OR 'range of movement':ab,ti OR 'passive range of motion':ab,ti OR 'articular arthrometry':ab,ti OR 'articular goniometry':ab,ti OR movement:ab,ti OR movements:ab,ti OR kinesis:ab,ti OR kineses:ab,ti OR 'scapular kinematics':ab,ti OR 'scapulothoracic kinematics':ab,ti OR dyskinesis:ab,ti OR dyskinesiss:ab,ti OR dyskinesis:ab,ti OR 'abnormal movement':ab,ti OR 'abnormal movements':ab,ti OR 'psychomotor performance':ab,ti OR 'psychomotor performances':ab,ti OR 'visual motor coordination':ab,ti OR 'visual motor coordinations':ab,ti OR 'perceptual motor performance':ab,ti OR 'perceptual motor performances':ab,ti OR 'sensory motor performance':ab,ti OR 'sensory motor performances':ab,ti AND [humans]/lim AND [embase]/lim
#2. Target population	('rotator cuff' AND [humans]/lim AND [embase]/lim OR 'biceps brachii muscle' AND [humans]/lim AND [embase]/lim OR 'cervical spine' AND [humans]/lim AND [embase]/lim OR 'sternoclavicular joint' AND [humans]/lim AND [embase]/lim OR 'clavicle' AND [humans]/lim AND [embase]/lim OR 'scapula'/exp AND [humans]/lim AND [embase]/lim OR 'shoulder'/de AND [humans]/lim AND [embase]/lim OR 'glenoid cavity':ab,ti OR shoulder:ab,ti OR shoulders:ab,ti OR 'shoulder joint':ab,ti OR 'shoulder joints':ab,ti OR 'glenohumeral joint':ab,ti OR 'glenohumeral joints':ab,ti OR 'glenoid cavities':ab,ti OR 'glenoid fossa of the scapula':ab,ti OR 'sternoclavicular joint':ab,ti OR 'sternoclavicular joints':ab,ti OR biceps:ab,ti OR 'rotator cuff':ab,ti OR 'rotator cuffs':ab,ti OR 'cervical vertebrae':ab,ti OR 'cervicothoracic spine':ab,ti OR cervicothoracic:ab,ti AND [humans]/lim AND [embase]/lim) OR (('rotator cuff' AND [humans]/lim AND [embase]/lim OR 'biceps brachii muscle' AND [humans]/lim AND [embase]/lim OR 'cervical spine' AND [humans]/lim AND [embase]/lim OR 'sternoclavicular joint' AND [humans]/lim AND [embase]/lim OR 'clavicle' AND [humans]/lim AND [embase]/lim OR 'scapula'/exp AND [humans]/lim AND [embase]/lim OR 'shoulder'/de AND [humans]/lim AND [embase]/lim OR 'glenoid cavity':ab,ti OR shoulder:ab,ti OR shoulders:ab,ti OR 'shoulder joint':ab,ti OR 'shoulder joints':ab,ti OR 'glenohumeral joint':ab,ti OR 'glenohumeral joints':ab,ti OR 'glenoid cavities':ab,ti OR 'glenoid fossa of the scapula':ab,ti OR 'sternoclavicular joint':ab,ti OR 'sternoclavicular joints':ab,ti OR biceps:ab,ti OR 'rotator cuff':ab,ti OR 'rotator cuffs':ab,ti OR 'cervical vertebrae':ab,ti OR 'cervicothoracic spine':ab,ti OR cervicothoracic:ab,ti AND [humans]/lim AND [embase]/lim) AND ('shoulder pain':ab,ti OR 'shoulder impingement syndrome':ab,ti OR 'shoulder fractures':ab,ti OR 'shoulder fracture':ab,ti OR 'shoulder dislocation':ab,ti OR 'shoulder injury':ab,ti OR 'shoulder injuries':ab,ti OR impingement:ab,ti OR 'proximal humeral fracture':ab,ti OR 'proximal humeral fractures':ab,ti OR 'glenohumeral dislocation':ab,ti OR 'glenohumeral dislocations':ab,ti OR 'glenohumeral subluxation':ab,ti OR 'glenohumeral subluxations':ab,ti OR 'glenohumeral instability':ab,ti OR 'brachial plexus neuritis':ab,ti OR 'brachial plexus neuritides':ab,ti OR 'parsonage-aldren-turner-syndrome':ab,ti OR 'parsonage-aldren-turner-syndromes':ab,ti OR 'parsonage-turner-syndrome':ab,ti OR 'parsonage-turner-syndromes':ab,ti OR 'cervicobrachial neuralgia':ab,ti OR 'cervicobrachial neuralgias':ab,ti OR 'bursitides adhesive':ab,ti OR 'brachial neuralgia':ab,ti OR 'brachial neuralgias':ab,ti OR 'capsulitis adhesive':ab,ti OR sprengel:ab,ti OR bursitis:ab,ti OR 'scapular dyskinesis':ab,ti OR tendinopathy:ab,ti OR tendinitis:ab,ti OR tendonitis:ab,ti OR tendonitides:ab,ti OR tendinosis:ab,ti AND [humans]/lim AND [embase]/lim OR 'tendinitis'/de AND [humans]/lim AND [embase]/lim OR 'shoulder impingement syndrome'/de AND [humans]/lim AND [embase]/lim OR 'bursitis'/de AND [humans]/lim AND [embase]/lim OR 'brachial plexus' AND [humans]/lim AND [embase]/lim OR 'shoulder dislocation'/de AND [humans]/lim AND [embase]/lim OR 'shoulder pain'/de AND [humans]/lim AND [embase]/lim))

Table continues on page E6.

APPENDIX A

	Search
#3. Type of instrument	'differential diagnosis' AND [humans]/lim AND [embase]/lim OR 'periodic medical examination' AND [humans]/lim AND [embase]/lim OR 'physical examination' AND [humans]/lim AND [embase]/lim OR 'diagnostic error' AND [humans]/lim AND [embase]/lim OR 'diagnostic accuracy' AND [humans]/lim AND [embase]/lim OR 'delayed diagnosis' AND [humans]/lim AND [embase]/lim OR 'exercise test' AND [humans]/lim AND [embase]/lim OR 'questionnaire'/exp OR 'named inventories, questionnaires and rating scales'/exp OR 'psychometry'/exp OR 'outcome assessment'/exp OR 'pain assessment'/exp OR 'disability'/exp OR 'validity'/exp OR 'reliability'/exp OR 'anatomic landmarks':ab,ti OR 'anatomic landmark':ab,ti OR 'physical examination':ab,ti OR 'physical examinations':ab,ti OR 'physical examinations and diagnoses':ab,ti OR 'routine diagnostic test':ab,ti OR 'routine diagnostic tests':ab,ti OR 'diagnostic test':ab,ti OR 'diagnostic tests':ab,ti OR 'delayed diagnosis':ab,ti OR 'delayed diagnoses':ab,ti OR 'late diagnosis':ab,ti OR 'late diagnoses':ab,ti OR 'differential diagnosis':ab,ti OR 'differential diagnoses':ab,ti OR 'diagnostic error':ab,ti OR 'diagnostic errors':ab,ti OR misdiagnosis:ab,ti OR misdiagnoses:ab,ti OR 'diagnostic imaging':ab,ti OR 'medical imaging':ab,ti AND [humans]/lim AND [embase]/lim
#4. Measurement properties	'questionnaire' /exp OR 'named inventories, questionnaires and rating scales'/exp OR 'psychometry'/exp OR 'outcome assessment'/exp OR 'pain assessment'/exp OR 'disability'/exp OR 'validity'/exp OR 'reliability'/exp
#5. Combining elements	#1 AND #2 AND #3 AND #4

APPENDIX B

FULL TABLES

TABLE 1. Study Characteristics and Methodological Quality Ratings of Included Studies

	Study Characteristics			Methodological Quality[a]	
Study	**Measurement Instrument, Material**	**Reference Standard, Comparator Instrument, Calculated Statistics**	**Population and Tester Characteristics**	**Box F[b]**	**Box H[c]**
Crotty and Smith[6d]	Determination of change in scapular position over time to detect muscle fatigue (established when a change of ≥1.5 cm was present) by (1) the DiVeta[9] palpation measurement method or (2) the Kibler[21] palpation measurement method Tape measure	Construct to be measured: muscle fatigue–based scapular dyskinesis NR Analysis of variance	Participants: n = 10 (all male) asymptomatic athletes (swimming) age NR Testers: n = 3 MDs (physical medicine and rehabilitation)	NA	NA
Greenfield et al[15]	Palpation of bony landmarks to assess (1) the scapular protraction ratio and (2) scapular rotation angle NR	Criterion: measurement of distances between identical bony landmarks on PA-view X-rays Correlations	Participants: n = 60 (asymptomatic or right scapula; 34 male, 26 female); age, 39 y (17-65 y) Testers: n = 4; occupation NR	NA	Poor: item 5, "other important methodological flaws in the design or execution of the study"
Johnson et al[20]	Determination of scapular upward rotation angle at (1) rest and at (2) 60°, (3) 90°, and (4) 120° of thoracohumeral elevation Digital protractor	Criterion: (1) inclinometer and magnetic tracking device under static-arm conditions, (2) inclinometer and magnetic tracking device during active arm elevation Magnetic tracking device Correlations	Participants: n = 39 students and patients (16 symptomatic, 23 asymptomatic; sex NR); age, 35 ± 11 y Tester: n = 1; occupation NR	Poor: item 4, "unclear what was expected"	Good: item 2, "not described, but it can be deduced how missing items were handled"
Koslow et al[23]	LSST to quantitatively measure "scapular stabilizer strength" Tape measure	Criterion: presence of scapular symmetry (>1.5-cm side-to-side difference) in asymptomatic athletes Specificity	Participants: n = 71 (33 male, 38 female) asymptomatic athletes (1-arm-dominant sports); age, 19.3 y (18-22 y) Tester: n = 1; occupation NR	NA	Good: item 4, "no evidence provided, but assumable that the criterion used can be considered an adequate 'gold standard'"
Morrissey et al[32]	Shoulder medial rotation test to detect a scapular give to predict shoulder impingement by palpation Motion-tracking system	Criterion: coracoid translation >6 mm during humeral rotation, measured with 3-D ultrasound Correlations	Participants: n = 11 (3 male, 8 female) asymptomatic; age, 24 ± 4 y (19-47 y) Tester: n = 1; PT (manual therapy)	Poor: item 9, "other important methodological flaws in the design or execution of the study." The thorax does not seem to be referenced in 3-D ultrasound measurement	Poor: item 4, "the criterion used cannot be considered an adequate 'gold standard.'" The transition from humeral measurement to scapular measurement with 3-D ultrasound was unclear, so the study was categorized as hypotheses testing for construct validity

Table continues on page E8.

APPENDIX B

	Study Characteristics			Methodological Quality[a]	
Study	Measurement Instrument, Material	Reference Standard, Comparator Instrument, Calculated Statistics	Population and Tester Characteristics	Box F[b]	Box H[c]
Nadeau et al[33]	Determination of active (1) scapular elevation, (2) retraction, and (3) protraction ROM Tape measure and goniometer	Criterion: 3-D kinematic data Correlations	Participants: n = 30 (15 male, 15 female) asymptomatic university employees and students; age, 45.3 ± 11.5 y Testers: n = 2 PTs	NA	Good: item 4, "no evidence provided, but assumable that the criterion used can be considered an adequate 'gold standard'"
Nijs et al[34]	Determination of scapular position by (1) acromion-to-table distance and (2) scapula-to-spine distance, and determination of asymmetrical scapular posture (>1.5-cm left-to-right difference) by (3) LSST Tape measure	Construct to be measured: scapular dysfunction Criterion: (1) pain VAS,[14] (2) SDQ[57] Correlations	Participants: n = 29 (10 male, 19 female) symptomatic; age, 56.6 ± 14.9 y (18-81 y) Testers: n = 2; occupation NR	Fair: item 4, "hypotheses vague or not formulated, but possible to deduce what was expected" and item 8, "some information on measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population"	NA
O'Shea et al[35]	Determination of scapular position by (1) "C7 protractor method" and (2) "T8 protractor method" Metal right-angle protractor	Criterion/comparator: 2-D camera analysis Correlations	Participants: n = 34 (21 male, 13 female) symptomatic; age, 44.53 ± 16.71 y Testers: n = 2 PTs	Poor: item 8, "no information on the measurement properties of the comparator instrument(s)"	Poor: item 4, "criterion used cannot be considered an adequate 'gold standard,'" so the study was categorized as hypotheses testing for construct validity
Park et al[38]	Determination of scapular dyskinesis by 3-D wing CT analysis 16- or 64-MDCT	Construct to be measured: scapular dyskinesis Criterion/comparator: 4-type scapular dyskinesis observational method Analysis of variance	Participants: n = 89 (sex and age NR) symptomatic athletes Testers: n = 7 (6 orthopaedic surgeons and 1 athletic trainer)	Poor: item 4, "unclear what was expected"	Poor: item 4, "criterion used cannot be considered an adequate 'gold standard,'" so the study was categorized as hypotheses testing for construct validity
Peterson et al[39]	Determination of forward shoulder posture by (1) acromion-to-wall distance and (2) scapula-to-spine distance (1) Double square technique, (2) Baylor square technique, (3) tape measure	Criterion: measurement of identical bony landmarks on lateral-view X-rays Correlations	Participants: n = 49 (25 male, 24 female) asymptomatic; age, 30 y (20-48 y) Tester: n = 1 PT	NA	Poor: item 5, "other important methodological flaws in the design or execution of the study"
Sobush et al[48]	Lennie test (palpation method) to assess (1) linear scapula-to-spine distance, (2) scapular rotation angle, and (3) scapular symmetry Scoliometer	Criterion: measurement of identical bony landmarks on PA-view X-rays Correlations	Participants: n = 15 (all female) asymptomatic; age NR Testers: n = 3 (1 PT and 2 PT students)	NA	Good

Table continues on page E9.

APPENDIX B

Study	Study Characteristics: Measurement Instrument, Material	Study Characteristics: Reference Standard, Comparator Instrument, Calculated Statistics	Study Characteristics: Population and Tester Characteristics	Methodological Quality[a]: Box F[b]	Methodological Quality[a]: Box H[c]
Tate et al[49]	Scapular dyskinesis test to assess scapular dyskinesis, characterized by either scapular winging or dysrhythmia (or both) Visual estimation	Criterion: 3-D kinematic data, measured with an electromagnetic-based motion-capture system, from participants with normal scapular motion versus participants with a positive scapular dyskinesis test Odds ratios	Participants: n = 66 (50 male, 16 female) athletes (symptom status NR); age, 20 ± 2.6 y Testers: n = 2 (1 PT student and 1 athletic trainer)	Poor: item 4, "unclear what was expected"	Poor: item 4, "criterion used cannot be considered an adequate 'gold standard,'" so the study was categorized as hypotheses testing for construct validity
Tucker and Ingram[54]	Digital inclinometer to assess scapular upward rotation at (1) rest and at (2) 60°, (3) 90°, and (4) 120° of thoracohumeral elevation 2-D inclinometer	Criterion: kinematic data measured with a digital protractor Correlations	Participants: n = 30 (all male) asymptomatic; age, 21.9 ± 2.3 y Tester: n = 1; occupation NR	Fair: item 5, "other minor methodological flaws in the design or execution of the study (eg, only data presented on a comparison with an instrument that measures another construct)"	NA
Uhl et al[55]	Observational method to assess scapular dysfunction during humeral elevation by using (1) the 4-type scapular dyskinesis classification system (I, inferior scapular angle prominence; II, medial scapular border prominence; III, superior scapular angle prominence; IV, normal) and (2) the dichotomous scapular dyskinesis classification system (at least 1 of types I, II, or III present versus type IV)	Criterion: asymmetric scapular motion Sensitivity, specificity, positive predictive value, negative predictive value, and accuracy	Participants: n = 56 (35 male, 21 female) symptomatic (n = 35) and asymptomatic (n = 21); age NR Testers: n = 2 (1 orthopaedic surgeon and 1 not specified)	Poor: item 9, "other important methodological flaws in the design or execution of the study"	Poor: item 4, "criterion used cannot be considered an adequate 'gold standard.'" The occurrence of asymmetry was used as a gold standard to assess scapular dysfunction, while the article states that the asymmetries occurred in both symptomatic and asymptomatic participants. When asymmetry occurred, no data were provided to support which scapula could be designated as deviated, so the study was categorized as hypothesis testing for construct validity

Abbreviations: CT, computed tomography; LSST, lateral scapular slide test; MD, medical doctor; MDCT, multidetector computed tomography; NA, not applicable; NR, not reported; PA, posterior to anterior; PT, physical therapist; ROM, range of movement; SDQ, Shoulder Disability Questionnaire; VAS, visual analog scale.

[a]*Using the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) checklist (lowest scored item).*

[b]*Hypothesis testing for construct validity.*

[c]*Criterion validity.*

[d]*Box I: responsiveness. This study was rated as poor: item 11, "no description of the constructs measured by the comparator instrument(s)" and item 13, "other important methodological flaws in the design or execution of the study." Pre-post comparison of the scapula-to-thoracic spine distance was measured and related to scapular muscle fatigue. Muscle fatigue was not measured with a comparator instrument.*

APPENDIX B

TABLE 2. Quality of Evidence for Measurements to Assess Scapular Posture and Positioning

Measurement Instrument		Criterion Validity		Hypotheses Testing for Construct Validity		
Test Condition/Study	**Measurement, Material**	**Reference Standard**	**Level of Evidence, COSMIN Score, Measurement Property Rating**	**Construct to Be Measured, Comparator**	**Hypotheses**	**Level of Evidence, COSMIN Score, Measurement Property Rating**
Scapular Protraction/Retraction Posture						
Participants standing upright with arms dangling alongside						
Sobush et al[48]	Thoracic spine-to-superior scapular angle distance (Lennie test) Scoliometer	Measurement of distances between identical landmarks displayed on AP-view X-rays	No evidence Good Indeterminate	NA	NA	NA
	Thoracic spine-to-inferior scapular angle distance (Lennie test) Scoliometer	Measurement of distances between identical landmarks displayed on AP-view X-rays	Moderate Good Sufficient	NA	NA	NA
	Thoracic spine-to-scapular root distance (Lennie test) Scoliometer	Measurement of distances between identical landmarks displayed on AP-view X-rays	Moderate Good Sufficient	NA	NA	NA
Peterson et al[39]	Linear distance from the C7 spinous process to the anterior acromion process Baylor square device	Horizontal distance between the anterior acromion process and the perpendicular line from C7 measured on lateral-view X-rays	No evidence Poor Sufficient	NA	NA	NA
	Horizontal distance from the medial scapular border to the T3 spinous process Tape measure	Horizontal distance between the anterior acromion process and the perpendicular line from C7 measured on lateral-view X-rays	No evidence Poor Insufficient	NA	NA	NA
	Linear distance from the wall to the anterior acromion process Double square device	Measurement of horizontal distance between the anterior acromion process and the perpendicular line from C7 displayed on lateral-view X-rays	No evidence Poor Insufficient	NA	NA	NA

Table continues on page E11.

APPENDIX B

Measurement Instrument		Criterion Validity		Hypotheses Testing for Construct Validity		
Test Condition/Study	Measurement, Material	Reference Standard	Level of Evidence, COSMIN Score, Measurement Property Rating	Construct to Be Measured, Comparator	Hypotheses	Level of Evidence, COSMIN Score, Measurement Property Rating
Greenfield et al[15]	Linear distance from the thoracic spine corresponding to the root of the scapular spine to the tip of the acromion, divided by the linear distance from the root of the scapular spine to the tip of the acromion	Measurement of distances between identical landmarks displayed on AP-view X-rays	No evidence Poor Sufficient	NA	NA	NA
Scapular Elevation/Depression Posture						
Participants standing upright with arms dangling alongside						
O'Shea et al[35]	"C7 method": vertical distance between C7 and the superior aspect of the scapular spine Protractor	NA	NA	Scapular position Comparator: 2-D analysis of marker distances with camera equipment and software	NR	No evidence Poor Sufficient
	"T8 method": vertical distance between T8 and the inferior scapular angle Protractor	NA	NA	Scapular position Comparator: 2-D analysis of marker distances with camera equipment and software	NR	No evidence Poor Sufficient
Asymmetrical Scapular Posture						
Participants standing upright with arms dangling alongside						
Sobush et al[48]	Difference between the vertical position of the inferior angles of the dominant and nondominant scapulae (Lennie test) Scoliometer	Measurement of distances between identical landmarks displayed on AP-view X-rays	Moderate Good Insufficient/ indeterminate	NA	NA	NA

Table continues on page E12.

APPENDIX B

Measurement Instrument		Criterion Validity		Hypotheses Testing for Construct Validity		
Test Condition/Study	Measurement, Material	Reference Standard	Level of Evidence, COSMIN Score, Measurement Property Rating	Construct to Be Measured, Comparator	Hypotheses	Level of Evidence, COSMIN Score, Measurement Property Rating
Scapular Positioning: Upward Rotation Angles						
Participants standing upright with arms dangling alongside						
Johnson et al[20]	Digital protractor (inclinometer)	3-D kinematic data derived from a magnetic tracking device	Moderate Good Sufficient	NA	NA	NA
	Digital protractor (inclinometer)	NA	NA	Scapular upward rotation Comparator: dynamic measurements with a magnetic tracking device	NR	No evidence Poor Sufficient
Tucker and Ingram[54]	2-D inclinometer	NA	NA	Scapular upward rotation Comparator: inclinometer (digital protractor)	"The electrical inclinometer will produce valid static measurements for scapular upward rotation in the plane of the scapula"	Low Fair Sufficient
Greenfield et al[15]	Tangents of linear measurements: distance between the inferior scapular angle and the corresponding mark on the thoracic spine, divided by the linear distance between the marks on the thoracic spine corresponding to the root of the scapular spine and the inferior scapular angle	Measurement of distances between identical landmarks displayed on AP-view X-rays	No evidence Poor Sufficient	NA	NA	NA
Sobush et al[48]	Distance between the thoracic midline and the superior angle of the scapula, and the distance between the thoracic midline and the inferior angle of the scapula, and the length of the medial border of the scapula (Lennie test)	Measurement of distances between identical landmarks displayed on AP-view X-rays	Moderate Good Insufficient	NA	NA	NA

Table continues on page E13.

APPENDIX B

Measurement Instrument		Criterion Validity		Hypotheses Testing for Construct Validity		
Test Condition/Study	**Measurement, Material**	**Reference Standard**	**Level of Evidence, COSMIN Score, Measurement Property Rating**	**Construct to Be Measured, Comparator**	**Hypotheses**	**Level of Evidence, COSMIN Score, Measurement Property Rating**
60° of elevation[a]						
Johnson et al[20]	Digital protractor (inclinometer)	Magnetic tracking device	Moderate Good Sufficient	NA	NA	NA
	Digital protractor (inclinometer)	NA	NA	Scapular upward rotation Comparator: dynamic kinematic data derived from a 3-D electromagnetic tracking device	NR	No evidence Poor Sufficient
Tucker and Ingram[54]	2-D inclinometer	NA	NA	Comparator: inclinometer (digital protractor)	"The electrical inclinometer will produce valid static measurements for scapular upward rotation in the plane of the scapula"	Low Fair Sufficient
90° of elevation[a]						
Johnson et al[20]	Digital protractor (inclinometer)	Static kinematic data derived from a 3-D electromagnetic tracking device	Moderate Good Sufficient	NA	NA	NA
	Digital protractor (inclinometer)	NA	NA	Scapular upward rotation Comparator: dynamic kinematic data derived from a 3-D electromagnetic tracking device	NR	No evidence Poor Sufficient
Tucker and Ingram[54]	2-D inclinometer	NA	NA	Scapular upward rotation Comparator: inclinometer (digital protractor)	"The electrical inclinometer will produce valid static measurements for scapular upward rotation in the plane of the scapula"	Low Fair Sufficient
120° of elevation[a]						
Johnson et al[20]	Digital protractor (inclinometer)	Static kinematic data derived from a 3-D electromagnetic tracking device	Moderate Good Sufficient	NA	NA	NA
	Digital protractor (inclinometer)	NA	NA	Scapular upward rotation Comparator: dynamic kinematic data derived from a 3-D electromagnetic tracking device	NR	No evidence Poor Sufficient

Table continues on page E14.

APPENDIX B

Measurement Instrument		Criterion Validity		Hypotheses Testing for Construct Validity		
Test Condition/Study	Measurement, Material	Reference Standard	Level of Evidence, COSMIN Score, Measurement Property Rating	Construct to Be Measured, Comparator	Hypotheses	Level of Evidence, COSMIN Score, Measurement Property Rating
Tucker and Ingram[54]	2-D inclinometer	NA	NA	Scapular upward rotation Comparator: digital protractor (inclinometer)	"The electrical inclinometer will produce valid static measurements for scapular upward rotation in the plane of the scapula"	Low Fair Sufficient

Abbreviations: AP, anterior to posterior; COSMIN, COnsensus-based Standards for the selection of health Measurement INstruments; NA, not applicable; NR, not reported.
[a]Thoracohumeral elevation in the scapular plane.

TABLE 3. Quality of Evidence for Measurements to Assess Scapular Movement

Measurement Instrument		Criterion Validity		Hypotheses Testing for Construct Validity		
Test Condition/Study	Measurement, Material	Reference Standard	Level of Evidence, COSMIN Score, Measurement Property Rating	Construct to Be Measured, Comparator	Hypotheses	Level of Evidence, COSMIN Score, Measurement Property Rating
			Scapular Active Range of Motion			
Elevation						
Nadeau et al[33]	Linear distance from neutral position to end range Tape measure	Kinematic data derived from a 3-D motion-analysis system	Moderate Good Insufficient	NA	NA	NA
	Linear distance from neutral position to end range Goniometer	Kinematic data derived from a 3-D motion-analysis system	Moderate Good Insufficient	NA	NA	NA
Retraction						
Nadeau et al[33]	Linear distance from neutral position to end range Tape measure	Kinematic data derived from a 3-D motion-analysis system	Moderate Good Insufficient	NA	NA	NA

Abbreviations: COSMIN, COnsensus-based Standards for the selection of health Measurement INstruments; NA, not applicable.

APPENDIX B

TABLE 4. Quality of Evidence for Measurement Instruments to Determine Scapular Dyskinesis

Measurement Instrument		Criterion Validity		Hypotheses Testing for Construct Validity		
Test Condition/Study	**Measurement, Material**	**Reference Standard**	**Level of Evidence, COSMIN Score, Measurement Property Rating**	**Construct to Be Measured, Comparator**	**Hypotheses**	**Level of Evidence, COSMIN Score, Measurement Property Rating**
Dichotomous Observational Method						
During rest and humeral elevation						
Uhl et al[55]	Observation of presence of posterior scapular displacement during motion and positioning Video analysis	NA	NA	Scapular dyskinesis Comparator: symmetry in scapular motion and positioning	"Observation-based clinical assessment methods of scapular asymmetry can yield good criterion validity"	No evidence Poor Sufficient
4-Type Observational Method						
During rest and humeral elevation						
Uhl et al[55]	Observation of presence of a specific posterior scapular displacement and asymmetry (types I-IV)[a] Video analysis	NA	NA	Scapular dyskinesis Comparator: symmetry in scapular motion and positioning	"Observation-based clinical assessment methods of scapular asymmetry can yield good criterion validity"	No evidence Poor Insufficient
3-D Wing CT Analysis						
Static, participants in supine position						
Park et al[38]	Scapular upward rotation, internal rotation, anterior tilting, superior translation, and protraction angles 3-D wing CT machine	NA	NA	Scapular dyskinesis Comparator: 4-type scapular dyskinesis observational method	NR	No evidence Poor Indeterminate
Scapular Dyskinesis Test						
Participant is standing						
Tate et al[49]	Rating of scapular movement pattern as "normal" or "obvious abnormality"[b] during weighted thoracohumeral flexion and abduction Videotaping	Kinematic data derived from a 3-D motion-analysis system	No evidence Poor Indeterminate	Scapular dyskinesis Comparator: pain scale of Penn Shoulder Score[25]	NR	No evidence Poor Insufficient

Table continues on page E16.

APPENDIX B

Measurement Instrument		Criterion Validity		Hypotheses Testing for Construct Validity		
Test Condition/Study	Measurement, Material	Reference Standard	Level of Evidence, COSMIN Score, Measurement Property Rating	Construct to Be Measured, Comparator	Hypotheses	Level of Evidence, COSMIN Score, Measurement Property Rating
			Lateral Scapular Slide Test			
Position 1: participant's arms dangling relaxed at the sides						
Koslow et al[23]	Measurement of side-to-side difference in linear scapula-to-spine distance (inferior scapular angle and the closest spinous process) Tape measure	Scapular stabilizer strength Symmetrical scapular positioning (<1.5-cm side-to-side difference) in asymptomatic athletes[21]	Moderate Good Insufficient	NA	NA	NA
Nijs et al[34]	Measurement of side-to-side difference in linear scapula-to-spine distance (inferior scapular angle and the closest spinous process) Tape measure	NA	NA	Scapular dyskinesis Comparator: (1) pain VAS and (2) SDQ	"Test outcome would differ between the symptomatic and asymptomatic sides"	Low Fair Insufficient
Position 2: participant's hands placed on the hips with the fingers directed anteriorly and thumbs posteriorly						
Koslow et al[23]	Measurement of side-to-side difference in linear scapula-to-spine distance (inferior scapular angle and the closest spinous process) Tape measure	Scapular stabilizer strength Symmetrical scapular positioning (<1.5-cm side-to-side difference) in asymptomatic athletes[21]	Moderate Good Insufficient	NA	NA	NA
Nijs et al[34]	Measurement of side-to-side difference in linear scapula-to-spine distance (inferior scapular angle and the closest spinous process) Tape measure	NA	NA	Scapular dyskinesis Comparator: (1) pain VAS and (2) SDQ	"Test outcome would differ between the symptomatic and asymptomatic sides"	Low Fair Insufficient

Table continues on page E17.

APPENDIX B

Measurement Instrument		Criterion Validity		Hypotheses Testing for Construct Validity		
Test Condition/Study	Measurement, Material	Reference Standard	Level of Evidence, COSMIN Score, Measurement Property Rating	Construct to Be Measured, Comparator	Hypotheses	Level of Evidence, COSMIN Score, Measurement Property Rating
Position 3: participant's arms were positioned at 90° of elevation, with maximal internal rotation of the glenohumeral joint						
Koslow et al[23]	Measurement of side-to-side difference in linear scapula-to-spine distance (inferior scapular angle and the closest spinous process) Tape measure	Scapular stabilizer strength Symmetrical scapular positioning (<1.5-cm side-to-side difference) in asymptomatic athletes[21]	Moderate Good Insufficient	NA	NA	NA
Nijs et al[34]	Measurement of side-to-side difference in linear scapula-to-spine distance (inferior scapular angle and the closest spinous process) Tape measure	NA	NA	Scapular dyskinesis Comparator: (1) pain VAS and (2) SDQ	"Test outcome would differ between the symptomatic and asymptomatic sides"	Low Fair Insufficient
Multitest regimen (positions 1, 2, and 3)						
Koslow et al[23]	Measurement of side-to-side difference in linear scapula-to-spine distance (inferior scapular angle and the closest spinous process) Tape measure	Weakness/decreased performance of scapular muscles Symmetrical scapular positioning (<1.5-cm side-to-side difference) in asymptomatic athletes[21]	Moderate Good Insufficient	NA	NA	NA
Shoulder Medial Rotation Test						
Participant supine with arm abducted 90°						
Morrissey et al[32]	Palpation measurement of glenohumeral and scapular translation during arm rotation Motion-tracking system	NA	NA	Scapular dysfunction associated with impingement Comparator: translation (46 mm) of the scapula before 70° of medial arm rotation, measured with 3-D ultrasound	NR	No evidence Poor Sufficient

Table continues on page E18.

APPENDIX B

Measurement Instrument		Criterion Validity		Hypotheses Testing for Construct Validity		
Test Condition/Study	Measurement, Material	Reference Standard	Level of Evidence, COSMIN Score, Measurement Property Rating	Construct to Be Measured, Comparator	Hypotheses	Level of Evidence, COSMIN Score, Measurement Property Rating
Measurement of Scapular Malposition						
Participants standing upright with arms dangling alongside, with scapula in (1) neutral position and (2) maximally retracted						
Koslow et al[23]	Scapula-to-spine distance (distance from T4 spinous process to medial scapular border) Tape measure	NA	NA	Scapular dyskinesis Comparator: (1) pain VAS and (2) SDQ	"Test outcome would differ between the symptomatic and asymptomatic sides"	Low Fair Insufficient
Participants in supine position, with the scapula (1) in neutral position and (2) maximally retracted						
Koslow et al[23]	Acromion-to-table distance Caliper	NA	NA	Scapular dyskinesis Comparator: (1) pain VAS and (2) SDQ	"Test outcome would differ between the symptomatic and asymptomatic sides"	Low Fair Insufficient

Abbreviations: COSMIN, COnsensus-based Standards for the selection of health Measurement INstruments; CT, computed tomography; NA, not applicable; NR, not reported; SDQ, Shoulder Disability Questionnaire; VAS, visual analog scale.
[a]*Type I, prominence of the inferior medial scapular angle; type II, prominence of the entire medial border; type III, prominence of the superior scapular border; type IV, normal.*
[b]*Normal, either test motions are rated as normal or 1 motion is rated as normal and the other as having subtle scapular dysrhythmia or winging; obvious abnormality, either flexion or abduction is rated as having obvious scapular dysrhythmia or winging.*

TABLE 5. Quality of Evidence for the Responsiveness of Measurement Instruments to Assess Scapular Dyskinesis

Measurement Instrument		Responsiveness	
Test Condition/Study	Measurement	Construct to Be Measured, Comparator	Level of Evidence, COSMIN Score, Measurement Property Rating
Static, participant's arms in neutral position			
Crotty and Smith[6]	Change in linear distance from the inferior scapular angle to the corresponding thoracic spinous process, or to the T7 or T8 spinous process (Kibler technique[21]), before and after swimming. Test is positive when pre-post difference is ≥1.5 cm	Muscle fatigue–based scapular dyskinesis Comparator NR	No evidence Poor Indeterminate
	Change in linear distance between the T3 spinous process and the posterolateral acromial angle (DiVeta technique[9]) before and after swimming. Test is positive when pre-post difference is ≥1.5 cm	Muscle fatigue–based scapular dyskinesis Comparator NR	No evidence Poor Indeterminate

Abbreviations: COSMIN, COnsensus-based Standards for the selection of health Measurement INstruments; NR, not reported.

APPENDIX C

EXCLUDED STUDIES AND REASONS FOR EXCLUSION

Study	Reason for Exclusion
Bourne D, Choo A, Regan W, MacIntyre D, Oxland T. Accuracy of digitization of bony landmarks for measuring change in scapular attitude. *Proc Inst Mech Eng H*. 2009;223:349-361. https://doi.org/10.1243/09544119JEIM480	Measurement instrument not clinically applicable (laboratory setting)
Bourne DA, Choo AM, Regan WD, MacIntyre DL, Oxland TR. The placement of skin surface markers for non-invasive measurement of scapular kinematics affects accuracy and reliability. *Ann Biomed Eng*. 2011;39:777-785. https://doi.org/10.1007/s10439-010-0185-1	Measurement instrument not clinically applicable (laboratory setting)
Brochard S, Lempereur M, Rémy-Néris O. Accuracy and reliability of three methods of recording scapular motion using reflective skin markers. *Proc Inst Mech Eng H*. 2011;225:100-105. https://doi.org/10.1243/09544119JEIM830	Measurement instrument not clinically applicable (laboratory setting)
Brochard S, Lempereur M, Rémy-Néris O. Double calibration: an accurate, reliable and easy-to-use method for 3D scapular motion analysis. *J Biomech*. 2011;44:751-754. https://doi.org/10.1016/j.jbiomech.2010.11.017	Measurement instrument not clinically applicable (laboratory setting)
Chu Y, Akins J, Lovalekar M, Tashman S, Lephart S, Sell T. Validation of a video-based motion analysis technique in 3-D dynamic scapular kinematic measurements. *J Biomech*. 2012;45:2462-2466. https://doi.org/10.1016/j.jbiomech.2012.06.025	Measurement instrument not clinically applicable (laboratory setting)
Deng HR, Shih YF. Test validity and intra-rater reliability in the measurement of scapular position sense in asymptomatic young adults. *Man Ther*. 2015;20:503-507. https://doi.org/10.1016/j.math.2015.02.002	Joint position sense No modality of interest
Gomes PF, Sesselmann M, Faria CD, Araújo PA, Teixeira-Salmela LF. Measurement of scapular kinematics with the moiré fringe projection technique. *J Biomech*. 2010;43:1215-1219. https://doi.org/10.1016/j.jbiomech.2009.12.015	Measurement instrument not clinically applicable (laboratory setting)
Hickey BW, Milosavljevic S, Bell ML, Milburn PD. Accuracy and reliability of observational motion analysis in identifying shoulder symptoms. *Man Ther*. 2007;12:263-270. https://doi.org/10.1016/j.math.2006.05.005	Involves thoracohumeral assessment. No data provided on clinical assessment of scapular posture and/or motion
Lempereur M, Brochard S, Mao L, Rémy-Néris O. Validity and reliability of shoulder kinematics in typically developing children and children with hemiplegic cerebral palsy. *J Biomech*. 2012;45:2028-2034. https://doi.org/10.1016/j.jbiomech.2012.05.020	Measurement instrument not clinically applicable (laboratory setting)
Lewis J, Green A, Reichard Z, Wright C. Scapular position: the validity of skin surface palpation. *Man Ther*. 2002;7:26-30. https://doi.org/10.1054/math.2001.0405	Assessment of embalmed cadaver shoulders No population of interest
Lovern B, Stroud LA, Evans RO, Evans SL, Holt CA. Dynamic tracking of the scapula using skin-mounted markers. *Proc Inst Mech Eng H*. 2009;223:823-831. https://doi.org/10.1243/09544119JEIM554	Viability study
Matsui K, Shimada K, Andrew PD. Deviation of skin marker from bone target during movement of the scapula. *J Orthop Sci*. 2006;11:180-184. https://doi.org/10.1007/s00776-005-1000-y	Measurement instrument not clinically applicable (laboratory setting)
Mattson JM, Russo SA, Rose WC, Rowley KM, Richards JG. Identification of scapular kinematics using surface mapping: a validation study. *J Biomech*. 2012;45:2176-2179. https://doi.org/10.1016/j.jbiomech.2012.05.048	Measurement instrument not clinically applicable (laboratory setting)
Prinold JA, Shaheen AF, Bull AM. Skin-fixed scapula trackers: a comparison of two dynamic methods across a range of calibration positions. *J Biomech*. 2011;44:2004-2007. https://doi.org/10.1016/j.jbiomech.2011.05.010	Measurement instrument not clinically applicable (laboratory setting)
Roren A, Fayad F, Roby-Brami A, et al. Precision of 3D scapular kinematic measurements for analytic arm movements and activities of daily living. *Man Ther*. 2013;18:473-480. https://doi.org/10.1016/j.math.2013.04.005	Measurement instrument not clinically applicable (laboratory setting)
Warner MB, Chappell PH, Stokes MJ. Measuring scapular kinematics during arm lowering using the acromion marker cluster. *Hum Mov Sci*. 2012;31:386-396. https://doi.org/10.1016/j.humov.2011.07.004	Measurement instrument not clinically applicable (laboratory setting)

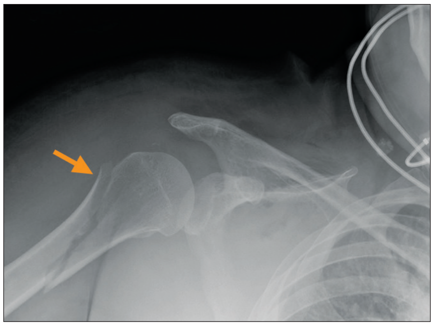


FIGURE 1. Anteroposterior radiograph of the right shoulder following arthroscopic distal clavicle resection and biceps tenodesis, demonstrating a minimally displaced comminuted fracture of the proximal humeral metadiaphysis (arrow).

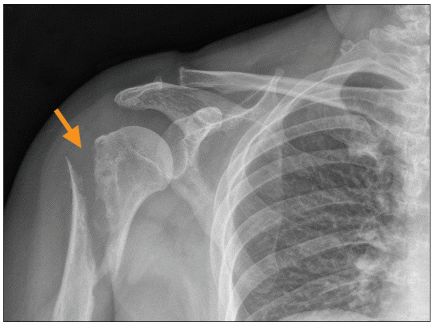


FIGURE 2. Anteroposterior radiograph of the right shoulder 14 weeks following arthroscopic surgery, demonstrating an oblique fracture of the proximal humerus with increased lateral displacement (arrow).

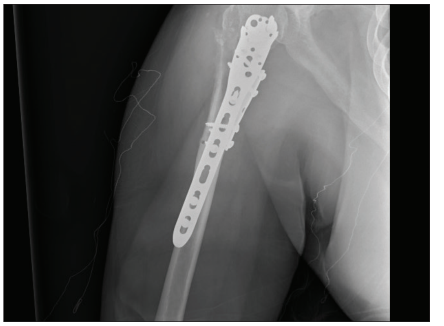


FIGURE 3. Anteroposterior internal rotation radiograph of the right shoulder following plate and screw fixation.

Proximal Humerus Fracture Following Arthroscopic Biceps Tenodesis

AIMEE N. JACOBS, PT, DPT, CSCS, *US Army, Tripler Army Medical Center, HI.*
JON A. UMLAUF, PT, DPT, DSc, FAAOMPT, *US Army, Fort Sam Houston, TX.*
JOSHUA R. KNISS, PT, DPT, OCS, *US Army, Fort Campbell, KY.*

A 60-YEAR-OLD WOMAN WITH CHRONIC atraumatic shoulder pain underwent arthroscopic biceps tenodesis, subacromial decompression, and distal clavicle excision. Upon presenting to the physical therapy clinic 7 days following surgery, she reported constant pain of 7/10 on the numeric pain-rating scale (NPRS). Her prior medical history included type 2 diabetes, hypertension, and osteoporosis. Physical examination revealed extensive ecchymosis, extending from her chest to her elbow, and significant pain that limited passive shoulder flexion and abduction to 10°.

Following the examination, the physical therapist reviewed the patient's postoperative radiographs, noted a comminuted but minimally displaced fracture of the right proximal humeral metaphysis (**FIGURE 1**), and immediately contacted the orthopaedic surgeon. The surgeon evaluated the patient and prescribed nonoperative fracture management with a sling and no active or passive glenohumeral motion for 6 weeks. After 6 weeks, the patient resumed supervised biweekly outpatient physical therapy treatment, with well-tolerated motion and strengthening. Within the 6- to 14-week postoperative time frame, her pain (NPRS) ranged from 3/10 at rest to 8/10 with sudden movements and at night. At 14 weeks post surgery, radiographs revealed increased displacement of the fracture (**FIGURE 2**). Due to failing conservative management, the patient underwent open reduction and internal fixation (**FIGURE 3**). She achieved union of the fracture without hardware complication 6 months post surgery.

Proximal humerus fractures are a rare complication from biceps tenodesis, with a reported incidence of less than 0.1%.[3] Conservative management and surgical fixation for displaced proximal humerus fractures yield the same functional outcomes in older adults.[1] Nonunion is reported in 10% to 20% of patients, but that risk is higher with the presence of known risk factors, including diabetes.[2] Despite the low prevalence, clinicians should consider surgical complications as a differential diagnosis when a patient's postoperative recovery fails to improve as expected and refer for imaging to assess for these complications. ◉ *J Orthop Sports Phys Ther 2020;50(11):649. doi:10.2519/jospt.2020.9497*

References
1. Beks RB, Ochen Y, Frima H, et al. Operative versus nonoperative treatment of proximal humeral fractures: a systematic review, meta-analysis, and comparison of observational studies and randomized controlled trials. *J Shoulder Elbow Surg*. 2018;27:1526-1534. https://doi.org/10.1016/j.jse.2018.03.009
2. Hernandez RK, Do TP, Critchlow CW, Dent RE, Jick SS. Patient-related risk factors for fracture-healing complications in the United Kingdom General Practice Research Database. *Acta Orthop*. 2012;83:653-660. https://doi.org/10.3109/17453674.2012.747054
3. Overmann AL, Colantonio DF, Wheatley BM, Volk WR, Kilcoyne KG, Dickens JF. Incidence and characteristics of humeral shaft fractures after subpectoral biceps tenodesis. *Orthop J Sports Med*. 2019;7:2325967119833420. https://doi.org/10.1177/2325967119833420

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 or Department of Defense.

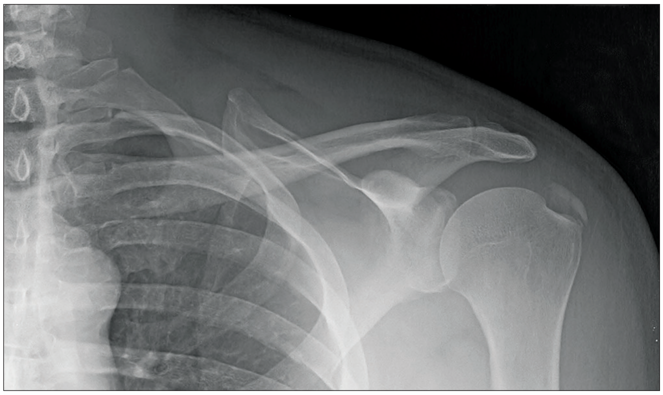


FIGURE 1. Anteroposterior external rotation radiograph of the left shoulder, demonstrating mild acromioclavicular joint arthropathy and extensive dense soft tissue calcification at the insertion site of the rotator cuff.

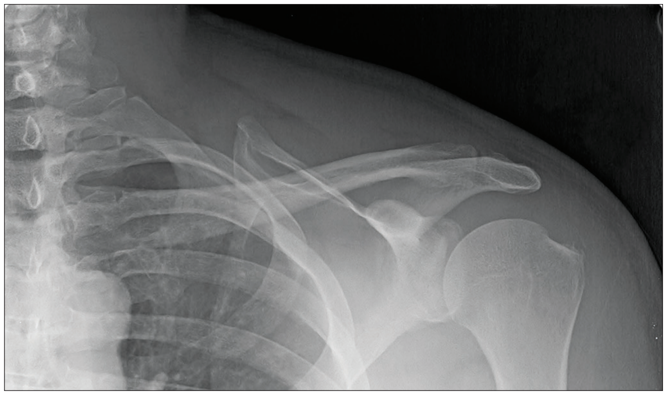


FIGURE 2. Anteroposterior external rotation radiograph of the left shoulder, demonstrating marked improvement in the previously noted soft tissue calcifications identified at the insertion site of the rotator cuff, with only trace amounts of calcification remaining.

Calcific Tendinopathy of the Rotator Cuff Treated With Acetic Acid Iontophoresis

MARYCRIS MEDINA-GANDIONCO, PT, DPT, *Physical Therapy Service, Brigadier General Crawford F. Sams US Army Health Clinic/Medical Department Activity-Japan, Camp Zama, Japan.*
ROBERT A. BRIGGS, PT, DPT, PhD, OCS, *Army-Baylor University/US Army Medical Center of Excellence, Fort Sam Houston, TX.*

A 62-year-old right hand–dominant man with a 20-year history of left shoulder pain and 6-month insidious, progressively worsening symptoms was referred to physical therapy with clinical and radiographic evidence of left supraspinatus tendon calcification (**FIGURE 1**). The patient enjoyed tennis, push-ups, and resistance training but had discontinued these activities because of his increasing shoulder symptom severity.

At evaluation, the patient reported pain ranging from 7/10 to 9/10 on the visual analog scale, and scored 54% on the Shoulder Pain and Disability Index (SPADI). Left shoulder active range of motion was limited to 60° of flexion and abduction, with empty end feel due to muscle guarding. Passively, the patient tolerated 90° of shoulder flexion with increasing pain.

Due to severity and irritability of pain, the patient was prescribed rest, isometrics, and pain-free range of motion. He was treated with iontophoresis, utilizing 3% acetic acid, for 3 visits per week for 5 weeks. Radiographs, repeated 1 week post treatment and 9 weeks since his initial radiographs, demonstrated marked resorption of the calcific deposit (**FIGURE 2**). At that point, the patient demonstrated 160° of active left shoulder flexion, 2/10 pain, and a SPADI score of 28% disability. A progressive resistive exercise program was initiated.

At 16 weeks post evaluation, the patient demonstrated full active range of motion of the left shoulder, 4+/5 shoulder girdle strength, a SPADI score of 5% disability, and reported a pain-free return to prior activities.

Calcific deposition, with accompanying shoulder pain, is common.[1-3] While understanding of specific physiological processes of calcification resorption remains incomplete, increased solubility of calcifications with acetic acid is theorized to be beneficial.[1,3] In this individual, concurrent resorption of the calcification and resolution of pain coincided with treatment including iontophoresis. However, we cannot conclude that iontophoresis with acetic acid was directly responsible for this outcome. This treatment modality appears to be a potentially beneficial, noninvasive treatment, warranting further consideration. ⦿ *J Orthop Sports Phys Ther 2020;50(11):650. doi:10.2519/jospt.2020.9270*

References

1. Fernández Cuadros ME, Pérez Moro OS, Álava Rabasa S, García González JM, Mirón Canelo JA. Calcifying tendonitis of the shoulder: risk factors and effectiveness of acetic acid iontophoresis and ultrasound. *Middle East J Rehabil Health Stud*. 2016;3:e41112. https://doi.org/10.17795/mejrh-41112
2. Fernández-Cuadros ME, Pérez-Moro OS, Diez-Ramos F, Albaladejo-Florin MJ. Acetic acid iontophoresis and ultrasound effectiveness on calcifying tendonitis of the shoulder, elbow, wrist, hip, knee and ankle: a non-randomized multicenter control trial. *Orthop Rheum Open Access J*. 2016;3:555611. https://doi.org/10.19080/OROAJ.2016.03.555611
3. Kachewar SG, Kulkarni DS. Calcific tendinitis of the rotator cuff: a review. *J Clin Diagn Res*. 2013;7:1482-1485. https://doi.org/10.7860/JCDR/2013/4473.3180

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, US Air Force, or Department of Defense.

ZOE A. MICHALEFF, PT, PhD[1] • JOSHUA R. ZADRO, PT, PhD[2,3] • ADRIAN C. TRAEGER, PT, PhD[2,3]
MARY O'KEEFFE, PT, PhD[2-4] • SIMON DÉCARY, PT, PhD[5-7]

Overcoming Overuse Part 2: Defining and Quantifying Health Care Overuse for Musculoskeletal Conditions

Health care overuse, commonly referred to as *overuse*, is a health service (clinic visit, test, or treatment) that provides no net benefit or causes harm to a patient or the wider population.[4,10] Overuse is typically considered a problem within medicine, and less so within ancillary health services. Between 10% and 30% of health care might be overuse. All treatments have the potential to cause harm in terms of physical, psychological, social, financial, and treatment burden.[4] Many physical therapists will understand the traditional definition of overuse. But what does overuse look like in practice? In part 2 of the Overcoming Overuse series, we (1) define overuse on a continuum from overuse to appropriate care, (2) consider how the definition of overuse depends on the perspective of the physical therapist, society, and the patient, and (3) discuss ways health care overuse can be measured.

A Continuum From Overuse to Appropriate Care

Clinical practice is complex, and quantifying health care as either overuse or appropriate in physical therapy is, as with other professions, not black and white. Appropriate care occurs along a continuum (**FIGURE**).[4] At one end of the continuum is overuse: care that is ineffective, inefficient (cost-effectiveness relative to alternatives), and misaligned with the patient's values and preferences.[11] At the other end is appropriate care: clearly effective (beneficial based on best available evidence), efficient, and aligned with the patient's values and preferences. Between the two extremes of overuse and appropriate care lies the "gray zone," the area in which most real-world practice is located, with all its subtleties and nuances. The "gray zone" includes tests or treatments that offer only small benefits; have incomplete or inconclusive evidence for benefits, harms, and cost-effectiveness; where the evidence is not generalizable to the patient; or where the patient's preferences don't align with best evidence. It is in the "gray zone" where defining, identifying, and measuring overuse are challenging.

Defining overuse depends on the perspective of the person viewing the problem.[11] Consumers, clinicians, health care institutions/organizations, policy makers, industry, and government likely all have different criteria when defining overuse and appropriate health care. In accordance with the framework proposed by Verkerk et al,[11] we consider overuse of musculoskeletal health care in terms of care that is ineffective, inefficient, and misaligned.[11] For each section, we include physical therapy–specific examples and encourage readers to reflect on their own practice (**TABLE 1**).

SUMMARY: In this series on "Overcoming Overuse," we explore the issue of health care overuse and how it may be identified in musculoskeletal physical therapy. In part 2, we frame health care overuse as a continuum from overuse to appropriate care, and consider how to measure overuse. We describe how overuse can be defined within a framework of care that is ineffective, inefficient, and misaligned, depending on the perspective of the person delivering or receiving care—the clinician, society, or patient. To ensure that musculoskeletal health care is of high value and sustainable, we encourage physical therapists to reflect on their practice. *J Orthop Sports Phys Ther 2020;50(11):588-591. doi:10.2519/jospt.2020.0109*

KEY WORDS: *appropriate care, health care, musculoskeletal, physical therapy, provider, services*

Health Care Overuse: Ineffective, Inefficient, and Misaligned Care

Ineffective care considers overuse from the physical therapist's perspective and focuses on evidence-based practice. Ineffective care includes any test or treatment that, based on high-quality evidence, provides

[1]Institute for Evidence-Based Healthcare, Faculty of Health Sciences and Medicine, Bond University, Robina, Australia. [2]School of Public Health, Sydney Medical School, The University of Sydney, Camperdown, Australia. [3]Institute for Musculoskeletal Health, Sydney Local Health District, Camperdown, Australia. [4]School of Allied Health, Faculty of Education and Health Sciences, University of Limerick, Limerick, Ireland. [5]Centre de recherche sur les soins et les services de première ligne de l'Université Laval, Quebec, Canada. [6]Canada Research Chair in Shared Decision Making and Knowledge Translation, Université Laval, Quebec, Canada. [7]Department of Family Medicine and Emergency Medicine, Faculty of Medicine, Université Laval, Quebec, Canada. No funding was received for this editorial. 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 Zoe A. Michaleff, Institute for Evidence-Based Healthcare, Faculty of Health Sciences and Medicine, Bond University, 14 University Drive, Robina, QLD 4226 Australia. E-mail: zmichale@bond.edu.au

little to no beneficial effect, is not cost-effective compared to available alternatives, or for which the risk of harm outweighs the probable benefit.[11] Examples include prolonged bed rest, electrotherapy, back supports, imaging, or injections for low back pain,[7] and arthroscopic surgery for degenerative knee disorders (eg, osteoarthritis) or rotator cuff–related shoulder pain.[3] Strategies aimed at reducing ineffective care target clinicians and include de-implementation initiatives (eg, the Choosing Wisely do-not-do recommendations, available at https://www.choosingwisely.org), audit and feedback, and multicomponent implementation strategies (eg, guidelines dissemination, peer comparison, and education).

Inefficient care can be summarized by the phrase "less is more." Inefficient care considers overuse from a societal perspective. It includes care that is de-

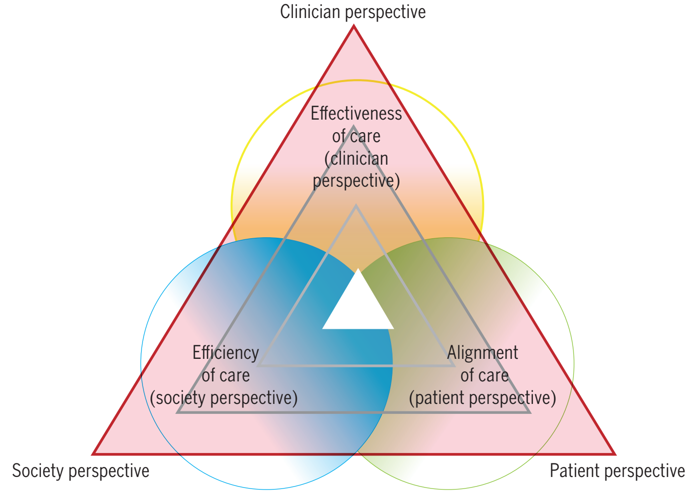


FIGURE. The continuum from overuse to appropriate care and the 3 perspectives to understand health care overuse (effectiveness, efficiency, and alignment of care). The small white triangle denotes appropriate care, gray shading denotes the "gray zone," and red shading denotes overuse. Dark yellow shading denotes effective care and light yellow shading denotes ineffective care, dark blue shading denotes efficient care and light blue shading denotes inefficient care, and dark green shading denotes aligned care and light green shading denotes misaligned care with the patient's values and preferences.

TABLE 1 Appropriate Care and Perceptions of Overuse From the Perspectives of the Patient, Physical Therapist, and Society

	Description
Case 1	A 58-year-old patient presents to a local physical therapist with a 3-week history of acute low back pain
Patient preferences	Electrotherapy and massage (based on advice from a friend)
Treatment delivered by physical therapist	Reassurance and advice to remain active
Treatment outcome	Pain resolved within 2 weeks of seeing the physical therapist
Clinician perspective	The physical therapist considers overuse to be care that best evidence suggests provides no benefit, or for which the risk of harms outweighs the probable benefit. The physical therapist provides care that is concordant with guidelines for acute low back pain (eg, advice to remain active and reassurance) but also uses spinal manipulative therapy, a treatment that falls within the "gray zone." Spinal manipulative therapy, when used judiciously over a short period and supported by strong clinical reasoning, may be considered appropriate care. The physical therapist avoids overuse by not providing the treatments requested by the patient and offering evidence-based alternatives
Society perspective	Insurance companies or policy makers consider overuse to be care that is not cost-effective. The insurance company weighs the costs and benefits of the treatment and considers whether alternatives are more cost-effective. For example, are there interventions that require fewer visits to health professionals, shorter interventions, or interventions that can be performed at home without supervision from a health professional?
Patient perspective	The patient considers overuse to be care that does not align with his or her values and preferences. The patient requests electrotherapy and massage, which he or she does not consider to be overuse, but instead is provided with reassurance, advice to remain active, and spinal manipulative therapy. From the patient's perspective, he or she has not received care that aligns with his or her values and preferences
Case 2	A 45-year-old patient presents to a physical therapist with an 8-year history of progressive knee pain and activity limitation due to osteoarthritis
Patient preferences	Advice on activity modifications and supervised exercise
Treatment delivered by physical therapist	Advice on activity modifications and 10 sessions of supervised exercise therapy plus provision of a home exercise program over 8 weeks
Treatment outcome	No difference in pain and disability at 8 weeks
Clinician perspective	The physical therapist provides care that is effective, according to randomized trials, and concordant with guidelines for knee osteoarthritis (ie, education and exercise therapy). Even though the patient did not respond to treatment, the physical therapist avoids overuse, from both a clinician and patient perspective
Society perspective	Physical therapy (inclusive of exercise and education) is cost-effective compared to enhanced physical therapy interventions (individually tailored and supervised exercise) and costs less than surgery. It is likely that the patient would be an appropriate candidate for surgery due to the ongoing symptoms and nonresponse to physical therapy. In this case, the care was not overuse from a society perspective
Patient perspective	The patient received care that aligned with his or her values and preferences, but his or her condition did not improve. The patient would probably consider the care to be appropriate, even though he or she experienced no benefit

livered in a way that increases costs without improving the patient's outcomes, particularly when compared to alternatives that involve lower treatment volume (ie, frequency/duration), are less complex/invasive, or can be delivered in less costly settings (**TABLE 2**).[10,11] Inefficient care may occur where care provision is not contingent on outcomes and is poorly coordinated between health care providers. Strategies aimed at reducing inefficient care target policy and include removing coverage for low-value care, reorganizing care pathways, and improving communication between health care providers.[10,11]

Misaligned care considers overuse from the patient's perspective and is care that does not align with the patient's values and preferences.[11] That is, a mismatch exists between care that is recommended in clinical practice guidelines and care that aligns with the patient's perspective.[5] This mismatch highlights the need to listen and understand the patient's perspective for 2 reasons: (1) the patient's values and preferences can act as a driver of overuse (**TABLE 1**) and will be explored further in part 3 of the series, and (2) misaligned care provides the opportunity to engage the patient as an active collaborator—especially in scenarios where care includes treatments that fall within the "gray zone" (**FIGURE**).[5] Shared decision making (further explored in part 5 of the series) is a strategy to engage the patient in a discussion about treatment decisions and, in turn, to overcome overuse.

Quantifying Health Care Overuse

Currently, measuring overuse is limited by a lack of systematic collection of detailed patient-level data.[10] Many systems lack data related to clinical decision making (ie, why a specific treatment was delivered) and patient preferences.[10] This level of detail is necessary to determine the appropriateness of care.

Approaches to measuring overuse are classified as direct or indirect.[4,10] Direct measurement includes use of medical registries or patient records to determine the specific care provided and patient outcomes. For example, audits of clinical records of people with acute low back pain show that approximately 70% of physical therapists provide appropriate care, including advice to keep active, and that 16% may overuse ineffective electrotherapy modalities.[12] In the absence of direct measures, indirect measures can identify potential areas of overuse, such as variations in health care delivery within and between countries or regions that are not attributable to differences in the populations or health systems.[4] Indirect measurement includes the use of quality indicators from primary care and hospitals (eg, administrative data or sur-

TABLE 2 EXAMPLES OF INEFFICIENT CARE IN PHYSICAL THERAPY, DESCRIBED IN TERMS OF VOLUME, COST, COMPLEXITY, AND CARE SETTING

Inefficiency/Condition	Study Design, Sample	Highly Inefficient Option	Efficient Alternative Option	Outcomes
Volume (intensity, duration)				
Chronic whiplash[6]	RCT n = 172	20 × 1-h individually tailored and supervised exercise sessions over 12 wk	1 × 30-min advice session and option of telephone support	No significant between-group difference for pain, disability, and range of motion at 14 wk, 6 mo, and 12 mo
Cost				
Early rehabilitation after lumbar disc surgery[8]	RCT n = 169	1-2 × 30-min individual, physical therapist–led exercise therapy sessions over 6-8 wk	No treatment	Cost utility (societal perspective): no significant between-group difference for any clinical outcome, quality-adjusted life-years, or societal costs at 26 wk
Complex/invasive				
Uncomplicated boxer's fracture (neck of fifth metacarpal)[9]	RCT n = 97	Plaster cast immobilization	Buddy taping of the ring and little fingers	No significant between-group difference for hand function, pain, satisfaction, return to sport, or health-related quality of life at 12 wk. Patients in the buddy taping group had a shorter length of stay in the emergency department and returned to work faster
Degenerative knee disorders (eg, degenerative meniscal tears)[3]	SR n = 13 RCTs	Arthroscopic knee surgery (including debridement and/or partial meniscectomy)	Nonsurgical management (exercise therapy, injections, medication)	Moderate- to high-quality evidence that arthroscopic knee surgery has a very small short-term (3 mo) benefit on pain, function, and quality of life compared to conservative management. In the long term (2 y), no significant between-group difference was found for pain or function
Care setting				
Rehabilitation following knee arthroplasty[1]	SR n = 6 RCTs	Outpatient physical therapy (eg, 2 × 1-h sessions per week for 2-12 wk)	Physical therapy provided in the home (including home exercise, telerehabilitation, home visits)	Moderate- to high-quality evidence of no significant difference in pain and function between outpatient physical therapy and home-based exercise

Abbreviations: RCT, randomized controlled trial; SR, systematic review.

veys of patients/clinicians to identify the type and amount of care delivered).[12] The Australian Atlas of Healthcare Variation uses indirect measures to demonstrate regional variations in the use of surgery for musculoskeletal conditions (eg, knee replacements, spinal decompression, and fusion).[2] Improving our ability to identify and measure overuse is critical to progress.[10]

Am I Contributing to or Reducing Health Care Overuse?

We encourage readers to reflect on their practice from the 3 perspectives of health care overuse (clinician, society, and patient) and consider to what degree their practice is helping to overcome this problem. As physical therapists, if we are aware of factors that may contribute to overuse, reflect on our practice, and aim to deliver treatments considered appropriate from multiple perspectives (**FIGURE**), we are heading toward overcoming overuse. Delivering care that is effective, efficient, and aligns with the patient's values and preferences will ensure that physical therapists remain leaders in managing musculoskeletal conditions. ◉

STUDY DETAILS

AUTHOR CONTRIBUTIONS: All authors conceived the idea. Dr Michaleff wrote the first draft. All authors contributed intellectual content, assisted with revisions, and approved the final version of this manuscript.

DATA SHARING: There are no data in this manuscript.

PATIENT AND PUBLIC INVOLVEMENT: Patients and the public were not involved in this editorial.

REFERENCES

1. Artz N, Elvers KT, Lowe CM, Sackley C, Jepson P, Beswick AD. Effectiveness of physiotherapy exercise following total knee replacement: systematic review and meta-analysis. *BMC Musculoskelet Disord*. 2015;16:15. https://doi.org/10.1186/s12891-015-0469-6
2. Australian Commission on Safety and Quality in Health Care. The Second Australian Atlas of Healthcare Variation. Sydney, Australia: Australian Commission on Safety and Quality in Health Care; 2017.
3. Brignardello-Petersen R, Guyatt GH, Buchbinder R, et al. Knee arthroscopy versus conservative management in patients with degenerative knee disease: a systematic review. *BMJ Open*. 2017;7:e016114. https://doi.org/10.1136/bmjopen-2017-016114
4. Brownlee S, Chalkidou K, Doust J, et al. Evidence for overuse of medical services around the world. *Lancet*. 2017;390:156-168. https://doi.org/10.1016/S0140-6736(16)32585-5
5. Lim YZ, Chou L, Au RT, et al. People with low back pain want clear, consistent and personalised information on prognosis, treatment options and self-management strategies: a systematic review. *J Physiother*. 2019;65:124-135. https://doi.org/10.1016/j.jphys.2019.05.010
6. Michaleff ZA, Maher CG, Lin CW, et al. Comprehensive physiotherapy exercise programme or advice for chronic whiplash (PROMISE): a pragmatic randomised controlled trial. *Lancet*. 2014;384:133-141. https://doi.org/10.1016/S0140-6736(14)60457-8
7. NSW Agency for Clinical Innovation. Management of People With Acute Low Back Pain: Model of Care. Chatswood, Australia: New South Wales Ministry of Health; 2016.
8. Oosterhuis T, Ostelo RW, van Dongen JM, et al. Early rehabilitation after lumbar disc surgery is not effective or cost-effective compared to no referral: a randomised trial and economic evaluation. *J Physiother*. 2017;63:144-153. https://doi.org/10.1016/j.jphys.2017.05.016
9. Pellatt R, Fomin I, Pienaar C, et al. Is buddy taping as effective as plaster immobilization for adults with an uncomplicated neck of fifth metacarpal fracture? A randomized controlled trial. *Ann Emerg Med*. 2019;74:88-97. https://doi.org/10.1016/j.annemergmed.2019.01.032
10. Scott IA, Duckett SJ. In search of professional consensus in defining and reducing low-value care. *Med J Aust*. 2015;203:179-181. https://doi.org/10.5694/mja14.01664
11. Verkerk EW, Tanke MAC, Kool RB, van Dulmen SA, Westert GP. Limit, lean or listen? A typology of low-value care that gives direction in de-implementation. *Int J Qual Health Care*. 2018;30:736-739. https://doi.org/10.1093/intqhc/mzy100
12. Zadro J, O'Keeffe M, Maher C. Do physical therapists follow evidence-based guidelines when managing musculoskeletal conditions? Systematic review. *BMJ Open*. 2019;9:e032329. https://doi.org/10.1136/bmjopen-2019-032329

[LITERATURE REVIEW]

BRUNO T. SARAGIOTTO, PhD[1-3] • IURI FIORATTI, PT[1,3] • ANNE TIEDEMANN, PhD[3]
MARK J. HANCOCK, PhD[4] • TIÊ P. YAMATO, PhD[1,2] • SAMUEL S.Y. WANG, BMed, MD[5]
JOSEPHINE Y. CHAU, PhD[4,6] • CHUNG-WEI CHRISTINE LIN, PhD[3]

The Effectiveness of Strategies to Promote Walking in People With Musculoskeletal Disorders: A Systematic Review With Meta-analysis

OBJECTIVE: To evaluate the effect of walking promotion strategies on physical activity, pain, and function in people with musculoskeletal disorders.

DESIGN: Intervention systematic review with meta-analysis.

LITERATURE SEARCH: We performed the searches in MEDLINE, Embase, the Cochrane Central Register of Controlled Trials, and the Physiotherapy Evidence Database (PEDro) from inception to August 2019.

STUDY SELECTION CRITERIA: We included randomized controlled trials evaluating interventions that promote walking in people with musculoskeletal disorders.

DATA SYNTHESIS: We used the PEDro scale for assessing risk of bias and the Grading of Recommendations Assessment, Development and Evaluation approach to evaluate the quality of evidence. We expressed pooled effects for between-group differences as mean differences or standardized mean differences and 95% confidence intervals, or as risk ratios and 95% confidence intervals, using random-effects meta-analyses.

RESULTS: Twelve eligible trials (n = 1456 participants) were identified. There was moderate- to very low–quality evidence of no difference in physical activity levels for walking promotion interventions when compared to minimal interventions, and a significant effect favoring walking promotion when compared with usual care in the short term. There was moderate-quality evidence that walking promotion was modestly effective for reducing pain and improving function compared with minimal intervention and usual care. There was no difference in pain and function for walking promotion compared to supervised exercise. Walking promotion was not associated with different rates of adverse events compared to control conditions.

CONCLUSION: Strategies to promote walking did not increase physical activity in people with musculoskeletal disorders. Walking promotion was associated with small improvements in pain and function compared to minimal intervention and usual care. *J Orthop Sports Phys Ther 2020;50(11):597-606. doi:10.2519/jospt.2020.9666*

KEY WORDS: *chronic pain, health promotion, physical activity*

Physical inactivity is associated with over 5 million deaths worldwide each year, accounting for the fourth leading risk factor for mortality.[18,30] Physical inactivity is responsible for up to 10% of the burden of the most common noncommunicable diseases, including coronary heart disease, diabetes, and cancer.[18] International guidelines recommend that adults do at least 150 minutes of moderate-intensity or 75 minutes of vigorous-intensity physical activity per week, or a combination of both.[31] However, a large proportion of the population fails to achieve the recommendation. The global prevalence of insufficient activity in adults is estimated at 23% (range, 4.1%-65%), but the prevalence is up to 70% in people with musculoskeletal disorders.[7,16]

[1]Masters and Doctoral Programs in Physical Therapy, Universidade Cidade de São Paulo, Tatuapé, Brazil. [2]Centre for Pain, Health and Lifestyle, Australia. [3]Institute for Musculoskeletal Health, School of Public Health, Faculty of Medicine and Health, The University of Sydney, Camperdown, Australia. [4]Faculty of Medicine, Health and Human Sciences, Macquarie University, Macquarie Park, Australia. [5]Prince of Wales Clinical School, Faculty of Medicine, University of New South Wales, Kensington, Australia. [6]Prevention Research Collaboration, Sydney School of Public Health, Charles Perkins Centre, The University of Sydney, Camperdown, Australia. This review was registered with PROSPERO (registration number CRD42017073641). No direct funding was received for this manuscript. Dr Saragiotto reports grants from the São Paulo Research Foundation outside the submitted work. Dr Chau reports grants from the National Heart Foundation of Australia, personal fees from the World Health Organization, personal fees from Bill Bellew Consulting Associates, and nonfinancial support from Marsh Australia outside the submitted work. Dr Lin reports nonfinancial support from Pfizer Australia outside the submitted work, and holds a Career Development Fellowship from the National Health and Medical Research Council (Australia). 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 Bruno T. Saragiotto, Masters and Doctoral Programs in Physical Therapy, Universidade Cidade de São Paulo, Rua Cesário Galeno, 448/475, Tatuapé, SP 03071-000 Brazil. E-mail: bruno.saragiotto@gmail.com

People with musculoskeletal disorders (eg, osteoarthritis, back and neck pain) usually report 2 main barriers to engaging in more physical activity: pain and the belief that physical activity will worsen their symptoms.[3] Kinesiophobia and anxiety may facilitate higher disability and a greater risk of other noncommunicable diseases. Strategies that include walking promotion are effective in increasing physical activity in the general population and in improving overall health.[13,26] However, it is unclear whether walking promotion is effective for people with musculoskeletal disorders.

Walking is a low-cost, easily accessible, and feasible form of physical activity, requiring no previous training or equipment.[2,17] Promoting walking to people with musculoskeletal disorders may improve physical activity levels, improve functional status, and reduce pain intensity. Increasing physical activity by encouraging people to walk more may be a feasible approach to reducing the burden of musculoskeletal disorders and a simple strategy to implement in a contemporary public health policy.[26] We aimed to evaluate the effectiveness of walking promotion strategies on physical activity and clinical outcomes (ie, pain, disability, and adverse events) in people with musculoskeletal disorders.

METHODS

This systematic review is reported according to the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines.[19] The review protocol was registered prospectively in the PROSPERO database (CRD42017073641).

Data Sources and Searches

We searched the Cochrane Central Register of Controlled Trials, Embase (via Ovid), MEDLINE (via Ovid), and Physiotherapy Evidence Database (PEDro) databases, with no restriction on the year of publication or language. Searches were conducted in December 2017 and updated in August 2019. We also searched the reference lists of relevant systematic reviews and included studies. Two reviewers (B.S. and T.Y. or S.W. or I.F.) independently screened titles and abstracts for potentially eligible studies, then examined the full texts of potentially relevant papers for inclusion. Disagreements were resolved through discussion or by a third reviewer if required (C.L. or M.H.). The search strategy for the MEDLINE database is outlined in **APPENDIX A** (available at www.jospt.org).

Study Selection

We included randomized controlled trials evaluating interventions to promote walking (eg, not supervised walking), delivered by any mode (eg, phone consultation, face-to-face advice, or internet based) in adults (18 years of age or older) with musculoskeletal disorders, identified according to clinical judgment or accepted diagnostic criteria. Musculoskeletal disorders included any condition related to the musculoskeletal system (eg, low back pain, osteoarthritis, shoulder pain, fibromyalgia) but excluded conditions due to (suspected) serious pathology (eg, cancer, cauda equina) or conditions that require urgent medical management (eg, fracture). We only considered trials that included mixed clinical populations when separate data for participants with musculoskeletal disorders could be obtained. We included trials that recruited participants from primary, secondary, or tertiary care who either sought care or were recruited from the community. We compared promotion of walking with no treatment, minimal intervention (eg, brief advice, education), usual care, and any other active treatment not aiming to promote walking or physical activity (eg, supervised exercise).

Outcome Measures

The primary outcome was physical activity or walking volume (measured objectively or subjectively). The secondary outcomes were pain intensity, physical function, and adverse events.

Data Extraction and Risk of Bias Assessment

Data regarding trial characteristics and estimates of effect were extracted by 2 reviewers (B.S. and T.Y. or S.W. or I.F.) using a pilot-tested data-extraction form, and any disagreements were resolved by discussion or arbitration by a third reviewer (C.L. or M.H.) if required.

Risk of bias of the included studies was assessed using the 11-point PEDro scale, in which higher scores represent lower risk of bias. The PEDro scale has acceptable reliability[21] and validity,[6,20] and there is evidence that the scale is reliable across disciplines and correlates with other risk of bias tools.[32] If the included study was indexed in PEDro (www.pedro.org.au), the PEDro score was downloaded directly from the database. If the included study was not indexed in PEDro, 2 independent and trained reviewers performed the assessment, with disagreements resolved by discussion or arbitration by a third reviewer.

Missing Data

We contacted authors to verify key study characteristics and obtain missing numerical outcome data. One trial[14] did not report means and standard deviations for pain intensity, and the authors provided the data. Missing standard deviations were computed from other statistics, such as standard errors, confidence intervals (CIs), or *P* values, using the Review Manager calculator (https://training.cochrane.org/resource/revman-calculator) if necessary. One trial[4] reported only median and range, so we estimated mean and standard deviation using the approximation method described by Wan et al.[29] We prioritized data extraction from intention-to-treat analyses over per-protocol or as-treated analyses.

Data Synthesis and Analysis

Where similar outcome measures were used, we expressed pooled effects of continuous variables as mean difference and 95% CI for between-group differences. Where continuous outcome measures were different between studies, we used standardized mean difference (SMD) and 95% CI. For dichotomous outcomes, we used risk ratio and 95% CI. We pooled effect estimates using a random-effects model for the following periods: short-term (less than 3 months post randomization), intermediate-term (3 months to less than 12 months post randomization), and long-term (12 months or more post randomization) follow-up. Pain scales were converted into a 0-to-100 scale and expressed as mean difference and 95% CI. Effect sizes were quantified as small (SMD, less than 0.20; mean difference, less than 10% of the scale), medium (SMD, 0.20-0.50; mean difference, 10%-20% of the scale), or large (SMD, greater than 0.50; mean difference, greater than 20% of the scale).[5,27] Meta-analyses were performed using Review Manager 5.4 (The Nordic Cochrane Center, Copenhagen, Denmark).

We used the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach for the assessment of quality of evidence, which assesses 5 domains. For each domain that was not met, we reduced the quality of evidence by 1 level: (1) study design and risk of bias: the evidence was downgraded if more than 25% of the participants were from trials with a high risk of bias (ie, PEDro score less than 6); (2) inconsistency of results: the evidence was downgraded if there was significant heterogeneity on visual inspection of the forest plot (ie, minimal or no overlap of CIs) or I^2 statistics (greater than 50% may indicate substantial heterogeneity); (3) imprecision: the evidence was downgraded if fewer than 400 participants were included in the comparison for continuous data and there were fewer than 300 events for dichotomous data[24]; (4) indirectness (generalizability of the findings): the evidence was downgraded if greater than 50% of the participants were outside the target group; and (5) other bias (eg, publication bias).

We planned a subgroup analysis of features of the risk of bias assessment (allocation concealment and use of intention-to-treat analysis) and specific populations (eg, rheumatoid arthritis, low back pain, fibromyalgia) by calculating the effect sizes for trials that met the criterion.

RESULTS

The database search retrieved 4541 records after duplicates were removed. After screening titles, abstracts, and full texts, we included 12 randomized controlled trials (n = 1456 participants) (**FIGURE 1**).[1,2,4,8-12,14,15,23,28]

Characteristics of Included Trials

Participants The 12 included trials (**TABLE 1**) were from 8 countries (United States, New Zealand, Sweden, Australia, Denmark, Japan, Ireland, and United Kingdom). Total sample sizes ranged from 33 to 246 participants, with the mean age of trial participants ranging from 45 to 73 years. The trials included patients with fibromyalgia,[1,8] knee or hip osteoarthritis,[9,11,28] rheumatoid arthritis,[2,4,14] and chronic low back pain.[10,12,15,23]

Intervention Eight trials[2,10-12,14,15,23,28] included interventions that exclusively promoted walking. The interventions in 4 trials[1,4,8,9] also promoted general aerobic exercise. The content of the interventions was similar among trials, including advice and education on the benefits of walking and physical activity, motivation to perform more steps per day, and goal setting. Six trials[2,4,9,12,15,23] included an individualized prescription of walking for the participants to perform during the study period. Ten trials delivered in-person interventions,[2,4,8-12,14,23,28] 1 was a telephone intervention,[1] and 1 delivered the intervention by a study-specific website.[15] Eleven trials[1,2,8-12,14,15,23,28] used

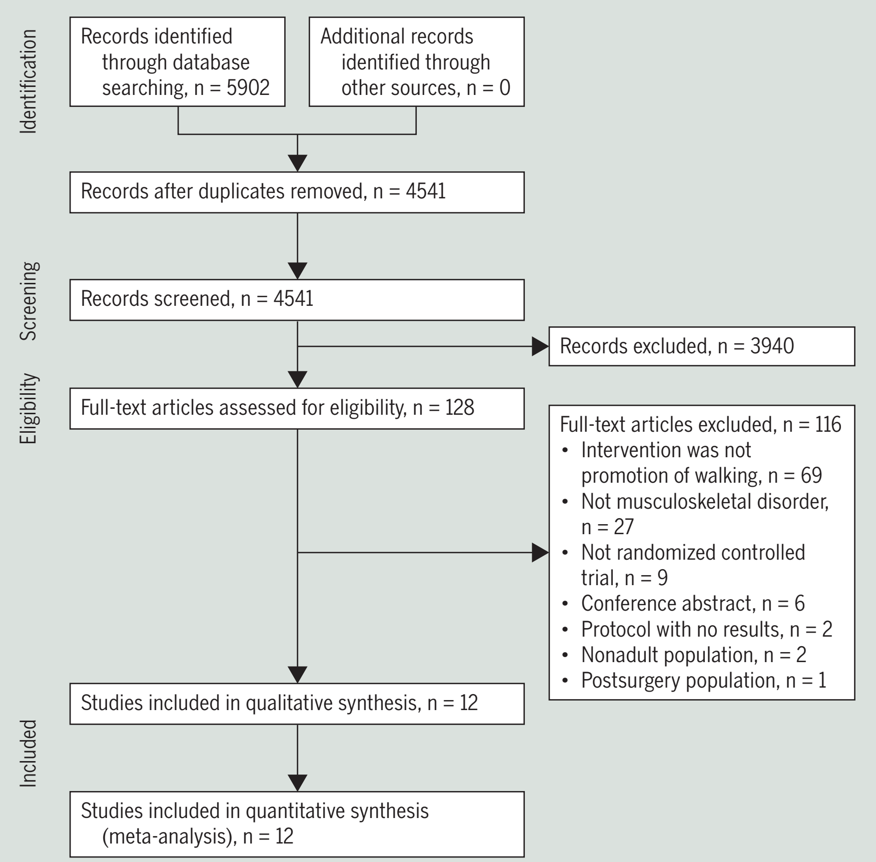


FIGURE 1. Flow chart of the inclusion process of the review.

pedometer or accelerometer devices to dose the intervention and provide feedback to participants. Walking promotion was compared with usual care (3 trials), minimal interventions (8 trials), and supervised exercises (2 trials). Two trials included more than 2 comparison groups (minimal intervention and supervised exercise), and we included both comparisons in the review.

Outcome Measures The most common outcome measures of walking or physical activity levels were step count, time spent in activity, and frequency of activities recorded using a diary or accelerometer/pedometer device. Measures of walking and physical activity level were collected in 11 trials, and 1 trial did not report between-group differences at follow-up. All the included trials measured pain intensity using the visual analog scale (0-100) or the numeric rating scale (0-10). Measures of function were reported by 11 trials and included a range of functional scales specific to each condition (eg, Fibromyalgia Impact Questionnaire, Western Ontario and McMaster Universities Osteoarthritis Index for osteoarthritis, Health Assessment Questionnaire disability index for rheumatoid arthritis, Roland-Morris Disability Questionnaire for low back pain, and Oswestry Disability Index for low back pain).

Risk of Bias The main limitations included lack of assessor blinding in 9 trials and absence of allocation concealment in

TABLE 1 Characteristics of the Included Trials

Study, Country	Sample Characteristics[a]	Intervention	Comparison	Follow-up
Ang et al[1] United States	216 participants with FM Female, n = 207 (96%) Age, 45.9 ± 11.2 y	Motivational interviewing by a health practitioner (6 telephone calls over a 12-wk period)	Didactic health information	12 wk, 3 mo, and 6 mo
Baxter et al[2] New Zealand	33 participants with RA Sex NR Age, 63 ± 11.5 y	Self-directed walking program (1 face-to-face session, then participants were encouraged to do the program 3-4 times per week for 6 wk)	1 appointment of nutritional advice	6 wk
Brodin et al[4] Sweden	228 participants with early RA Female, n = 169 (74%) Age, 55 ± 14 y	1-y coaching program including a prescription of aerobic exercise (delivered by physical therapists and trainers)	Education	3, 6, 9, and 12 mo
Fontaine et al[8] United States	73 participants with FM Female, n = 70 (96%) Age, 47.2 ± 11.1 y	Group-based motivational sessions to find ways to help participants increase physical activity levels	FM education	12 wk, 6 mo, and 12 mo
Halbert et al[9] Australia	69 participants with OA Female, n = 41 (59%) Age, 69 ± 5.9 y	Individualized physical activity advice	Nutrition advice pamphlet	3, 6, and 12 mo
Hartvigsen et al[10] Denmark	136 participants with chronic LBP Female, n = 98 (72%) Age, 46.7 ± 11 y	Instructions on Nordic walking	1. Advice to remain active and maintain daily function level 2. Supervised Nordic walking	2, 6, and 12 mo
Hiyama et al[11] Japan	40 participants with knee OA All female Age, 72.9 ± 5.45 y	Walking group (instructed to increase the number of steps per day)	Usual care (conventional physical therapy once per week)	4 wk
Hurley et al[12] Ireland	246 participants with chronic LBP Female, n = 167 (68%) Age, 45.4 ± 11.4 y	Individualized walking prescription	1. Usual physical therapy 2. Exercise classes	6 mo
Katz et al[14] United States	96 participants with RA Female, n = 84 (88%) Age, 54.8 ± 13.4 y	Pedometer, individualized daily step targets, and educational booklet	Education	10 and 21 wk
Krein et al[15] United States	229 participants with chronic LBP Female, n = 54 (24%) Age, 51.5 ± 12.6 y	Website walking promotion	Usual care	6 and 12 mo
McDonough et al[23] United Kingdom	56 participants with chronic LBP Female, n = 31 (55%) Age, 48.9 ± 6.2 y	Education/advice plus an 8-wk pedometer-driven walking program	Education/advice	9 wk and 6 mo
Talbot et al[28] United States	34 participants with knee OA Female, n = 26 (76.5%) Age, 70.2 ± 5.7 y	Home-based pedometer-driven walking program	12 h of the Arthritis Self-Management Program	12 wk and 6 mo

Abbreviations: FM, fibromyalgia; LBP, low back pain; NR, not reported; OA, osteoarthritis; RA, rheumatoid arthritis.
[a]*Values for age are mean ± SD.*

7 trials (**TABLE 2**). None of the trials blinded patients or therapists.

Effect of Walking Promotion on Physical Activity Meta-analysis was only possible for the outcome of number of steps per day. Compared to minimal intervention, we found very low–quality evidence of no effect of walking promotion on number of steps per day (×1000) at short-term follow-up (mean difference, –0.78; 95% CI: –2.12, 0.55; 3 trials), moderate-quality evidence of no effect at intermediate-term follow-up (mean difference, –1.10; 95% CI: –2.77, 0.56; 2 trials), and very low–quality evidence of no effect at long-term follow-up (mean difference, –0.69; 95% CI: –2.21, 0.83; 1 trial) (**FIGURE 2**). The quality of evidence was downgraded due to risk of bias, inconsistency, and imprecision for short-term follow-up, downgraded due to imprecision for intermediate-term follow-up, and downgraded due to risk of bias, inconsistency, and imprecision for long-term follow-up.

Compared to usual care, walking promotion had very low–quality evidence of an effect at short-term follow-up (mean difference, –3.08; 95% CI: –4.03, –2.12; 1 trial), and very low–quality evidence of no effect at intermediate- and long-term follow-ups (mean difference, –0.69; 95% CI: –1.48, 0.11; 1 trial and mean difference, 0.08; 95% CI: –0.70, 0.85; 1 trial, respectively) (**TABLE 3**).

Three trials reported other outcomes related to walking and physical activity lev-

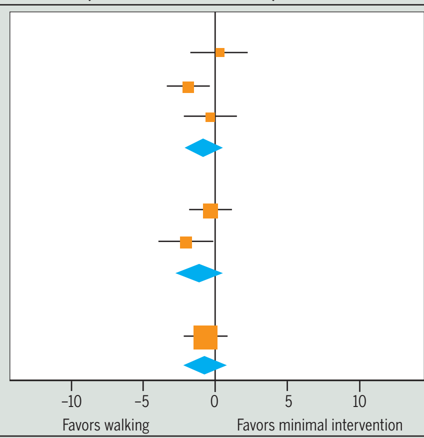
TABLE 2 Risk of Bias of Included Trials According to the PEDro Scale[a]

	Item[b]											
Study	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**Score**
Ang et al[1]	Yes	Yes	No	Yes	No	No	Yes	Yes	Yes	Yes	Yes	7
Baxter et al[2]	Yes	Yes	Yes	Yes	No	No	No	Yes	No	Yes	Yes	6
Brodin et al[4]	Yes	Yes	No	Yes	No	No	Yes	No	Yes	Yes	Yes	6
Fontaine et al[8]	No	Yes	No	Yes	No	No	No	No	No	Yes	Yes	4
Halbert et al[9]	Yes	Yes	No	Yes	No	No	No	Yes	No	Yes	Yes	5
Hartvigsen et al[10]	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	7
Hiyama et al[11]	Yes	Yes	Yes	Yes	No	No	No	Yes	No	Yes	Yes	6
Hurley et al[12]	Yes	Yes	No	Yes	No	No	Yes	No	Yes	Yes	Yes	6
Katz et al[14]	Yes	Yes	Yes	No	No	No	No	Yes	No	Yes	Yes	5
Krein et al[15]	Yes	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	7
McDonough et al[23]	Yes	Yes	No	Yes	No	No	No	Yes	Yes	Yes	Yes	6
Talbot et al[28]	Yes	Yes	No	Yes	No	No	No	Yes	No	Yes	Yes	5

Abbreviation: PEDro, Physiotherapy Evidence Database.
[a]*All quality scores were downloaded from the PEDro website. Item 1 is not included in the scoring calculation.*
[b]*Items: 1, Eligibility criteria were specified; 2, Subjects were randomly allocated to groups; 3, Allocation was concealed; 4, The groups were similar at baseline; 5, There was blinding of subjects; 6, There was blinding of therapists; 7, There was blinding of assessors; 8, There were measures of key outcomes from more than 85% of subjects; 9, Analyses were performed by intention-to-treat principles; 10, Between-group statistical comparisons were performed; 11, The study provided point measures and measures of variability.*

Steps Per Day (×1000)

	Walking		Minimal Intervention			
Subgroup/Study	**Mean ± SD**	**Total, n**	**Mean ± SD**	**Total, n**	**Weight**	**MD IV, Random (95% Confidence Interval)**
Short term						
Baxter et al[2]	–5.107 ± 2.676	11	–5.411 ± 2.896	22	28.6%	0.30 (–1.69, 2.30)
Fontaine et al[8]	–5.648 ± 3.564	30	–3.752 ± 1.897	23	39.8%	–1.90 (–3.39, –0.40)
Talbot et al[28]	–4.337 ± 2.903	17	–3.972 ± 2.563	17	31.6%	–0.36 (–2.21, 1.48)
Subtotal[a]		58		62	100.0%	–0.78 (–2.12, 0.55)
Intermediate term						
Fontaine et al[8]	–4.496 ± 3.228	30	–4.142 ± 2.286	23	56.3%	–0.35 (–1.84, 1.13)
Katz et al[14]	–6.675 ± 3.706	31	–4.609 ± 3.608	26	43.7%	–2.07 (–3.97, –0.16)
Subtotal[b]		61		49	100.0%	–1.10 (–2.77, 0.56)
Long term						
Fontaine et al[8]	–4.589 ± 3.19	30	–3.897 ± 2.46	23	100.0%	–0.69 (–2.21, 0.83)
Subtotal[c]		30		23	100.0%	–0.69 (–2.21, 0.83)



[a]*Heterogeneity: τ^2 = 0.58, χ^2 = 3.44, df = 2 (P = .18), I^2 = 42%. Test for overall effect: z = 1.15 (P = .25).*
[b]*Heterogeneity: τ^2 = 0.71, χ^2 = 1.93, df = 1 (P = .16), I^2 = 48%. Test for overall effect: z = 1.30 (P = .19).*
[c]*Heterogeneity: not applicable. Test for overall effect: z = 0.89 (P = .37).*
Abbreviations: IV, inverse variance; MD, mean difference.

FIGURE 2. Forest plot of the comparison between promotion of walking and minimal intervention for the outcome of steps per day.

el (**APPENDIX B**, available at www.jospt.org). Ang et al[1] found a significant effect of walking promotion programs on time spent in moderate-to-vigorous physical activity at 12-week follow-up (mean difference, 1.5 h/wk; 95% CI: 0.39, 2.60) when compared with health information, but did not find any intervention effect at intermediate- or long-term follow-up or for other physical activity outcomes (ie, minimum increment change in moderate-to-vigorous physical activity or accelerometer-based moderate-to-vigorous physical activity). Two trials[4,9] did not find any difference between walking promotion and minimal intervention for walking or physical activity outcomes.

The planned subgroup analyses for risk of bias and type of condition were not possible because we had insufficient trials per comparison.

Effect of Walking Promotion on Pain There was moderate-quality evidence of a small effect of walking promotion on reducing pain compared to minimal intervention at short-term follow-up (mean difference, –3.98; 95% CI: –7.16, –0.80; 8 trials). There was no effect for intermediate-term (mean difference, –1.60; 95% CI: –4.81, 1.62; 7 trials) and long-term follow-ups (mean difference, –1.92; 95% CI: –5.77, 1.92; 4 trials) (**FIGURE 3**). When compared to usual care, there was moderate-quality evidence of a small effect of walking promotion on reducing pain at intermediate-term follow-up (mean difference, –4.81; 95% CI: –9.46, –0.16; 2 trials), but no effect at the short- and long-term follow-ups (mean difference, –2.09; 95% CI: –4.91, 0.72 and mean difference, –1.39; 95% CI: –6.07, 3.29; 2 trials, respectively) (**TABLE 3**). The quality of evidence was downgraded due to imprecision.

When compared to supervised exercise, there was moderate-quality evidence of no effect on pain intensity at short- and intermediate-term follow-ups, and low-quality evidence of no effect for long-term follow-up (**TABLE 3**). The quality of the evidence was downgraded due to imprecision at short- and intermediate-term follow-ups, and downgraded due to inconsistency and imprecision for long-term follow-up.

Effect of Walking Promotion on Function There was moderate-quality evidence that walking promotion was superior to minimal intervention at short-term (SMD, –0.24; 95% CI: –0.47, –0.01; 5 trials), intermediate-term (SMD, –0.16; 95% CI: –0.33, 0.01; 6 trials), and long-term follow-ups (SMD, –0.20; 95% CI: –0.39, –0.01; 4 trials), with small effect sizes (**FIGURE 4**). When compared to usual care, there was moderate-quality evidence of an effect favoring walking promotion at intermediate-term follow-up (SMD, –0.30; 95% CI: –0.52, –0.08; 2 trials) and no differences at short- and long-term follow-ups, with low- and moderate-quality evidence, respectively (**TABLE 3**). The quality of evidence was downgraded due to imprecision and inconsistency at short-term follow-up, and downgraded due to imprecision at intermediate- and long-term follow-ups. When compared to supervised exercise, there was moderate-quality evidence of no effect of walking promotion on function (**TABLE 3**). The quality of evidence was downgraded due to imprecision.

Adverse Events Five trials[2,12,14,15,23] reported on minor adverse events for at least 1 of the groups. Meta-analysis was precluded, as only 1 trial reported adverse events for both groups, but no dif-

TABLE 3 SUMMARY OF FINDINGS AND QUALITY OF EVIDENCE FOR WALKING PROMOTION VERSUS USUAL CARE AND WALKING PROMOTION VERSUS SUPERVISED EXERCISE FOR THE OUTCOMES OF PAIN AND FUNCTION

Comparison, Outcome, Time Point	Participants, n (trials, n)	Effect Size[a]	Quality of Evidence (GRADE)
Walking promotion versus usual care			
Steps/d (×1000)			
Short term	40 (1)	–3.08 (–4.03, –2.12)	Very low
Intermediate term	229 (1)	–0.69 (–1.48, 0.11)	Very low
Long term	229 (1)	0.08 (–0.70, 0.85)	Very low
Pain			
Short term	137 (2)	–2.09 (–4.91, 0.72)	Moderate
Intermediate term	326 (2)	–4.81 (–9.46, –0.16)	Moderate
Long term	326 (2)	–1.39 (–6.07, 3.29)	Moderate
Function			
Short term	135 (2)	–0.17 (–0.86, 0.52)	Low
Intermediate term	324 (2)	–0.30 (–0.52, –0.08)	Moderate
Long term	324 (2)	–0.21 (–0.45, 0.03)	Moderate
Walking promotion versus supervised exercise			
Pain			
Short term	191 (2)	0.44 (–12.67, 13.56)	Moderate
Intermediate term	191 (2)	–1.99 (–14.48, 10.50)	Moderate
Long term	191 (2)	–1.74 (–17.98, 14.51)	Low
Function			
Short term	187 (2)	0.14 (–0.15, 0.42)	Moderate
Intermediate term	187 (2)	0.17 (–0.12, 0.46)	Moderate
Long term	187 (2)	0.11 (–0.18, 0.40)	Moderate

Abbreviation: GRADE, Grading of Recommendations Assessment, Development and Evaluation.
[a]Values for pain are mean difference (95% confidence interval) on a 0-to-100 scale, and values for function are standardized mean difference (95% confidence interval). A negative effect size represents an effect in favor of the walking promotion group.

ference was reported.[15] Adverse events of walking promotion included worsening pain, calluses, ankle sprain, and allergic reaction to the metal clip of the pedometer. No serious adverse events were reported.

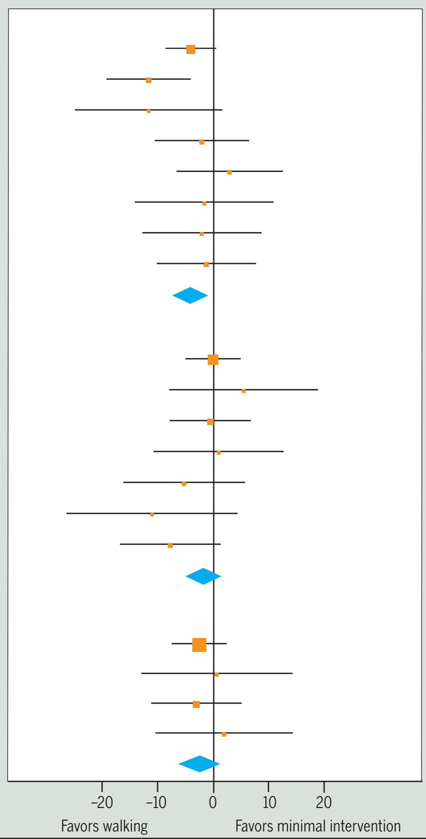
DISCUSSION

WE IDENTIFIED 12 TRIALS EVALUATING strategies to promote walking in people with musculoskeletal disorders. Strategies to promote walking did not increase walking level compared to minimal intervention, but they significantly increased short-term walking volume compared with usual care, based on very low–quality evidence. We found moderate-quality evidence that walking promotion may improve pain and function compared to minimal intervention and usual care.

Our systematic review was prospectively registered, followed the PRISMA recommendations, and included the GRADE system to appraise the quality of the evidence. We presented novel findings on walking promotion in people with musculoskeletal disorders. A previous review[22] reported that pedometer-driven walking programs were effective for promoting physical activity among patients with musculoskeletal disorders, contrary to our findings. However, meta-analysis was not conducted, and the authors did not judge the quality of the evidence. Further, the previous review included 7 trials, of which 6 are included in our review. Other reviews focused on supervised walking for different chronic pain conditions. The review by O'Connor et al[25] found significant improvements in pain and function in patients with chronic musculoskeletal pain when compared to various control conditions, such as education, usual care, and other forms of exercise. These findings corroborate our results, indicating that there may be some clinical benefits of walking.

Pain Intensity (0-100)

Subgroup/Study	Walking Mean ± SD	Walking Total, n	Minimal Intervention Mean ± SD	Minimal Intervention Total, n	Weight	MD IV, Random (95% Confidence Interval)
Short term						
Ang et al[1]	−12 ± 17.5	107	−8 ± 16.7	109	32.4%	−4.00 (−8.56, 0.56)
Baxter et al[2]	−30.7 ± 11.3	11	−19.1 ± 8.7	22	14.8%	−11.60 (−19.20, −4.00)
Fontaine et al[8]	48.4 ± 23.2	30	60 ± 25.4	23	5.4%	−11.60 (−24.89, 1.69)
Halbert et al[9]	20.5 ± 18.5	37	22.5 ± 17.5	32	12.2%	−2.00 (−10.50, 6.50)
Hartvigsen et al[10]	46 ± 23.5	46	43 ± 23.2	45	9.9%	3.00 (−6.60, 12.60)
Katz et al[14]	29.4 ± 24.9	31	31 ± 25.3	31	6.1%	−1.60 (−14.10, 10.90)
McDonough et al[23]	−9 ± 21.6	39	−7 ± 17.5	17	8.0%	−2.00 (−12.73, 8.73)
Talbot et al[28]	−1.2 ± 11.7	17	0 ± 14.7	17	11.2%	−1.20 (−10.13, 7.73)
Subtotal[a]		318		296	100.0%	−3.98 (−7.16, −0.80)
Intermediate term						
Ang et al[1]	−12 ± 18.6	107	−12 ± 18.7	109	41.8%	0.00 (−4.97, 4.97)
Fontaine et al[8]	54.9 ± 21	30	49.4 ± 27.1	23	5.8%	5.50 (−7.88, 18.88)
Halbert et al[9]	17 ± 16.5	37	17.5 ± 14.5	32	19.3%	−0.50 (−7.81, 6.81)
Hartvigsen et al[10]	41 ± 30.3	46	40 ± 26.6	45	7.5%	1.00 (−10.71, 12.71)
Katz et al[14]	19.9 ± 20.7	31	25.1 ± 23.3	31	8.6%	−5.20 (−16.17, 5.77)
McDonough et al[23]	−16 ± 30.8	39	−5 ± 25.3	17	4.3%	−11.00 (−26.43, 4.43)
Talbot et al[28]	−4 ± 10.7	17	3.7 ± 15.8	17	12.6%	−7.70 (−16.77, 1.37)
Subtotal[b]		307		274	100.0%	−1.60 (−4.81, 1.62)
Long term						
Brodin et al[4]	34.8 ± 18.8	94	37.3 ± 18.8	134	60.1%	−2.50 (−7.46, 2.46)
Fontaine et al[8]	51.6 ± 22	30	50.9 ± 27.2	23	8.0%	0.70 (−12.92, 14.32)
Halbert et al[9]	18.5 ± 18	37	21.5 ± 16.5	32	22.3%	−3.00 (−11.14, 5.14)
Hartvigsen et al[10]	44 ± 30.3	46	42 ± 29.9	45	9.7%	2.00 (−10.37, 14.37)
Subtotal[c]		207		234	100.0%	−1.92 (−5.77, 1.92)



[a]*Heterogeneity:* $\tau^2 = 2.69$, $\chi^2 = 8.00$, $df = 7$ ($P = .33$), $I^2 = 13\%$. *Test for overall effect:* $z = 2.46$ ($P = .01$).
[b]*Heterogeneity:* $\tau^2 = 0.00$, $\chi^2 = 5.33$, $df = 6$ ($P = .50$), $I^2 = 0\%$. *Test for overall effect:* $z = 0.97$ ($P = .33$).
[c]*Heterogeneity:* $\tau^2 = 0.00$, $\chi^2 = 0.65$, $df = 3$ ($P = .89$), $I^2 = 0\%$. *Test for overall effect:* $z = 0.98$ ($P = .33$).
Abbreviations: IV, inverse variance; MD, mean difference.

FIGURE 3. Forest plot of the comparison between promotion of walking and minimal intervention for the outcome of pain intensity.

It is unclear why strategies to promote walking are effective in increasing physical activity in the general population[26] but not consistently in people with musculoskeletal disorders. Physical activity strategies for people with musculoskeletal disorders may need to be different from those for the general population, due to barriers to physical activity participation such as fear of movement and catastrophizing behavior. Further, we included a wide range of strategies to promote walking (eg, group based, remotely delivered, advice, coaching, motivational interviews) and, due to the limited number of trials, we were unable to explore the impact of intervention dose, mode, and delivery type on physical activity outcomes. Our results were based on very low–quality evidence, so further studies are likely to change this conclusion.

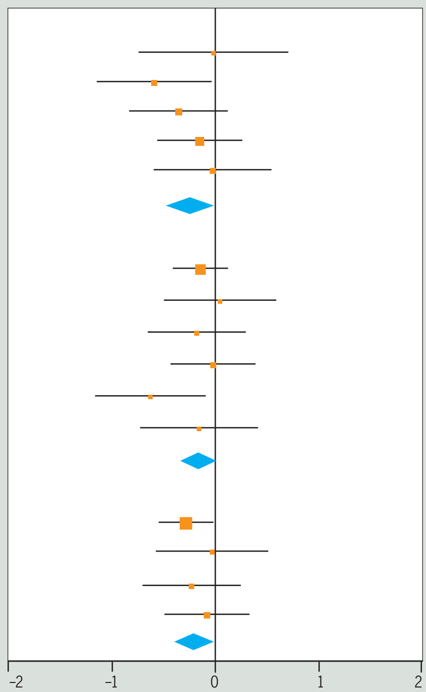
The mechanisms by which walking promotion strategies may improve clinical outcomes, despite not consistently increasing physical activity levels, in people with musculoskeletal disorders are unclear. One possibility is that the social component of the intervention had an effect on clinical outcomes (eg, perceived pain). While the comparison groups had minimal interaction with others (eg, pamphlet, self-management, 1 advice session), most interventions involved some sort of interaction with other people or health professionals over an extended period (eg, clinician, coaching, group sessions). None of the trials provided clear information or quantified the amount of social interaction.

Limitations

Due to the different measures reported for assessing the primary outcome (ie, physical activity or walking), meta-analysis was only possible for 1 outcome (number of steps per day). We were unable to explore the effects of different types of walking promotion interventions or features of risk of bias due to the small number of trials. It is possible that different strategies could elicit different effects. We considered a diverse range of musculoskeletal disorders, including multisystem disorders such as fibromyalgia and rheumatoid arthritis. Although this is a common approach, we were unable to identify whether the level of physical activity and walking differed among these types of conditions, and whether that may have affected our results. All included tri-

Function

Subgroup/Study	Walking Mean ± SD	Walking Total, n	Minimal Intervention Mean ± SD	Minimal Intervention Total, n	Weight	SMD IV, Random (95% Confidence Interval)
Short term						
Baxter et al[2]	0 ± 0.6	11	0.01 ± 0.6	22	10.3%	−0.02 (−0.74, 0.71)
Fontaine et al[8]	55.6 ± 18.7	30	67.1 ± 19.9	23	17.5%	−0.59 (−1.14, −0.03)
Halbert et al[9]	11.9 ± 9.6	37	15.8 ± 12.2	32	23.7%	−0.35 (−0.83, 0.12)
Hartvigsen et al[10]	42 ± 20.2	46	45 ± 19.9	45	31.9%	−0.15 (−0.56, 0.26)
McDonough et al[23]	26.4 ± 16.6	39	26.8 ± 14.3	17	16.6%	−0.02 (−0.59, 0.54)
Subtotal[a]		163		139	100.0%	−0.24 (−0.47, −0.01)
Intermediate term						
Ang et al[1]	−1.7 ± 2.1	107	−1.4 ± 2.1	109	40.8%	−0.14 (−0.41, 0.12)
Fontaine et al[8]	65 ± 17	30	64 ± 24.5	23	9.9%	0.05 (−0.50, 0.59)
Halbert et al[9]	11.3 ± 9.3	37	13.1 ± 10.8	32	13.0%	−0.18 (−0.65, 0.30)
Hartvigsen et al[10]	42.5 ± 25.2	46	43 ± 23.2	45	17.2%	−0.02 (−0.43, 0.39)
Katz et al[14]	−0.3 ± 0.5	31	−0.01 ± 0.4	26	10.2%	−0.63 (−1.16, −0.09)
McDonough et al[23]	23.7 ± 16.3	39	26.2 ± 14.9	17	8.9%	−0.16 (−0.73, 0.42)
Subtotal[b]		290		252	100.0%	−0.16 (−0.33, 0.01)
Long term						
Brodin et al[4]	0.75 ± 0.4	94	0.88 ± 0.5	134	50.9%	−0.28 (−0.55, −0.02)
Fontaine et al[8]	64.4 ± 20.9	30	65.1 ± 25.8	23	12.1%	−0.03 (−0.57, 0.51)
Halbert et al[9]	11.4 ± 10.7	37	13.8 ± 10.2	32	15.9%	−0.23 (−0.70, 0.25)
Hartvigsen et al[10]	42 ± 23.5	46	44 ± 26.6	45	21.1%	−0.08 (−0.49, 0.33)
Subtotal[c]		207		234	100.0%	−0.20 (−0.39, −0.01)



[a]*Heterogeneity: $\tau^2 = 0.00$, $\chi^2 = 2.85$, df = 4 (P = .58), $I^2 = 0\%$. Test for overall effect: z = 2.03 (P = .04).*
[b]*Heterogeneity: $\tau^2 = 0.00$, $\chi^2 = 3.94$, df = 5 (P = .56), $I^2 = 0\%$. Test for overall effect: z = 1.81 (P = .07).*
[c]*Heterogeneity: $\tau^2 = 0.00$, $\chi^2 = 1.08$, df = 3 (P = .78), $I^2 = 0\%$. Test for overall effect: z = 2.06 (P = .04).*
Abbreviations: IV, inverse variance; SMD, standardized mean difference.

FIGURE 4. Forest plot of the comparison between promotion of walking and minimal intervention for the outcome of function.

als were from high-income countries, so our results may not be generalizable to medium- or low-income countries.

Clinical and Research Implications

Even if current strategies to promote walking are not effective for increasing physical activity or walking volume, they may still be worthwhile for improving clinical outcomes (eg, pain, function). The fact that no differences in clinical outcomes were observed between walking and the more resource-intensive and costly supervised exercise suggests that walking promotion may be a worthwhile treatment alternative to supervised exercise, particularly in resource-poor or remote settings.

Physical activity campaigns typically have a one-size-fits-all approach, but our results suggest that effective strategies to promote physical activity in the general population do not necessarily increase physical activity in people with musculoskeletal disorders. This population is at increased risk of physical inactivity and other noncommunicable diseases, so further research is needed to find effective interventions to increase physical activity. Promoting physical activity among people with musculoskeletal disorders may require different strategies involving a combination of activities (eg, walking plus another form of cardiovascular exercise), or there may be a need to address the psychosocial aspect (fear of activity) together with physical activity promotion.

Future high-quality randomized controlled trials with appropriate sample sizes are needed. It is important that future trials include objective measures of physical activity that are not subject to the reporting or recall bias associated with self-report methods. Future trials should also be conducted in low- and middle-income countries to fill this gap in the literature.

CONCLUSION

STRATEGIES TO PROMOTE WALKING did not increase physical activity or walking volume in people with musculoskeletal disorders compared to minimal interventions. Walking promotion provided small improvements in pain and function compared to minimal interventions and usual care. There was moderate- to low-quality evidence of no difference between walking promotion and supervised exercise for clinical outcomes. ◉

KEY POINTS

FINDINGS: Strategies to promote walking did not consistently increase physical activity in people with musculoskeletal disorders. Strategies to promote walking were more effective for improving clinical outcomes (ie, pain and disability) compared to minimal interventions and usual care, and were as effective as supervised exercises.

IMPLICATIONS: Promoting physical activity in people with musculoskeletal disorders may require strategies other than walking to account for the complexity of the pain experience in this population. Walking may be a worthwhile treatment alternative to supervised exercise to manage pain and disability, particularly in more resource-poor or remote settings.

CAUTION: We found different measures for assessing the primary outcome; thus, meta-analysis was only possible for the outcome of number of steps per day. We were unable to identify whether the level of physical activity or clinical outcomes differed among the different types of conditions included.

STUDY DETAILS

AUTHOR CONTRIBUTIONS: All authors were involved in the conception and design of the review. Drs Saragiotto, Yamato, Wang, and Lin and Mr Fioratti developed the search strategy and performed study selection. Drs Saragiotto, Yamato, and Wang and Mr Fioratti extracted data from included studies. Drs Saragiotto, Yamato, Lin, Hancock, Tiedemann, and Chau and Mr Fioratti were involved in the data analysis. All authors were involved in the interpretation and discussion of results. Dr Saragiotto drafted the manuscript, and all authors revised the draft. All authors approved the final version of the article. All authors had access to all data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

DATA SHARING: Data are available on request.

PATIENT AND PUBLIC INVOLVEMENT: Patients were not involved in this review.

REFERENCES

1. Ang DC, Kaleth AS, Bigatti S, et al. Research to encourage exercise for fibromyalgia (REEF): use of motivational interviewing, outcomes from a randomized-controlled trial. *Clin J Pain*. 2013;29:296-304. https://doi.org/10.1097/AJP.0b013e318254ac76
2. Baxter SV, Hale LA, Stebbings S, Gray AR, Smith CM, Treharne GJ. Walking is a feasible physical activity for people with rheumatoid arthritis: a feasibility randomized controlled trial. *Musculoskeletal Care*. 2016;14:47-56. https://doi.org/10.1002/msc.1112
3. Boutevillain L, Dupeyron A, Rouch C, Richard E, Coudeyre E. Facilitators and barriers to physical activity in people with chronic low back pain: a qualitative study. *PLoS One*. 2017;12:e0179826. https://doi.org/10.1371/journal.pone.0179826
4. Brodin N, Eurenius E, Jensen I, Nisell R, Opava CH, Para Study Group. Coaching patients with early rheumatoid arthritis to healthy physical activity: a multicenter, randomized, controlled study. *Arthritis Rheum*. 2008;59:325-331. https://doi.org/10.1002/art.23327
5. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
6. de Morton NA. The PEDro scale is a valid measure of the methodological quality of clinical trials: a demographic study. *Aust J Physiother*. 2009;55:129-133. https://doi.org/10.1016/s0004-9514(09)70043-1
7. Farr JN, Going SB, Lohman TG, et al. Physical activity levels in patients with early knee osteoarthritis measured by accelerometry. *Arthritis Rheum*. 2008;59:1229-1236. https://doi.org/10.1002/art.24007
8. Fontaine KR, Conn L, Clauw DJ. Effects of lifestyle physical activity in adults with fibromyalgia: results at follow-up. *J Clin Rheumatol*. 2011;17:64-68. https://doi.org/10.1097/RHU.0b013e31820e7ea7
9. Halbert J, Crotty M, Weller D, Ahern M, Silagy C. Primary care–based physical activity programs: effectiveness in sedentary older patients with osteoarthritis symptoms. *Arthritis Rheum*. 2001;45:228-234. https://doi.org/10.1002/1529-0131(200106)45:3<228::AID-ART253>3.0.CO;2-2
10. Hartvigsen J, Morsø L, Bendix T, Manniche C.

Supervised and non-supervised Nordic walking in the treatment of chronic low back pain: a single blind randomized clinical trial. *BMC Musculoskelet Disord*. 2010;11:30. https://doi.org/10.1186/1471-2474-11-30
11. Hiyama Y, Yamada M, Kitagawa A, Tei N, Okada S. A four-week walking exercise programme in patients with knee osteoarthritis improves the ability of dual-task performance: a randomized controlled trial. *Clin Rehabil*. 2012;26:403-412. https://doi.org/10.1177/0269215511421028
12. Hurley DA, Tully MA, Lonsdale C, et al. Supervised walking in comparison with fitness training for chronic back pain in physiotherapy: results of the SWIFT single-blinded randomized controlled trial (ISRCTN17592092). *Pain*. 2015;156:131-147. https://doi.org/10.1016/j.pain.0000000000000013
13. Kassavou A, Turner A, French DP. Do interventions to promote walking in groups increase physical activity? A meta-analysis. *Int J Behav Nutr Phys Act*. 2013;10:18. https://doi.org/10.1186/1479-5868-10-18
14. Katz P, Margaretten M, Gregorich S, Trupin L. Physical activity to reduce fatigue in rheumatoid arthritis: a randomized controlled trial. *Arthritis Care Res (Hoboken)*. 2018;70:1-10. https://doi.org/10.1002/acr.23230
15. Krein SL, Kadri R, Hughes M, et al. Pedometer-based internet-mediated intervention for adults with chronic low back pain: randomized controlled trial. *J Med Internet Res*. 2013;15:e181. https://doi.org/10.2196/jmir.2605
16. Larsson C, Ekvall Hansson E, Sundquist K, Jakobsson U. Impact of pain characteristics and fear-avoidance beliefs on physical activity levels among older adults with chronic pain: a population-based, longitudinal study. *BMC Geriatr*. 2016;16:50. https://doi.org/10.1186/s12877-016-0224-3
17. Lee IM, Buchner DM. The importance of walking to public health. *Med Sci Sports Exerc*. 2008;40:S512-S518. https://doi.org/10.1249/MSS.0b013e31817c65d0
18. Lee IM, Shiroma EJ, Lobelo F, et al. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *Lancet*. 2012;380:219-229. https://doi.org/10.1016/S0140-6736(12)61031-9
19. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ*. 2009;339:b2700. https://doi.org/10.1136/bmj.b2700
20. Macedo LG, Elkins MR, Maher CG, Moseley AM, Herbert RD, Sherrington C. There was evidence of convergent and construct validity of Physiotherapy Evidence Database quality scale for physiotherapy trials. *J Clin Epidemiol*. 2010;63:920-925. https://doi.org/10.1016/j.jclinepi.2009.10.005
21. Maher CG, Sherrington C, Herbert RD, Moseley AM, Elkins M. Reliability of the PEDro scale for rating quality of randomized controlled trials. *Phys Ther*. 2003;83:713-721. https://doi.org/10.1093/ptj/83.8.713
22. Mansi S, Milosavljevic S, Baxter GD, Tumilty S, Hendrick P. A systematic review of studies using pedometers as an intervention for musculoskeletal diseases. *BMC Musculoskelet Disord*. 2014;15:231. https://doi.org/10.1186/1471-2474-15-231
23. McDonough SM, Tully MA, Boyd A, et al. Pedometer-driven walking for chronic low back pain: a feasibility randomized controlled trial. *Clin J Pain*. 2013;29:972-981. https://doi.org/10.1097/AJP.0b013e31827f9d81
24. Mueller PS, Montori VM, Bassler D, Koenig BA, Guyatt GH. Ethical issues in stopping randomized trials early because of apparent benefit. *Ann Intern Med*. 2007;146:878-881. https://doi.org/10.7326/0003-4819-146-12-200706190-00009
25. O'Connor SR, Tully MA, Ryan B, et al. Walking exercise for chronic musculoskeletal pain: systematic review and meta-analysis. *Arch Phys Med Rehabil*. 2015;96:724-734.e3. https://doi.org/10.1016/j.apmr.2014.12.003
26. Ogilvie D, Foster CE, Rothnie H, et al. Interventions to promote walking: systematic review. *BMJ*. 2007;334:1204. https://doi.org/10.1136/bmj.39198.722720.BE
27. Saragiotto BT, Maher CG, Yamato TP, et al. Motor control exercise for chronic non-specific low-back pain. *Cochrane Database Syst Rev*. 2016:CD012004. https://doi.org/10.1002/14651858.CD012004
28. Talbot LA, Gaines JM, Huynh TN, Metter EJ. A home-based pedometer-driven walking program to increase physical activity in older adults with osteoarthritis of the knee: a preliminary study. *J Am Geriatr Soc*. 2003;51:387-392. https://doi.org/10.1046/j.1532-5415.2003.51113.x
29. Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Med Res Methodol*. 2014;14:135. https://doi.org/10.1186/1471-2288-14-135
30. World Health Organization. Global Health Risks: Mortality and Burden of Disease Attributable to Selected Major Risks. Geneva, Switzerland: World Health Organization; 2009.
31. World Health Organization. Global Recommendations on Physical Activity for Health. Geneva, Switzerland: World Health Organization; 2010.
32. Yamato TP, Maher C, Koes B, Moseley A. The PEDro scale had acceptably high convergent validity, construct validity, and interrater reliability in evaluating methodological quality of pharmaceutical trials. *J Clin Epidemiol*. 2017;86:176-181. https://doi.org/10.1016/j.jclinepi.2017.03.002

APPENDIX A

SEARCH STRATEGY FOR THE MEDLINE DATABASE

MEDLINE (via Ovid)

Participants

1. Musculoskeletal pain.mp.
2. Musculoskeletal diseases.mp.
3. Musculoskeletal disorder$.mp
4. exp Back pain/
5. (Neck OR cervical) AND pain.mp.
6. (Pelvic OR Spine OR Spinal) AND pain.mp.
7. exp Sciatica/
8. Cervicogenic.mp.
9. (Spondylitis OR Spondylosis).mp.
10. exp Pain/ and exp Hip/
11. *Arthralgia/ or exp Osteoarthritis, Hip/
12. exp Knee Joint/ and exp Pain/
13. exp Osteoarthritis, Knee/ or exp Osteoarthritis/
14. exp Ankle Joint/ and exp Pain/
15. exp Shoulder Pain/
16. (Frozen adj Shoulder).mp.
17. exp Headache/ or *Tension-Type Headache/ or *Headache Disorders/
18. exp Fibromyalgia/
19. arthritis.mp.
20. Joint disease.mp.
21. (Joint adj Pain).mp.
22. Myalgia.mp.
23. 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 or 16 or 17 or 18 or 19 or 20 or 21 or 22

Intervention

24. exp Walking/
25. (walk$ or stroll$ or treadmill$ or locomot$ or stride$ or pace$ or pacing).mp.
26. Accelerometer$.mp.
27. Pedometer$.mp.
28. Activity monitor$.mp.
29. Step count$.mp
30. (Aerobic adj exercise).mp.
31. 24 or 25 or 26 or 27 or 28 or 29 or 30

Study Design

32. Randomised controlled trial.mp.
33. Randomized controlled trial.mp.
34. Controlled clinical trial.mp.
35. Double blind method.mp.
36. Single-blind method.mp.
37. Clinical trial.mp.
38. Random Allocation.mp.
39. Cross-over studies.mp.
40. (Clinic$ adj25 trial$).tw.
41. 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40
42. 23 and 31 and 41

APPENDIX B

PHYSICAL ACTIVITY/WALKING-RELATED OUTCOMES FOR THE COMPARISON BETWEEN WALKING PROMOTION AND MINIMAL INTERVENTION

Study	Outcome	Type of Measure	Follow-up	Between-Group Difference[a]
Ang et al[1]	Minimum increment change of MVPA (percent of participants with >30-min increment of MVPA)	Self-reported (CHAMPS survey)	6 mo	RR = 1.04 (0.79, 1.36)
	Change in time spent on MVPA (hours per week)	Self-reported (CHAMPS survey)	12 wk	MD, 1.5 (0.39, 2.60)[b]
			3 mo after intervention	MD, 0.3 (−0.81, 1.41)
			6 mo after intervention	MD, 0.8 (−0.31, 1.91)
	Change in time spent on MVPA (minutes per week)	Accelerometer	12 wk	MD, −20.7 (−6.57, 48.05)
			3 mo after intervention	MD, 20.4 (−54.91, 14.11)
			6 mo after intervention	MD, 12 (−36.53, 12.53)
Brodin et al[4]	Percent of participants' physical activity classified as sufficiently healthy (≥4 times per week of moderate and/or high intensity)	Self-reported questionnaire designed by study researchers	12 mo	RR = 1.22 (0.92, 1.62)
Halbert et al[9]	Minutes of walking per session	Self-reported	3 mo	MD, 5.00 (−15.83, 25.83)
			6 mo	MD, 3.00 (−12.80, 18.80)
			12 mo	MD, −2.00 (−25.92, 21.92)
	Minutes of vigorous exercise per session	Self-reported	3 mo	MD, 8.00 (−5.86, 21.86)
			6 mo	MD, 6.00 (−6.55, 18.55)
			12 mo	MD, 9.00 (−1.96, 19.96)
	Frequency of walking per week	Self-reported	3 mo	MD, 0.40 (−0.85, 1.65)
			6 mo	MD, 0.40 (−0.58, 1.38)
			12 mo	MD, −0.25 (−1.39, 0.89)
	Frequency of vigorous exercise per week	Self-reported	3 mo	MD, 0.70 (−0.55, 1.95)
			6 mo	MD, 0.50 (−0.33, 1.33)
			12 mo	MD, 0.40 (−0.74, 1.54)

Abbreviations: CHAMPS, Community Healthy Activities Model Program for Seniors; MD, mean difference; MVPA, moderate-to-vigorous physical activity; RR, risk ratio.

[a]*Values in parentheses are 95% confidence interval. Between-group differences were calculated based on the raw data available in the study, except for that reported by Halbert et al,[9] which was estimated from the bar charts of the study. Positive values indicate a difference favoring the intervention group, and negative values indicate a difference favoring the control group.*

[b]*Statistically significant difference.*

MARK P. WILHELM, PT, DPT, PhD[1] • MEGAN DONALDSON, PT, PhD[1] • DAVID GRISWOLD, DPT, PhD[2] • KENNETH E. LEARMAN, PT, PhD[2]
ALESSANDRA N. GARCIA, PT, PhD[3] • SHANE M. LEARMAN, BS[4] • JOSHUA A. CLELAND, PT, PhD[1]

The Effects of Exercise Dosage on Neck-Related Pain and Disability: A Systematic Review With Meta-analysis

Neck pain is a common reason for patients to seek medical care.[27] At least 1 in 3 people with neck pain experience persistent symptoms.[21] Clinical practice guidelines for managing neck pain recommend exercise therapy for reducing pain and disability.[8,30] However, evidence-based recommendations of specific exercise dosage for managing neck pain are lacking.[30]

Exercise therapy may provide both physical and mental benefits because it impacts numerous body systems, including cardiovascular, immune, neurologic, and musculoskeletal systems.[23] Clinical trials often use a multimodal approach, blending exercise therapy with other therapeutic interventions, including manual therapy and dry needling. These trials have shown small to moderate effects on neck pain and disability.[6] Studies investigating effects of exercise therapy as an individual treatment for neck pain and disability report consistent positive outcomes but variable effect sizes.[7,14,15]

OBJECTIVE: To (1) evaluate whether exercise therapy is effective for managing neck pain, and (2) investigate the relationship between exercise therapy dosage and treatment effect.

DESIGN: Intervention systematic review with meta-analysis and meta-regression.

LITERATURE SEARCH: An electronic search of 6 databases was completed for trials assessing the effects of exercise therapy on neck pain.

STUDY SELECTION CRITERIA: We included randomized controlled trials that compared exercise therapy to a no-exercise therapy control for treating neck pain. Two reviewers screened and selected studies, extracted outcomes, assessed article risk of bias, and rated the quality of evidence using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach.

DATA SYNTHESIS: Data were pooled using random-effects meta-analysis. We used meta-regression to analyze the effect of exercise dosage on neck pain and disability.

RESULTS: Fourteen trials were included in the review. Seven trials were at high risk of bias, 4 were at unclear risk of bias, and 3 were at low risk of bias. Exercise therapy was superior to control for reducing pain (visual analog scale mean difference, –15.32 mm) and improving disability (Neck Disability Index mean difference, –3.64 points). Exercise dosage parameters did not predict pain or disability outcomes.

CONCLUSION: Exercise was beneficial for reducing pain and disability, regardless of exercise therapy dosage. Therefore, optimal exercise dosage recommendations remain unknown. We encourage clinicians to use exercise when managing mechanical neck pain. *J Orthop Sports Phys Ther 2020;50(11):607-621. doi:10.2519/jospt.2020.9155*

KEY WORDS: *cervical spine, disability, dosage, pain, therapeutic exercise*

Although 2 systematic reviews have demonstrated the efficacy of exercise therapy for patients with neck pain,[7,15] neither attempted to assess the impact of exercise therapy dosage on pain and disability outcomes. Inconsistent effect sizes in the Cochrane review update[15] may reflect the different exercise therapy dosage, highlighting the need to study the exercise therapy dosage necessary to reach efficacy.

The purpose of our systematic review was to assess published randomized controlled trials to (1) evaluate whether exercise therapy was effective for managing neck pain, and (2) investigate the relationship between exercise therapy dosage and treatment effect.

METHODS

Protocol and Registration

The review was registered prospectively on April 18, 2017 with PROSPERO (CRD42017063956). The Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines were used to guide the conduct and reporting of this systematic review and meta-analysis.[29]

[1]Doctor of Physical Therapy Program, Department of Public Health and Community Medicine, Tufts University School of Medicine, Boston, MA. [2]Physical Therapy Department, Youngstown State University, Youngstown, OH. [3]Division of Physical Therapy, Department of Orthopaedic Surgery, Duke University School of Medicine, Durham, NC. [4]Department of Biological Sciences, Youngstown State University, Youngstown, OH. This project was not supported by grant funding, and was registered with PROSPERO (CRD42017063956). 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 Mark Wilhelm, 136 Harrison Avenue, Boston, MA 02111. E-mail: MarkPWilhelm@gmail.com

Search Strategy

An electronic search of PubMed, the Physiotherapy Evidence Database (PEDro), SPORTDiscus, CINAHL, Scopus, and the Cochrane Central Register of Controlled Trials was completed after registration with PROSPERO. Database searches included articles published from the time of inception until the searches were completed on May 10, 2017. An electronic search update was completed on November 26, 2018. The search strategy was developed with the assistance of a health sciences librarian and used Medical Subject Headings (MeSH) terms, key words, and text words associated with neck-related pain and exercise treatment. Search strategies were altered as necessary to complete searches in all databases. Database-specific filters restricted article searches to randomized controlled trials or controlled trials. The full search strategy for PubMed is available in **APPENDIX A** (available at www.jospt.org).

Inclusion Criteria

We included randomized controlled trials that (1) included individuals with acute, subacute, or chronic neck-related disorders; (2) assessed an exercise-based exercise therapy approach intended to treat the neck-related painful condition; (3) included a control group with a nonactivity-based treatment strategy; (4) reported outcomes using a validated pain and/or disability outcome measure; and (5) included exercise dosage details related to frequency, intensity, sets, repetitions, or other details to allow comparison across trials. We excluded trials (1) with a comparison group that received treatment including exercise, manual therapy, or active modalities, (2) on participants undergoing surgical intervention, (3) on participants receiving postsurgical rehabilitation, (4) with exercise therapy that only consisted of unsupervised home exercise programs, or (5) that were not published in the English language.

Study Selection and Data Extraction

Titles, abstracts, and full-text articles were independently screened in 3 sequential steps by 2 authors at each step. Disagreements between authors were resolved via consensus. A third author was available as a tiebreaker. Agreement at each screening step was determined using an unweighted Cohen's kappa.[25] Data were extracted by 2 authors and entered into a data-extraction form. The accuracy of data extraction was confirmed by a third author. We extracted (1) sample size, (2) participant demographics, (3) pain duration (in weeks), (4) exercise type (ie, endurance, strength), (5) exercise dosage details (ie, session duration in minutes, frequency per week), (6) duration of care (in weeks), (7) number of sessions, (8) follow-up time frames, (9) pain- and disability-related outcome measures utilized, and (10) measures of central tendency and dispersion for these pain and disability measures. If mean and standard deviation were not available but data were presented in figures, we used the WebPlotDigitizer to estimate these data.[33]

Risk of Bias

Two authors independently assessed the included trials using the Cochrane risk-of-bias tool. Disagreements were resolved by consensus. A third author was available as a tiebreaker. The Cochrane risk of bias tool assesses bias across 6 domains using 7 items: selection bias (random sequence generation and allocation concealment), performance bias (blinding of participants and personnel), detection bias (blinding of outcome assessment), attrition bias (incomplete outcome data), reporting bias (selective reporting), and other biases (other sources of bias). The 7 items are rated as "high," "low," or "unclear" risk.[17] We classified studies as having a high overall risk of bias if they were at high or unclear risk of bias for allocation concealment, random allocation, or incomplete outcome assessment.[13,40] Due to the difficulty of blinding participants and personnel from allocation to an exercise group assignment, high risk of bias in this domain did not automatically result in an overall assessment of high risk of bias. Interrater reliability was determined using an unweighted Cohen's kappa.[25]

Exercise Dosage

Because exercise dosage reporting varies across published rehabilitation literature, we operationally defined overall exercise dosage as the total number of minutes of supervised exercise performed by participants over the duration of care.

Data Synthesis and Analysis

We conducted separate meta-analyses for exercise therapy programs of intermediate duration (4-9 weeks) and long duration (10 weeks or greater). Data were pooled from the follow-up measure that was closest to the end of the exercise program. Pooled mean differences and 95% confidence intervals (CIs) were calculated for pain and disability outcomes. Pain and disability measures were converted to common 0-to-100 and 0-to-50 scales, respectively.

Heterogeneity across studies was investigated by the I^2 statistic (25%, low heterogeneity; 50%, medium; and 75% or greater, high[19]), overlap of CIs, and *P* value results to determine whether a fixed-effects or random-effects model was appropriate. Although heterogeneity of the separate meta-analyses ranged from low (0.0%) to high (86%), inverse-variance random-effects models were used for overall effect magnitude interpretation due to a large amount of variability in design factors among included trials. We interpreted the mean difference values as small effect (less than 10% of the scale), medium effect (mean difference of 10% to 20% of the scale), or large effect (greater than 20% of the scale).[34,36] Measures of treatment were considered statistically significant when the 95% CI excluded zero (for mean difference), and changes were considered clinically meaningful with a mean difference of ±10/100 for the pain visual analog scale (VAS) and a mean difference of 5/50 points for the Neck Disability Index (NDI) and Neck Pain and Disability scale (NPAD).[35,38]

We performed random-effects meta-regression for each outcome (pain and disability) to explore the difference in effects due to variability in exercise therapy dosage. Two meta-regression models were created, with each model using 1 of the following continuous moderator variables: average number of exercise minutes per week or total weeks of exercise. Pooled meta-analyses were completed using Review Manager 5.3 (The Nordic Cochrane Centre, Copenhagen, Denmark). Meta-regression models were performed using Comprehensive Meta-Analysis Version 3.0 (Biostat, Inc, Englewood, NJ). As an exploratory analysis, each study was removed from the overall pain and disability meta-analyses to assess whether removal of individual studies altered the statistical significance of the pooled mean difference values. During each exploratory analysis, single studies were removed that had the largest and smallest sample sizes, as well as the largest and smallest treatment dosages.

Judging Certainty of Evidence

We used the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach to judge the certainty of evidence.[17] We considered evidence as "high certainty" (further research is very unlikely to change the confidence in the estimate of effect), "moderate certainty" (further research is likely to have an important impact on the confidence in the estimate of effect), "low certainty" (further research is very likely to have an important impact on the confidence in the estimate of effect and is likely to change the estimate), or "very low certainty" (little confidence in the effect estimate).[13,18] We downgraded the certainty of evidence 1 level for each of 5 domains: (1) risk of bias (25% or greater of the trials were at high risk); (2) inconsistency (heterogeneity): point estimates varied widely across studies, CIs had minimal or no overlap, substantial heterogeneity ($I^2 \geq 50\%$); (3) imprecision of results (fewer than 300 events for dichotomous outcomes and fewer than 400 participants for continuous outcomes, similar to previous Cochrane reviews examining the effectiveness of exercise to treat spinal conditions[40]; the 95% CIs around the effect estimate included both the possibility of no effect [includes zero] and also an important benefit [mean difference of ±10/100 for the pain VAS and 5/50 points for the NDI and NPAD])[16,40]; (4) indirectness (inability to generalize); and (5) publication bias (selective publication of studies, assessed using funnel plot analysis and the Egger regression test when at least 10 studies were included in the meta-analysis).[13]

RESULTS

Trial Selection

The electronic database search yielded 4151 titles (**FIGURE 1**). After duplicates were removed, 2380 unique article titles were screened for eligibility by 2 reviewers, with a reliability of $\kappa = 0.47$ (95% CI: 0.43, 0.51) and 83% agreement. From these titles, 559 ab-

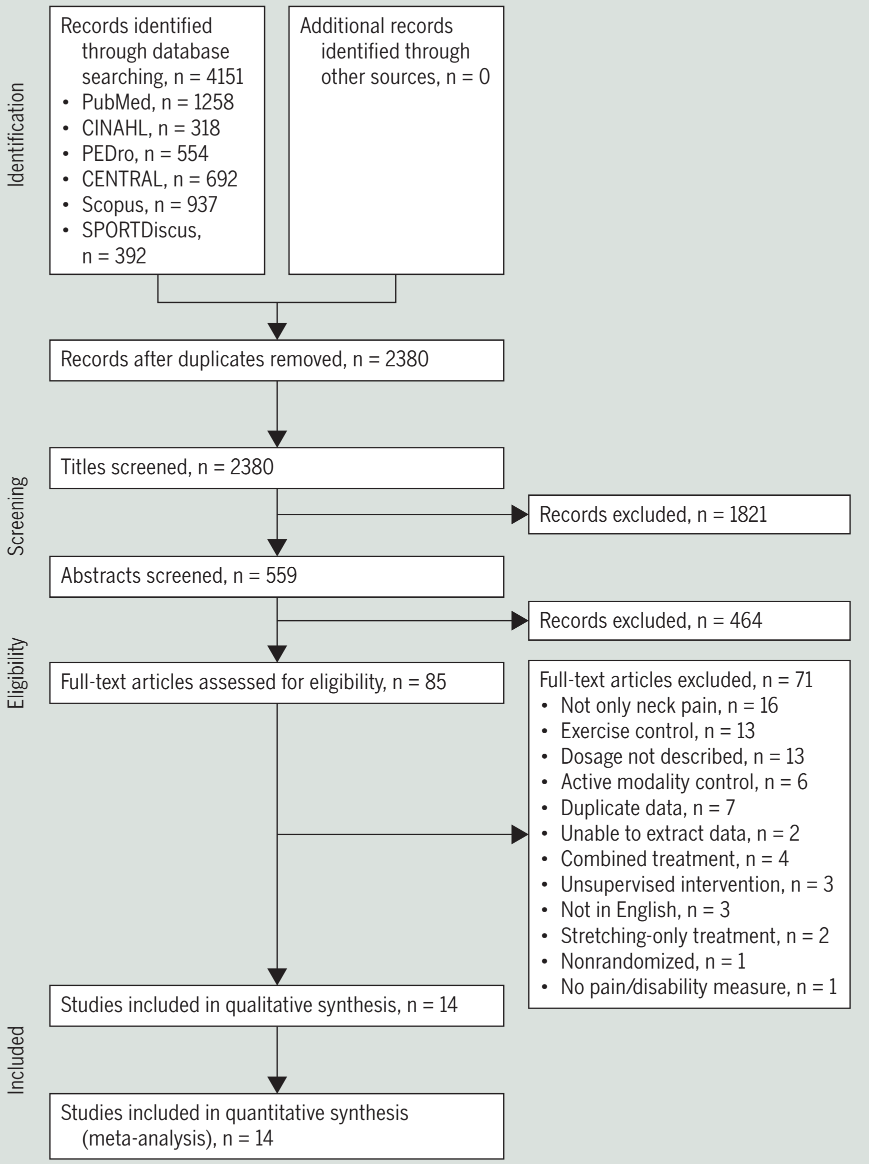


FIGURE 1. PRISMA flow diagram. Abbreviation: CENTRAL, Cochrane Central Register of Controlled Trials.

stracts were screened (κ = 0.46; 95% CI: 0.37, 0.56; 82% agreement), resulting in 85 full-text articles assessed for eligibility. Fourteen articles were included in the quantitative synthesis (κ = 0.48; 95% CI: 0.27, 0.69; 80% agreement) (**FIGURE 1**).

Trial Characteristics

Sample sizes ranged from 15 to 449 participants, with a combined sample size of 1708 participants (**TABLE 1**). Seven trials had 1 treatment group with a control group, 5 had 2 different treatment groups with a control group, and only 2 used 3 different exercise groups in addition to the control. Thirteen trials used a pain intensity measure. One trial used a pain-related measure of "self-rated health"; we did not pool these data with the other pain measures.[32] Nine trials used a disability-related measure. We could not extract sufficient data from 1 trial for inclusion in the meta-analysis.[2] We attempted to contact the authors but were unsuccessful. One trial, with 2 exercise groups, used the NPAD measure and was therefore not able to be included in the meta-analysis for the NDI.[39] As a result, data from these 2 exercise groups were combined in a separate subanalysis for the NPAD.

Risk of Bias

Seven trials were at high risk of bias, 4 were at unclear risk of bias, and 3 were at low risk of bias (**TABLE 2**). Trials were considered to have a high risk of bias due to high or unclear risk of bias in random allocation (3 trials), allocation concealment (5 trials), and incomplete outcome assessment (4 trials). Interrater reliability for risk of bias was κ = 0.84 (95% CI: 0.71, 0.96; absolute agreement, 94%).

Session Duration

Exercise duration ranged from 2 to 90 minutes (**TABLE 3**). Multiple trials included an exercise program that progressed in session duration over the course of care or presented ranges of exercise session duration.[5,26,31] One trial[5] used deep neck flexor and extensor exercises, starting at 15 minutes and progressing to 30 minutes. One trial[26] included a tai chi group, with sessions ranging from 75 to 90 minutes, and a neck exercise group, with sessions ranging from 60 to 75 minutes. One trial[31] included exercise sessions between 30 and 40 minutes.

Exercise Dosage and Adherence

Exercise frequency ranged from 1 to 7 days per week. The total number of supervised exercise sessions over the entire duration of care for the 14 included studies ranged from 4 to 180. The total duration of care for the included trials ranged from 4 to 20 weeks, and total exercise therapy dosage ranged from 70 to 1800 minutes. Total exercise dosage in minutes could not be calculated for 1 trial.[24] Two trials met adherence rates greater than 75% and had a low risk of bias.[9,11] Exercise dosage details for individual trials can be found in **TABLE 3**.

Effect of Exercise: Pain Intensity

The overall effects of exercise on pain outcomes were studied using 21 exercise groups from 13 different trials (n = 1080 participants). For exercise compared to control, there was moderate-certainty evidence (downgraded 1 level due to risk of bias) that exercise produced medium, clinically meaningful improvements in pain intensity (mean difference, –15.32 mm; 95% CI: –19.20, –11.44) at intermediate- to long-term follow-up (**TABLE 4, FIGURE 2**; **APPENDIX B**, available at www.jospt.org). Because the

TABLE 1 — Characteristics of Included Studies

Study, Setting, Country, Sample Size	Participants, Stage	Controls, Stage	Intervention	Outcome: Pain	Outcome: Disability	ITT	Effect Magnitude[a]
Ahlgren et al[1] University medical center Sweden n = 102	Strength: n = 29; age, 38 y; pain duration, 327.6 wk Endurance: n = 28; age, 38.5 y; pain duration, 338 wk Coordination: n = 25; age, 37.7 y; pain duration, 343.2 wk Chronic	n = 20; age, 38.9 y; pain duration, 400.4 wk Chronic	10-wk training program. Strength: pneumatic machines for upper body. Endurance: alternating intervals of upper-body cycle ergometry with light band-resisted exercises. Coordination training involved body awareness training with integration of movement with breathing	VAS	NA	No	Pain Strength, –16.0 (–36.29, 4.29); no effect Endurance, –7.0 (–25.86, 11.86); no effect Coordination, –8.0 (–26.99, 10.99); no effect
Andersen et al[3] Denmark n = 48	SST: n = 18; age, 44 y; BMI, 27 kg/m² GFT: n = 16; age, 45 y; BMI, 27 kg/m² Pain duration, >30 d of pain during the past y Subacute or chronic	n = 14; age, 42 y; BMI, 25 kg/m²; pain duration, >30 d of pain during the past y Subacute or chronic	10 wk of training. SST group worked on 5 dumbbell exercises for local strengthening of the upper body, whereas the GFT group worked on GFT for the lower body. The reference group received education on general health	VAS	NA	No	Pain SST, –22.0 (–38.82, –5.18); large effect GFT, 3.0 (–14.33, 20.33); no effect

Table continues on page 611.

TABLE 1	Characteristics of Included Studies (continued)						
Study, Setting, Country, Sample Size	**Participants, Stage**	**Controls, Stage**	**Intervention**	**Outcome: Pain**	**Outcome: Disability**	**ITT**	**Effect Magnitude[a]**
Andersen et al[4] Denmark n = 198	2 min of exercise: n = 66 (8 male, 58 female); age, 44 y; BMI, 25 kg/m^2 12 min of exercise: n = 66 (8 male, 58 female); age, 42 y; BMI, 24 kg/m^2 Pain duration, >30 d of pain during the past y Subacute or chronic	n = 66 (8 male, 58 female); age, 43 y; BMI, 23 kg/m^2; pain duration, >30 d of pain during the past y Subacute or chronic	10 wk of training. The 12-min group performed 5-6 sets of 8-12 repetitions of shoulder abduction. The 2-min group performed 1 set of shoulder abduction to failure. The control group received weekly e-mails of general health advice	NRS for neck, NRS for headache, tenderness score	NA	No	Pain 2-min group: −14.0 (−22.24, −5.76); medium effect 12-min group: −19.0 (−27.24, −10.76); medium effect
Andersen et al[2] 12 outpatient clinics Denmark n = 449	1 workout per week: n = 116 (44 male, 72 female); age, 47 y 3 workouts per week: n = 126 (39 male, 87 female); age, 46 y 9 workouts per week: n = 106 (45 male, 61 female); age, 45 y Pain duration and stage not stated	n = 101 (42 male, 59 female); age, 46 y; pain duration not stated	20-wk strengthening programs emphasizing dumbbell exercises. Groups were divided by time per session, not total time	NPRS	DASH	Yes	Pain 1 workout, −11.4 (−18.41, −4.39); medium effect 3 workouts, −18.8 (−26.04, −11.56); medium effect 9 workouts, −13.5 (−21.28, −5.72); medium effect
Andias et al[5] Secondary school Portugal n = 43	n = 21; age, 17.4 y; pain duration, at least 12 wk Chronic	n = 22; age, 15.9 y; pain duration, at least 12 wk Chronic	Neuroeducation and exercise for 3 sessions. Exercise was aimed at increasing the endurance and strength of deep neck flexors and scapular stabilizers	VAS	NA	NA	Pain −8.0 (−19.09, 3.09); no effect
Brage et al[9] Denmark n = 15	n = 8; age, 40.8 y; BMI, 23 kg/m^2; pain duration, 399.9 wk Chronic	n = 7; age, 42.1 y; BMI, 23.4 kg/m^2; pain duration, 1095.0 wk Chronic	8 wk of training were performed. Included combination of neck flexor/extensor training, balance, and upper extremity proprioception and aerobic training	NPRS	NDI	Yes, but not actually done	Pain −34.3 (−57.88, −10.72); large effect Disability −5.92 (−13.95, 2.11); no effect
de Araujo Cazotti et al[11] NA n = 64	n = 32 (26 male); age, 48.6 y; BMI, 25.7 kg/m^2; pain duration, 299.4 wk Chronic	n = 32; age, 49 y; BMI, 24.2 kg/m^2; pain duration, 373.1 wk Chronic	12 wk of Pilates, prioritizing breathing exercises, spine mobility, and shoulder girdle strengthening	VAS	NDI	Yes	Pain −41.7 (−50.95, −32.45); large effect Disability −7.03 (−9.84, −4.22); medium effect
Falla et al[12] Pain management centers Germany n = 42	n = 22; age, 39.1 y; pain duration, 520 wk Chronic	n = 20; age, 38.6 y; pain duration, 436.8 wk Chronic	8 wk of training were performed. Intervention included craniocervical endurance training using biofeedback for 6 wk, followed by 2 wk of gradual strengthening with resistance from the head for flexors and extensors	VAS	NDI	No	Pain −13.0 (−27.22, 1.22); no effect Disability −2.5 (−6.76, 1.76); no effect
Im et al[22] University setting Korea n = 15	n = 8 (5 male, 3 female); age, 35.5 y; pain duration, >12 wk Chronic	n = 7 (6 male, 1 female); age, 35.7 y; pain duration, >12 wk Chronic	4 wk of scapular stabilization exercises were performed	VAS	NDI	No	Pain −13.0 (−21.47, −4.53); medium effect Disability −5.6 (−11.78, 0.58); no effect

Table continues on page 612.

I^2 test was close to 50% and we did not observe wide point estimates across studies, with minimal or no overlap across CIs, we did not downgrade for inconsistency.

For pain intensity at intermediate-term (5 trials, n = 242 participants) follow-up, there was low-certainty evidence (downgraded 1 level due to risk of bias and imprecision due to sample size) that exercise produced medium, clinically meaningful improvements in pain intensity (mean difference, –13.36 mm; 95% CI: –18.57, –8.14) (**TABLE 4, FIGURE 3**).

For pain intensity at long-term follow-up (8 trials, n = 838 participants), there was low-certainty evidence (downgraded 1 level due to risk of bias and inconsistency)

TABLE 1 CHARACTERISTICS OF INCLUDED STUDIES (CONTINUED)

Study, Setting, Country, Sample Size	Participants, Stage	Controls, Stage	Intervention	Outcome: Pain	Outcome: Disability	ITT	Effect Magnitude[a]
Kuijper et al[24] 3 hospital settings the Netherlands n = 205	Collar: n = 69 (38 male, 31 female); age, 47 y; BMI, 26.5 kg/m^2 Physical therapy: n = 70 (34 male, 36 female); age, 46.7 y; BMI, 26.2 kg/m^2 Pain duration, <4 wk Acute	n = 66 (32 male, 34 female); age, 47.7 y; BMI, 26.8 kg/m^2; pain duration, <4 wk Acute	One group received 6 wk of a semi-rigid collar, which was weaned over the final 3 wk. The second group received 12 sessions of supervised exercises for mobility and stability of the neck, including strengthening of the deep and superficial neck muscles	VAS	NDI	Yes	Pain –14.9 (–26.00, –3.80); medium effect Disability –1.05 (–4.34, 2.24); no effect
Lauche et al[26] University hospital Germany n= 114	Tai chi: n = 38 (10 male, 28 female); age, 52 y; BMI, 27.2 kg/m^2 Neck exercises: n = 37 (6 male, 31 female); age,47.0 y; BMI, 25.8 kg/m^2 Pain duration, >12 wk Chronic	n = 39 (7 male, 32 female); age, 49.2 y; BMI, 26.4 kg/m^2; pain duration, >12 wk Chronic	12 wk of training were performed. One group did tai chi training and the second group performed neck strengthening, proprioception, dynamic mobilization, and stretching	VAS	NDI	Yes	Pain Tai chi, –9.4 (–21.79, 2.99); no effect Neck exercises, –16.6 (–27.35, –5.85); medium effect Disability Tai chi, –6.0 (–12.43, 0.43); no effect Neck exercises, –4.80 (–10.63, 1.03); no effect
Revel et al[31] France n = 60	n = 30 (8 male, 22 female); age, 47 y; pain duration, 104 wk Chronic	n = 30 (1 male, 29 female); age, 46.5 y; pain duration, 156 wk Chronic	15 individual exercise sessions performed twice a week during an 8-wk period	VAS	Ordinal self-report improvement scale	No	Pain –17.5 (–28.92, –6.08); medium effect
Ris et al[32] Denmark n = 200	n = 101 (32 male, 69 female); age, 45.1 y; pain duration, 465.0 wk Chronic	n = 99 (19 male, 80 female); age, 45.2 y; pain duration, 469.3 wk Chronic	All participants participated in 4 sessions of pain management education. The exercise group received eight 30-min sessions of instruction on exercise/physical training	NA	NDI	Yes	Disability –0.20 (–1.45, 1.05); no effect
von Trott et al[39] University medical center Germany n = 117	Qigong: n = 38; age, 75.9 y; BMI, 28 kg/m^2; pain duration, 1045.2 wk Exercise: n = 39; age, 76.0 y; BMI, 27 kg/m^2 Pain duration, 889.2 wk Chronic	n = 40; age, 75.7 y; BMI, 27.1 kg/m^2; pain duration, 1034.8 wk Chronic	3 mo of training were performed, 2 sessions per week, for a total of 24 sessions	VAS	NPAD	Yes	Pain Exercise, –10.40 (–26.08, 5.28); no effect Qigong, –7.5 (–24.85, 9.85); no effect Disability Exercise, –2.75 (–9.30, 3.80); no effect Qigong, –2.40 (–9.02, 4.22); no effect

Abbreviations: BMI, body mass index; DASH, Disabilities of the Arm, Shoulder and Hand questionnaire; GFT, general fitness training; ITT, intention to treat; NA, not applicable; NDI, Neck Disability Index; NPAD, Neck Pain and Disability scale; NPRS, numeric pain-rating scale; NRS, numeric rating scale; SST, specific strength training; VAS, visual analog scale.
[a]*Values in parentheses are 95% confidence interval.*

that exercise produced medium, clinically meaningful improvements in pain intensity (mean difference, –15.51 mm; 95% CI: –20.28, –10.74) (**TABLE 4, FIGURE 4**).

Effect of Exercise: Disability

The effect of exercise on disability (NDI) was assessed in 8 groups from 7 trials (n = 577 participants). For exercise compared to control, there was low-certainty evidence (downgraded 1 level due to risk of bias and inconsistency) that exercise produced small (not clinically meaningful) improvements in disability (mean difference, –3.64 points; 95% CI: –6.19, –1.09) (**TABLE 5, FIGURE 5**).

For disability at intermediate-term follow-up (4 trials, n = 199 participants), there was low-certainty evidence (downgraded 1 level due to risk of bias and imprecision due to sample size) that exercise produced small (not clinically meaningful) improvements compared to control (mean difference, –2.50 points; 95% CI: –4.80, –0.20) (**TABLE 5, FIGURE 6**).

For disability at long-term follow-up (5 exercise groups in 3 trials, n = 378 participants), there was very low–certainty evidence (downgraded 1 level due to risk of bias, inconsistency, and imprecision due to sample size and the 95% CI around the difference including zero and an important benefit) of no difference between exercise and control (mean difference, –4.23 points; 95% CI: –8.75, 0.29) (**TABLE 5, FIGURE 7**).

Additionally, within 2 combined groups (n = 101 participants), there was moderate-certainty evidence (downgraded 1 level due to imprecision due to sample size and the 95% CI around the difference including zero and an important benefit) of no difference between exercise and control when disability was assessed using the NPAD (mean difference, –2.58 points; 95% CI: –7.24, 2.08) (**TABLE 5, FIGURE 7**).

Meta-regression

There was no significant effect of exercise therapy dosage on treatment effect for pain or disability outcomes. Minutes of exercise per week (regression coefficient = 0.00; 95% CI: 0.00, 0.00) and total weeks of exercise (regression coefficient = 0.00; 95% CI: –0.05, 0.04) did not predict pain outcomes (**FIGURES 8A** and **8B**, respectively). Minutes of exercise per week (regression coefficient = 0.00; 95% CI: 0.00, 0.01) and total weeks of exercise (regression coefficient = 0.01; 95% CI: –0.03, 0.07) did not predict disability outcomes (**FIGURES 9A** and **9B**, respectively).

DISCUSSION

REGARDLESS OF THE DOSAGE, EXERCISE therapy was superior to control for improving pain and disability in people with neck pain (mean difference, –15.32 mm and –3.64 points, respectively). These findings are consistent with another systematic review with meta-analysis reporting significant moderate effects for exercise therapy on pain and disability for patients with chronic neck pain.[7] Our findings are also consistent with current clinical practice guideline recommendations to use exercise therapy for the management of mechanical neck pain.[8] However, no specific dosage trends were identified for maximizing outcomes in patients with neck pain and disability.

Our findings are similar to those of a recently published systematic review investigating the impact of exercise on knee disorders, in that optimal dosage variables are difficult to ascertain from the reporting of current clinical trials.[42] The recently published systematic review suggested that knee exercise programs consisting of 24 sessions generally had larger effect sizes.[42] Knee exercise programs within the range of 8 to 12 weeks in duration generally resulted in large effect sizes, whereas shorter and longer durations resulted in smaller effects.[42] We found that exercise therapy dosage did not significantly change the effect on neck pain or disability. Exercise was beneficial for reducing pain and disability in individuals with mechanical neck pain. We encourage clinicians to follow clinical practice guidelines and use exer-

TABLE 2 RISK OF BIAS WITHIN INCLUDED STUDIES

Study	Cochrane Risk of Bias Criteria[a]							
	A	B	C	D	E	F	G	Overall
Ahlgren et al[1]	H	H	H	H	L	L	U	High
Andersen et al[3]	L	U	U	U	H	H	L	High
Andersen et al[4]	L	L	H	L	L	L	L	Low
Andersen et al[2]	L	U	H	U	L	L	L	High
Andias et al[5]	L	L	H	H	L	L	L	Unclear
Brage et al[9]	L	L	H	L	L	L	L	Low
de Araujo Cazotti et al[11]	L	L	H	L	L	L	L	Low
Falla et al[12]	L	L	H	H	U	L	H	High
Im et al[22]	U	U	H	H	U	U	H	High
Kuijper et al[24]	L	L	H	U	L	L	L	Unclear
Lauche et al[26]	L	L	H	L	H	L	L	High
Revel et al[31]	U	U	H	H	L	U	L	High
Ris et al[32]	L	L	H	H	L	H	L	Unclear
von Trott et al[39]	L	L	H	U	L	U	U	Unclear

Abbreviations: H, high risk of bias; L, low risk of bias; U, unclear risk of bias.
[a]*Criteria: A, Random allocation; B, Allocation concealment; C, Participant and personnel blinding; D, Blinding of outcome assessors; E, Incomplete outcome data; F, Selective outcome reporting; G, Other bias sources.*

cise as the first-line treatment for managing neck pain.[8] As we were unable to determine a specific dosage, we suggest that clinicians tailor their exercise prescription (type and dosage) based on individual patients' clinical presentation and treatment response. Future studies should intentionally focus on exercise program dosage and intensity, similar to the study by Brage et al,[9] which yielded one of the largest mean difference values (–34.30).

Trials used different types or approaches of exercise therapy, including strengthening, motor control, endurance, aerobic, and mobility, alone or in combination. There were too few studies of each exercise type to allow a meaningful comparison of impact on outcomes. Therefore, we are unable to recommend a specific type of exercise therapy that is most beneficial

TABLE 3 Exercise Dosage Details

Study/Group	Session Duration, min	Frequency (n per week)	Average Time, min/wk	Care Duration, wk	Sessions, n	Follow-up	Adherence
Ahlgren et al[1] Coordination, strength, endurance	60	3	180	10	30		Coordination, 75.2%; strength, 77.1%; endurance, 79.6%
Andersen et al[3] GFT, SST	20	3	60	10	30		Not assessed
Andersen et al[4]							
2 min of exercise	2	5	10	10	50		65%
12 min of exercise	12	5	60	10	50		67%
Andersen et al[2]							
1 workout per week	60	1	60	20	20		49%
3 workouts per week	20	3	60	20	60		60%
9 workouts per week	7	9	63	20	180		60%
Andias et al[5] Endurance	15, progressing to 30	1, but only final 3 were exercise	17.5	4	4 sessions, but 3 were exercise		Not assessed
Brage et al[9] Mixed	90 education and 30 training	Education every 2 wk, training weekly plus 14 HEP sessions	30	8	12 (4 education, 8 training) plus 112 HEP	4 and 12 mo	>75%
de Araujo Cazotti et al[11] Pilates	60	2	120	12	24	6 mo	90.6% completed all sessions
Falla et al[12] Endurance	30 supervised, 10-20 unsupervised	1 supervised, 14 unsupervised	30	8	8 supervised and 112 unsupervised		Not assessed
Im et al[22] Strength	30	3	90	4	12		Not assessed
Kuijper et al[24]							
Mixed	Not specified	2 with HEP	UC	6	12	6 mo	52% exercised >10 min/d for 3 wk, and 43% for the next 3 wk
Intervention	Not specified	2	UC	6	12	6 mo	
Lauche et al[26]							
Tai chi	75-90	1	82.5	12	12	3 mo	Tai chi, 7.6/12 sessions
Neck exercise	60-75	1	67.5	12	12		Neck exercise, 5.4/12 sessions
Revel et al[31] Proprioception	30-40	2	65.6	8	15	10 wk	Not assessed
Ris et al[32] Mixed	30	HEP twice daily, physical training: 3	UC	16	48 training sessions and 224 HEP		Exercise group, 67%; control group, 61%
von Trott et al[39] Exercise, qigong	45	2	90	12	24		Not assessed

Abbreviations: GFT, general fitness training; HEP, home exercise program; SST, specific strength training; UC, unable to be calculated.

for people with neck pain. Variability of exercise therapy may have impacted our results by increasing the between-trial variability, while it may not have impacted the between-group variability.

The lack of reporting on exercise intensity makes it difficult to determine ideal exercise therapy recommendations. Therefore, clinical trials that clearly define exercise therapy parameters are necessary to determine optimal dosage and exercise prescription for the management of individuals with neck pain. Limited reporting of exercise dosage, including intensity, is common within clinical research trials.[20,39] To improve the quality

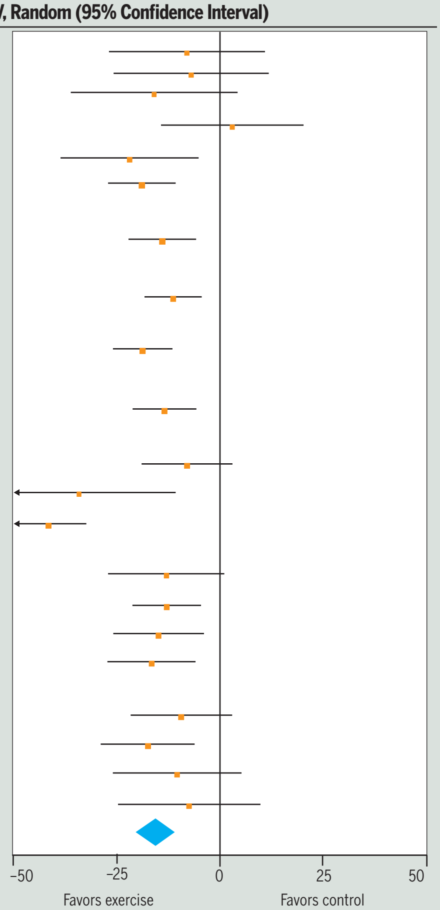
TABLE 4 SUMMARY OF FINDINGS FOR PAIN INTENSITY: EXERCISE COMPARED WITH CONTROL

Pain Intensity	Mean Difference/Effect Size[a]	Participants, n (n of studies)	Quality of Evidence (GRADE)
Overall (intermediate- to long-term follow-up)	−15.32 (−19.20, −11.44) Clinically meaningful	1080 (13)	Moderate[b]
Intermediate term (4-9 wk)	−13.36 (−18.57, −8.14) Clinically meaningful	242 (5)	Low[b,c]
Long term (≥10 wk)	−15.51 (−20.28, −10.74) Clinically meaningful	838 (8)	Low[b,d]

Abbreviation: GRADE, Grading of Recommendations Assessment, Development and Evaluation.
[a]Values in parentheses are 95% confidence interval.
[b]Downgraded 1 level due to risk of bias.
[c]Downgraded 1 level due to imprecision.
[d]Downgraded 1 level due to inconsistency.

Overall Pain Intensity (0-100)

Study	Exercise Mean ± SD	Exercise Total, n	Control Mean ± SD	Control Total, n	Weight	MD IV, Random (95% Confidence Interval)
Ahlgren et al[1] (coordination)	30 ± 17	25	38 ± 24	7	2.9%	−8.00 (−26.99, 10.99)
Ahlgren et al[1] (endurance)	31 ± 17	28	38 ± 24	7	2.9%	−7.00 (−25.86, 11.86)
Ahlgren et al[1] (strength)	22 ± 18	29	38 ± 24	6	2.6%	−16.00 (−36.29, 4.29)
Andersen et al[3] (GFT)	33 ± 12	16	30 ± 22	7	3.2%	3.00 (−14.33, 20.33)
Andersen et al[3] (SST)	8 ± 9	18	30 ± 22	7	3.4%	−22.00 (−38.82, −5.18)
Andersen et al[4] (12 min of exercise)	−18 ± 24.7	65	1 ± 16.3	32	6.5%	−19.00 (−27.24, −10.76)
Andersen et al[4] (2 min of exercise)	−13 ± 24.3	63	1 ± 16.3	32	6.5%	−14.00 (−22.24, −5.76)
Andersen et al[2] (1 workout per week)	−23.1 ± 10.5	43	−11.7 ± 12.8	16	7.1%	−11.40 (−18.41, −4.39)
Andersen et al[2] (3 workouts per week)	−30.5 ± 11.7	40	−11.7 ± 12.8	16	7.0%	−18.80 (−26.04, −11.56)
Andersen et al[2] (9 workouts per week)	−25.2 ± 12.2	27	−11.7 ± 12.8	16	6.7%	−13.50 (−21.28, −5.72)
Andias et al[5] (endurance)	−11 ± 18.3	21	−3 ± 18.8	22	5.2%	−8.00 (−19.09, 3.09)
Brage et al[9] (mixed)	25.7 ± 19	7	60 ± 27.3	8	2.1%	−34.30 (−57.88, −10.72)
de Araujo Cazotti et al[11] (Pilates)	13 ± 16.6	32	54.7 ± 20.9	32	6.0%	−41.70 (−50.95, −32.45)
Falla et al[12] (endurance)	36 ± 24	22	49 ± 23	20	4.1%	−13.00 (−27.22, 1.22)
Im et al[22] (strength)	31 ± 11	8	44 ± 5	7	6.4%	−13.00 (−21.47, −4.53)
Kuijper et al[24] (unknown)	36.2 ± 31	66	51.1 ± 32.7	61	5.2%	−14.90 (−26.00, −3.80)
Lauche et al[26] (neck exercise)	25.2 ± 18.3	37	41.8 ± 20	19	5.4%	−16.60 (−27.35, −5.85)
Lauche et al[26] (tai chi)	32.4 ± 22.5	38	41.8 ± 22.5	19	4.7%	−9.40 (−21.79, 2.99)
Revel et al[31] (proprioception)	−21.8 ± 25.2	30	−4.3 ± 19.6	30	5.1%	−17.50 (−28.92, −6.08)
von Trott et al[39] (exercise)	44.5 ± 25.7	35	54.9 ± 28.5	18	3.7%	−10.40 (−26.08, 5.28)
von Trott et al[39] (qigong)	47.4 ± 30.8	31	54.9 ± 28.5	17	3.2%	−7.50 (−24.85, 9.85)
Total[a]		681		399	100.0%	−15.32 (−19.20, −11.44)

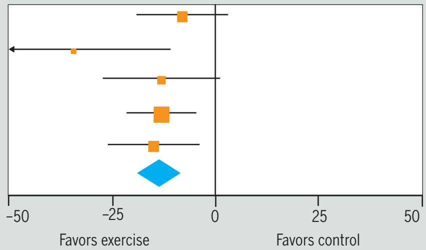


[a]Heterogeneity: $\tau^2 = 42.58$, $\chi^2 = 47.26$, $df = 20$ ($P = .0005$), $I^2 = 58\%$. Test for overall effect: $z = 7.74$ ($P<.001$).
Abbreviations: GFT, general fitness training; IV, inverse variance; MD, mean difference; SST, specific strength training.

FIGURE 2. Meta-analysis for exercise therapy versus control for overall pain intensity.

Intermediate-Term Pain Intensity (0-100)

Study	Exercise Mean ± SD	Exercise Total, n	Control Mean ± SD	Control Total, n	Weight	MD IV, Random (95% Confidence Interval)
Andias et al[5] (endurance)	−11 ± 18.3	21	−3 ± 18.8	22	22.0%	−8.00 (−19.09, 3.09)
Brage et al[9] (mixed)	25.7 ± 19	7	60 ± 27.3	8	4.9%	−34.30 (−57.88, −10.72)
Falla et al[12] (endurance)	36 ± 24	22	49 ± 23	20	13.4%	−13.00 (−27.22, 1.22)
Im et al[22] (strength)	31 ± 11	8	44 ± 5	7	37.7%	−13.00 (−21.47, −4.53)
Kuijper et al[24] (unknown)	36.2 ± 31	66	51.1 ± 32.7	61	22.0%	−14.90 (−26.00, −3.80)
Total[a]		124		118	100.0%	−13.36 (−18.57, −8.14)

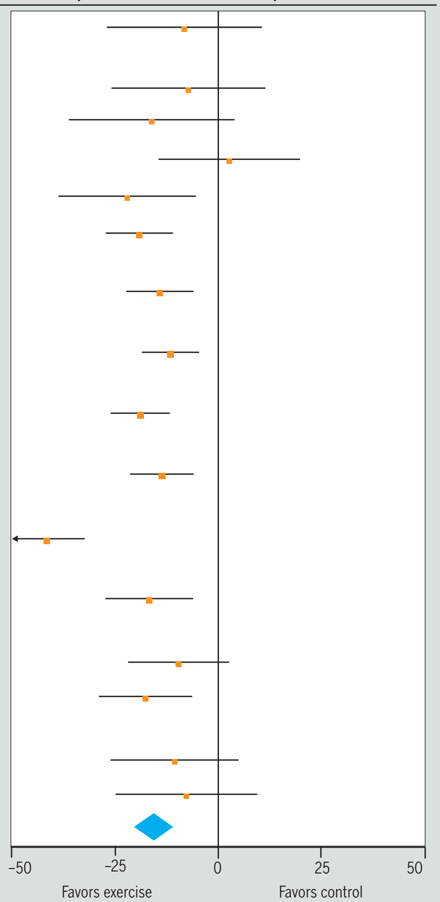


[a]Heterogeneity: $\tau^2 = 0.10$, $\chi^2 = 4.01$, $df = 4$ ($P = .40$), $I^2 = 0\%$. Test for overall effect: $z = 5.02$ ($P<.001$).
Abbreviations: IV, inverse variance; MD, mean difference.

FIGURE 3. Meta-analysis for exercise therapy versus control for intermediate-term pain intensity.

Long-Term Pain Intensity (0-100)

Study	Exercise Mean ± SD	Exercise Total, n	Control Mean ± SD	Control Total, n	Weight	MD IV, Random (95% Confidence Interval)
Ahlgren et al[1] (coordination)	30 ± 17	25	38 ± 24	7	4.0%	−8.00 (−26.99, 10.99)
Ahlgren et al[1] (endurance)	31 ± 17	28	38 ± 24	7	4.0%	−7.00 (−25.86, 11.86)
Ahlgren et al[1] (strength)	22 ± 18	29	38 ± 24	6	3.7%	−16.00 (−36.29, 4.29)
Andersen et al[3] (GFT)	33 ± 12	16	30 ± 22	7	4.5%	3.00 (−14.33, 20.33)
Andersen et al[3] (SST)	8 ± 9	18	30 ± 22	7	4.6%	−22.00 (−38.82, −5.18)
Andersen et al[4] (12 min of exercise)	−18 ± 24.7	65	1 ± 16.3	32	8.2%	−19.00 (−27.24, −10.76)
Andersen et al[4] (2 min of exercise)	−13 ± 24.3	63	1 ± 16.3	32	8.2%	−14.00 (−22.24, −5.76)
Andersen et al[2] (1 workout per week)	−23.1 ± 10.5	43	−11.7 ± 12.8	16	8.8%	−11.40 (−18.41, −4.39)
Andersen et al[2] (3 workouts per week)	−30.5 ± 11.7	40	−11.7 ± 12.8	16	8.7%	−18.80 (−26.04, −11.56)
Andersen et al[2] (9 workouts per week)	−25.2 ± 12.2	27	−11.7 ± 12.8	16	8.4%	−13.50 (−21.28, −5.72)
de Araujo Cazotti et al[11] (Pilates)	13 ± 16.6	32	54.7 ± 20.9	32	7.7%	−41.70 (−50.95, −32.45)
Lauche et al[26] (neck exercise)	25.2 ± 18.3	37	41.8 ± 20	19	7.0%	−16.60 (−27.35, −5.85)
Lauche et al[26] (tai chi)	32.4 ± 22.5	38	41.8 ± 22.5	19	6.3%	−9.40 (−21.79, 2.99)
Revel et al[31] (proprioception)	−21.8 ± 25.2	30	−4.3 ± 19.6	30	6.7%	−17.50 (−28.92, −6.08)
von Trott et al[39] (exercise)	44.5 ± 25.7	35	54.9 ± 28.5	18	5.0%	−10.40 (−26.08, 5.28)
von Trott et al[39] (qigong)	47.4 ± 30.8	31	54.9 ± 28.5	17	4.5%	−7.50 (−24.85, 9.85)
Total[a]		557		281	100.0%	−15.51 (−20.28, −10.74)



[a]Heterogeneity: $\tau^2 = 54.79$, $\chi^2 = 42.15$, $df = 15$ ($P = .0002$), $I^2 = 64\%$. Test for overall effect: $z = 6.37$ ($P<.001$).
Abbreviations: GFT, general fitness training; IV, inverse variance; MD, mean difference; SST, specific strength training.

FIGURE 4. Meta-analysis for exercise therapy versus control for long-term pain intensity.

of exercise intervention reporting, the Consensus on Exercise Reporting Template (CERT) guidelines and template should guide future research. Using the CERT guidelines will help improve transparency, replication, and clinical implementation of exercise-based treatment programs.[37] The Template for Intervention Description and Replication (TIDieR) guidelines were developed to improve general intervention protocol reporting.[20,41] We encourage future trials to report interventions according to the CERT and TIDieR guidelines, to enhance clinical replication and help identify optimal exercise dosage.

TABLE 5 SUMMARY OF FINDINGS FOR DISABILITY: EXERCISE COMPARED WITH CONTROL

Disability	Mean Difference/ Effect Size[a]	Participants, n (n of studies)	Quality of Evidence (GRADE)
Overall (intermediate- to long-term follow-up)	–3.64 (–6.19, –1.09) Small	678 (7)	Low[b,c]
Intermediate term (4-9 wk)	–2.50 (–4.80, –0.20) Small	199 (4)	Low[b,d]
Long term (≥10 wk): NDI	–4.23 (–8.75, 0.29) No difference	378 (3)	Very low[b-d]
Long term (≥10 wk): NPAD	–2.58 (–7.24, 2.08) No difference	101 (1)	Moderate[d]

Abbreviations: GRADE, Grading of Recommendations Assessment, Development and Evaluation; NDI, Neck Disability Index; NPAD, Neck Pain and Disability scale.
[a]Values in parentheses are 95% confidence interval.
[b]Downgraded 1 level due to risk of bias.
[c]Downgraded 1 level due to inconsistency.
[d]Downgraded 1 level due to imprecision.

Clinical Implications

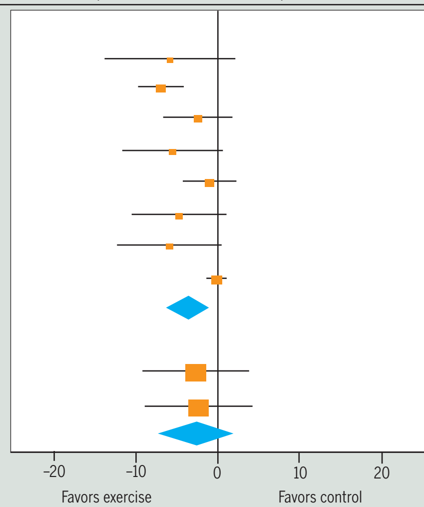
Exercise was superior to control for improving neck pain and disability. Consistent with guideline recommendations,[8] clinicians should consider incorporating exercise into management of neck pain. Exercises to stretch, strengthen, and increase endurance of the cervical muscles and supportive postural muscles may have an effect, but no specific exercise dosage appears to provide a more meaningful difference in improvement. Increasing the amount of exercise the patient receives in the clinic may not improve outcomes. These results can be applied to patients with chronic neck pain because the population of individuals included in this study had symptoms ranging from 4 to more than 1000 weeks.

Limitations

There are likely confounding factors impacting the overall recovery of patients with neck pain disorders, beyond the effects of exercise therapy, not accounted for in this review. Various psychosocial issues,[10] poor overall health,[10] higher

Overall Disability (0-50)

Instrument/Study	Exercise Mean ± SD	Exercise Total, n	Control Mean ± SD	Control Total, n	Weight	MD IV, Random (95% Confidence Interval)
Neck Disability Index						
Brage et al[9] (mixed)	12.71 ± 8.1	7	18.63 ± 7.71	8	6.8%	–5.92 (–13.95, 2.11)
de Araujo Cazotti et al[11] (Pilates)	3.56 ± 3.74	32	10.59 ± 7.19	32	16.6%	–7.03 (–9.84, –4.22)
Falla et al[12] (endurance)	14.1 ± 6.6	22	16.6 ± 7.4	20	13.2%	–2.50 (–6.76, 1.76)
Im et al[22] (strength)	15.8 ± 6.2	8	21.4 ± 6	7	9.4%	–5.60 (–11.78, 0.58)
Kuijper et al[24] (unknown)	13.9 ± 8.85	66	14.95 ± 10	61	15.4%	–1.05 (–4.34, 2.24)
Lauche et al[26] (neck exercise)	22.7 ± 9.3	37	27.5 ± 11.4	20	10.0%	–4.80 (–10.63, 1.03)
Lauche et al[26] (tai chi)	21.5 ± 12.2	38	27.5 ± 11.4	19	9.0%	–6.00 (–12.43, 0.43)
Ris et al[32] (mixed)	–1.3 ± 4.7	101	–1.1 ± 4.3	99	19.8%	–0.20 (–1.45, 1.05)
Subtotal[a]		311		266	100.0%	–3.64 (–6.19, –1.09)
Neck Pain and Disability scale						
von Trott et al[39] (exercise)	16.8 ± 12.75	35	19.55 ± 10.85	18	50.5%	–2.75 (–9.30, 3.80)
von Trott et al[39] (qigong)	17.15 ± 11.8	31	19.55 ± 10.85	17	49.5%	–2.40 (–9.02, 4.22)
Subtotal[b]		66		35	100.0%	–2.58 (–7.24, 2.08)

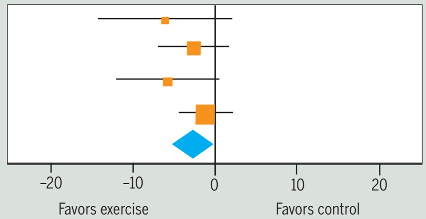


[a]Heterogeneity: τ^2 = 8.18, χ^2 = 25.06, df = 7 (P = .0007), I^2 = 72%. Test for overall effect: z = 2.79 (P = .005).
[b]Heterogeneity: τ^2 = 0.00, χ^2 = 0.01, df = 1 (P = .94), I^2 = 0%. Test for overall effect: z = 1.08 (P = .28). Test for subgroup differences: χ^2 = 0.15, df = 1 (P = .70), I^2 = 0%.
Abbreviations: IV, inverse variance; MD, mean difference.

FIGURE 5. Meta-analysis for exercise therapy versus control for overall disability.

Intermediate-Term Disability (0-50)

Study	Exercise Mean ± SD	Exercise Total, n	Control Mean ± SD	Control Total, n	Weight	MD IV, Random (95% Confidence Interval)
Brage et al[9] (mixed)	12.71 ± 8.1	7	18.63 ± 7.71	8	8.2%	−5.92 (−13.95, 2.11)
Falla et al[12] (endurance)	14.1 ± 6.6	22	16.6 ± 7.4	20	29.2%	−2.50 (−6.76, 1.76)
Im et al[22] (strength)	15.8 ± 6.2	8	21.4 ± 6	7	13.8%	−5.60 (−11.78, 0.58)
Kuijper et al[24] (mixed)	13.9 ± 8.85	66	14.95 ± 10	61	48.8%	−1.05 (−4.34, 2.24)
Total[a]		103		96	100.0%	−2.50 (−4.80, −0.20)

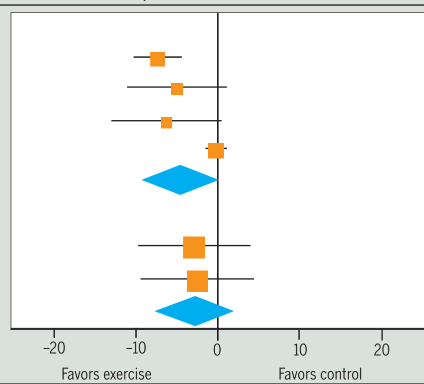


[a]*Heterogeneity: $\tau^2 = 0.00$, $\chi^2 = 2.41$, df = 3 (P = .49), $I^2 = 0\%$. Test for overall effect: z = 2.13 (P = .03).*
Abbreviations: IV, inverse variance; MD, mean difference.

FIGURE 6. Meta-analysis for exercise therapy versus control for intermediate-term disability.

Long-Term Disability (0-50)

Instrument/Study	Exercise Mean ± SD	Exercise Total, n	Control Mean ± SD	Control Total, n	Weight	MD IV, Random (95% Confidence Interval)
Neck Disability Index						
de Araujo Cazotti et al[11] (Pilates)	3.56 ± 3.74	32	10.59 ± 7.19	32	28.5%	−7.03 (−9.84, −4.22)
Lauche et al[26] (neck exercise)	22.7 ± 9.3	37	27.5 ± 11.4	20	20.9%	−4.80 (−10.63, 1.03)
Lauche et al[26] (tai chi)	21.5 ± 12.2	38	27.5 ± 11.4	19	19.4%	−6.00 (−12.43, 0.43)
Ris et al[32] (mixed)	−1.3 ± 4.7	101	−1.1 ± 4.3	99	31.2%	−0.20 (−1.45, 1.05)
Subtotal[a]		208		170	100.0%	−4.23 (−8.75, 0.29)
Neck Pain and Disability scale						
von Trott et al[39] (exercise)	16.8 ± 12.75	35	19.55 ± 10.85	18	50.5%	−2.75 (−9.30, 3.80)
von Trott et al[39] (qigong)	17.15 ± 11.8	31	19.55 ± 10.85	17	49.5%	−2.40 (−9.02, 4.22)
Subtotal[b]		66		35	100.0%	−2.58 (−7.24, 2.08)



[a]*Heterogeneity: $\tau^2 = 16.62$, $\chi^2 = 22.16$, df = 3 (P<.001), $I^2 = 86\%$. Test for overall effect: z = 1.84 (P = .07).*
[b]*Heterogeneity: $\tau^2 = 0.00$, $\chi^2 = 0.01$, df = 1 (P = .94), $I^2 = 0\%$. Test for overall effect: z = 1.08 (P = .28). Test for subgroup differences: $\chi^2 = 0.25$, df = 1 (P = .62), $I^2 = 0\%$.*
Abbreviations: IV, inverse variance; MD, mean difference.

FIGURE 7. Meta-analysis for exercise therapy versus control for long-term disability.

pain intensities and disability levels, and coexisting musculoskeletal problems[28] are all established prognostic factors for neck pain recovery. The known effect of exercise therapy based on exercise dosage cannot be definitively determined by our review.

The potential for publication bias exists, as we only included articles written in English. Our hand search was limited, as it did not include the reference sections of other articles, and therefore some relevant articles may not have been identified. Seven of the 14 included trials were considered to have a high risk of bias, 4 trials were considered to have an unclear risk of bias, and 3 trials were considered to have a low risk of bias. We were unable to analyze the impact of exercise type on outcomes secondary to heterogeneity of reporting and an insufficient sample of studies addressing different exercise types.

The heterogeneous nature of the exercise regimen and outcomes reported made it difficult to identify ideal exercise dosage for patients with neck pain. Furthermore, we could not control for all exercise dosage aspects. This includes controlling for program adherence when calculating the total dosage of exercise secondary to inconsistent adherence reporting. Finally, only 2 studies had both low risk of bias and a 75% adherence rate for patients completing the exercise protocol. Therefore, it appears that exercise adherence is a problematic design feature in trials involving patients with chronic neck pain. Future studies should continue to examine approaches to exercise prescription to improve overall adherence.

CONCLUSION

Exercise therapy was effective in reducing pain and disability in individuals with mechanical neck dis-

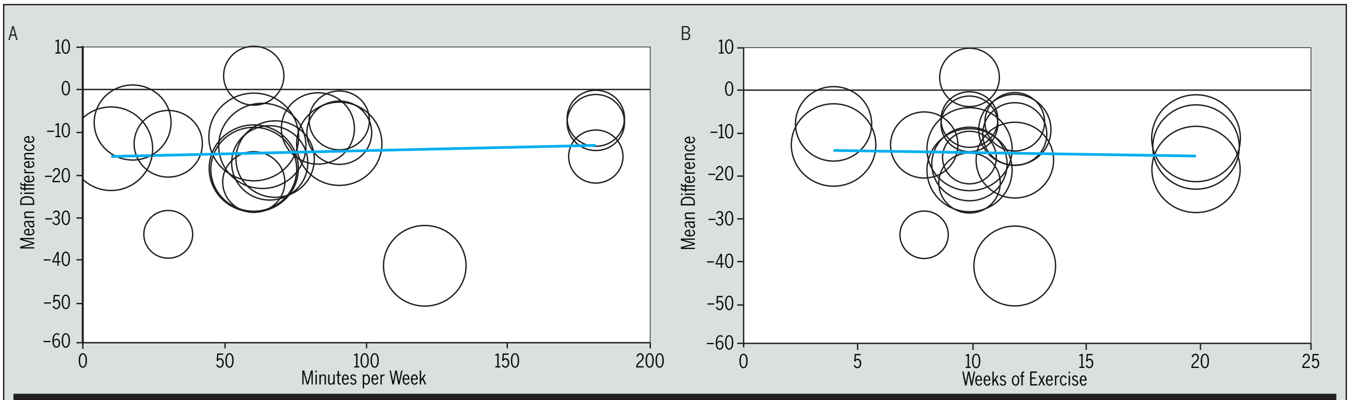


FIGURE 8. Meta-regression models of exercise dosage on treatment effect for pain intensity. (A) Minutes of exercise per week: regression coefficient = 0.00 (95% confidence interval: 0.00, 0.00). (B) Weeks of exercise: regression coefficient = 0.00 (95% confidence interval: −0.05, 0.04). Each circle represents 1 exercise-versus-control comparison.

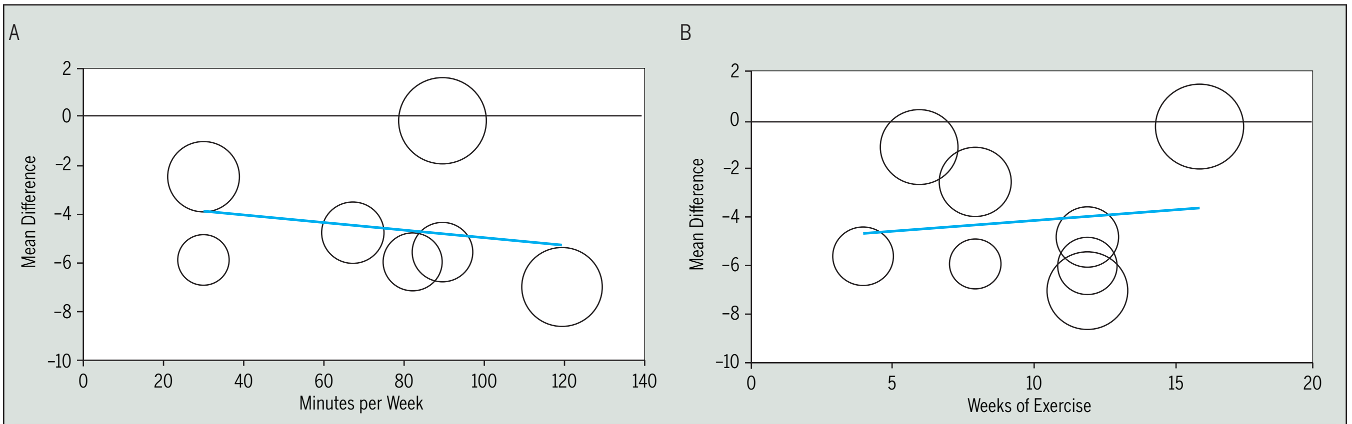


FIGURE 9. Meta-regression models of exercise dosage on treatment effect for disability. (A) Minutes of exercise per week: regression coefficient = 0.00 (95% confidence interval: 0.00, 0.01). (B) Weeks of exercise: regression coefficient = 0.01 (95% confidence interval: −0.03, 0.07). Each circle represents 1 exercise-versus-control comparison.

orders. However, the exact influence of exercise dosage remains to be determined, as the included trials were heterogeneous in the nature of the exercise regimens. Based on our findings, clinicians are encouraged to use exercise in the management of neck pain and disability.

KEY POINTS

FINDINGS: Regardless of exercise dosage (frequency, number of treatment sessions, duration of care), the exercise group demonstrated significantly greater improvements in pain and disability than the control group.

IMPLICATIONS: Exercise therapy was effective in reducing pain and disability in people with mechanical neck disorders.

CAUTION: The majority of studies included in this review were at high risk of bias. Therefore, the results should be generalized with caution.

STUDY DETAILS

AUTHOR CONTRIBUTIONS: All authors contributed to drafting and revising the content, gave final approval of the version to be published, and agreed to be accountable for all aspects of the work. Additionally, the following authors contributed to conception and design of the work: Drs Wilhelm, Donaldson, Griswold, Learman, and Cleland and Shane Learman. All authors contributed to acquisition, analysis, and interpretation of the data.

DATA SHARING: Data are available on request from the corresponding author. All data have been extracted from the included articles.

PATIENT AND PUBLIC INVOLVEMENT: There was no patient, public, or athlete involvement in this study.

REFERENCES

1. Ahlgren C, Waling K, Kadi F, Djupsjöbacka M, Thornell LE, Sundelin G. Effects on physical performance and pain from three dynamic training programs for women with work-related trapezius myalgia. *J Rehabil Med*. 2001;33:162-169. https://doi.org/10.1080/16501970118258
2. Andersen CH, Andersen LL, Gram B, et al. Influence of frequency and duration of strength training for effective management of neck and shoulder pain: a randomised controlled trial. *Br J Sports Med*. 2012;46:1004-1010. https://doi.org/10.1136/bjsports-2011-090813
3. Andersen LL, Kjær M, Søgaard K, Hansen L, Kryger AI, Sjøgaard G. Effect of two contrasting types of physical exercise on chronic neck muscle pain. *Arthritis Rheum*. 2008;59:84-91.

https://doi.org/10.1002/art.23256
4. Andersen LL, Saervoll CA, Mortensen OS, Poulsen OM, Hannerz H, Zebis MK. Effectiveness of small daily amounts of progressive resistance training for frequent neck/shoulder pain: randomised controlled trial. *Pain*. 2011;152:440-446. https://doi.org/10.1016/j.pain.2010.11.016
5. Andias R, Neto M, Silva AG. The effects of pain neuroscience education and exercise on pain, muscle endurance, catastrophizing and anxiety in adolescents with chronic idiopathic neck pain: a school-based pilot, randomized and controlled study. *Physiother Theory Pract*. 2018;34:682-691. https://doi.org/10.1080/09593985.2018.1423590
6. Beltran-Alacreu H, López-de-Uralde-Villanueva I, Fernández-Carnero J, La Touche R. Manual therapy, therapeutic patient education, and therapeutic exercise, an effective multimodal treatment of nonspecific chronic neck pain: a randomized controlled trial. *Am J Phys Med Rehabil*. 2015;94:887-897. https://doi.org/10.1097/PHM.0000000000000293
7. Bertozzi L, Gardenghi I, Turoni F, et al. Effect of therapeutic exercise on pain and disability in the management of chronic nonspecific neck pain: systematic review and meta-analysis of randomized trials. *Phys Ther*. 2013;93:1026-1036. https://doi.org/10.2522/ptj.20120412
8. Blanpied PR, Gross AR, Elliott JM, et al. Neck pain: revision 2017. *J Orthop Sports Phys Ther*. 2017;47:A1-A83. https://doi.org/10.2519/jospt.2017.0302
9. Brage K, Ris I, Falla D, Søgaard K, Juul-Kristensen B. Pain education combined with neck- and aerobic training is more effective at relieving chronic neck pain than pain education alone – a preliminary randomized controlled trial. *Man Ther*. 2015;20:686-693. https://doi.org/10.1016/j.math.2015.06.003
10. Carroll LJ, Hogg-Johnson S, Côté P, et al. Course and prognostic factors for neck pain in workers: results of the Bone and Joint Decade 2000-2010 Task Force on Neck Pain and Its Associated Disorders. *Spine (Phila Pa 1976)*. 2008;33:S93-S100. https://doi.org/10.1097/BRS.0b013e31816445d4
11. de Araujo Cazotti L, Jones A, Roger-Silva D, Ribeiro LHC, Natour J. Effectiveness of the Pilates method in the treatment of chronic mechanical neck pain: a randomized controlled trial. *Arch Phys Med Rehabil*. 2018;99:1740-1746. https://doi.org/10.1016/j.apmr.2018.04.018
12. Falla D, Lindstrøm R, Rechter L, Boudreau S, Petzke F. Effectiveness of an 8-week exercise programme on pain and specificity of neck muscle activity in patients with chronic neck pain: a randomized controlled study. *Eur J Pain*. 2013;17:1517-1528. https://doi.org/10.1002/j.1532-2149.2013.00321.x
13. Furlan AD, Malmivaara A, Chou R, et al. 2015 updated method guideline for systematic reviews in the Cochrane Back and Neck Group. *Spine (Phila Pa 1976)*. 2015;40:1660-1673. https://doi.org/10.1097/BRS.0000000000001061
14. Gross A, Kay TM, Paquin JP, et al. Exercises for mechanical neck disorders. *Cochrane Database Syst Rev*. 2015;1:CD004250. https://doi.org/10.1002/14651858.CD004250.pub5
15. Gross AR, Paquin JP, Dupont G, et al. Exercises for mechanical neck disorders: a Cochrane review update. *Man Ther*. 2016;24:25-45. https://doi.org/10.1016/j.math.2016.04.005
16. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines 6. Rating the quality of evidence—imprecision. *J Clin Epidemiol*. 2011;64:1283-1293. https://doi.org/10.1016/j.jclinepi.2011.01.012
17. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011;343:d5928. https://doi.org/10.1136/bmj.d5928
18. Higgins JPT, Thomas J, Chandler J, et al. *Cochrane Handbook for Systematic Reviews of Interventions*. 2nd ed. Chichester, UK: Wiley; 2019.
19. Higgins JPT, 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
20. Hoffmann TC, Glasziou PP, Boutron I, et al. Better reporting of interventions: Template for Intervention Description and Replication (TIDieR) checklist and guide. *BMJ*. 2014;348:g1687. https://doi.org/10.1136/bmj.g1687
21. Hoy DG, Protani M, De R, Buchbinder R. The epidemiology of neck pain. *Best Pract Res Clin Rheumatol*. 2010;24:783-792. https://doi.org/10.1016/j.berh.2011.01.019
22. Im B, Kim Y, Chung Y, Hwang S. Effects of scapular stabilization exercise on neck posture and muscle activation in individuals with neck pain and forward head posture. *J Phys Ther Sci*. 2016;28:951-955. https://doi.org/10.1589/jpts.28.951
23. Kay TM, Gross A, Goldsmith CH, et al. Exercises for mechanical neck disorders. *Cochrane Database Syst Rev*. 2012:CD004250. https://doi.org/10.1002/14651858.CD004250.pub4
24. Kuijper B, Tans JT, Beelen A, Nollet F, de Visser M. Cervical collar or physiotherapy versus wait and see policy for recent onset cervical radiculopathy: randomised trial. *BMJ*. 2009;339:b3883. https://doi.org/10.1136/bmj.b3883
25. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-174. https://doi.org/10.2307/2529310
26. Lauche R, Stumpe C, Fehr J, et al. The effects of tai chi and neck exercises in the treatment of chronic nonspecific neck pain: a randomized controlled trial. *J Pain*. 2016;17:1013-1027. https://doi.org/10.1016/j.jpain.2016.06.004
27. Liu R, Kurihara C, Tsai HT, et al. Classification and treatment of chronic neck pain: a longitudinal cohort study. *Reg Anesth Pain Med*. 2017;42:52-61. https://doi.org/10.1097/AAP.0000000000000505
28. McLean SA, Ulirsch JC, Slade GD, et al. Incidence and predictors of neck and widespread pain after motor vehicle collision among US litigants and nonlitigants. *Pain*. 2014;155:309-321. https://doi.org/10.1016/j.pain.2013.10.016
29. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med*. 2009;151:264-269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135
30. Parikh P, Santaguida P, Macdermid J, Gross A, Eshtiaghi A. Comparison of CPG's [sic] for the diagnosis, prognosis and management of non-specific neck pain: a systematic review. *BMC Musculoskelet Disord*. 2019;20:81. https://doi.org/10.1186/s12891-019-2441-3
31. Revel M, Minguet M, Gregoy P, Vaillant J, Manuel JL. Changes in cervicocephalic kinesthesia after a proprioceptive rehabilitation program in patients with neck pain: a randomized controlled study. *Arch Phys Med Rehabil*. 1994;75:895-899. https://doi.org/10.1016/0003-9993(94)90115-5
32. Ris I, Søgaard K, Gram B, Agerbo K, Boyle E, Juul-Kristensen B. Does a combination of physical training, specific exercises and pain education improve health-related quality of life in patients with chronic neck pain? A randomised control trial with a 4-month follow up. *Man Ther*. 2016;26:132-140. https://doi.org/10.1016/j.math.2016.08.004
33. Rohatgi A. WebPlotDigitizer. Available at: https://automeris.io/WebPlotDigitizer/. Accessed October 31, 2019.
34. Rubinstein SM, Terwee CB, Assendelft WJ, de Boer MR, van Tulder MW. Spinal manipulative therapy for acute low-back pain. *Cochrane Database Syst Rev*. 2012:CD008880. https://doi.org/10.1002/14651858.CD008880.pub2
35. Salaffi F, Stancati A, Silvestri CA, Ciapetti A, Grassi W. Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale. *Eur J Pain*. 2004;8:283-291. https://doi.org/10.1016/j.ejpain.2003.09.004
36. Saragiotto BT, Maher CG, Yamato TP, et al. Motor control exercise for chronic non-specific low-back pain. *Cochrane Database Syst Rev*. 2016:CD012004. https://doi.org/10.1002/14651858.CD012004
37. Slade SC, Dionne CE, Underwood M, et al. Consensus on Exercise Reporting Template (CERT): modified Delphi study. *Phys Ther*. 2016;96:1514-1524. https://doi.org/10.2522/ptj.20150668
38. Stratford PW, Riddle DL, Binkley JM, Spadoni G, Westaway MD, Padfield B. Using the Neck Disability Index to make decisions concerning individual patients. *Physiother Can*. 1999;51:107-112.
39. von Trott P, Wiedemann AM, Lüdtke R, Reißhauer A, Willich SN, Witt CM. Qigong and exercise therapy for elderly patients with chronic neck pain (QIBANE): a randomized controlled study. *J Pain*. 2009;10:501-508. https://doi.org/10.1016/j.jpain.2008.11.004
40. Wieland LS, Skoetz N, Pilkington K, Vempati R, D'Adamo CR, Berman BM. Yoga treatment for chronic non-specific low back pain. *Cochrane*

Database Syst Rev. 2017;1:CD010671. https://doi.org/10.1002/14651858.CD010671.pub2

41. Yamato T, Maher C, Saragiotto B, et al. The TIDieR checklist will benefit the physical therapy profession. *J Orthop Sports Phys Ther*. 2016;46:402-404. https://doi.org/10.2519/jospt.2016.0108

42. Young JL, Rhon DI, Cleland JA, Snodgrass SJ. The influence of exercise dosing on outcomes in patients with knee disorders: a systematic review. *J Orthop Sports Phys Ther*. 2018;48:146-161. https://doi.org/10.2519/jospt.2018.7637

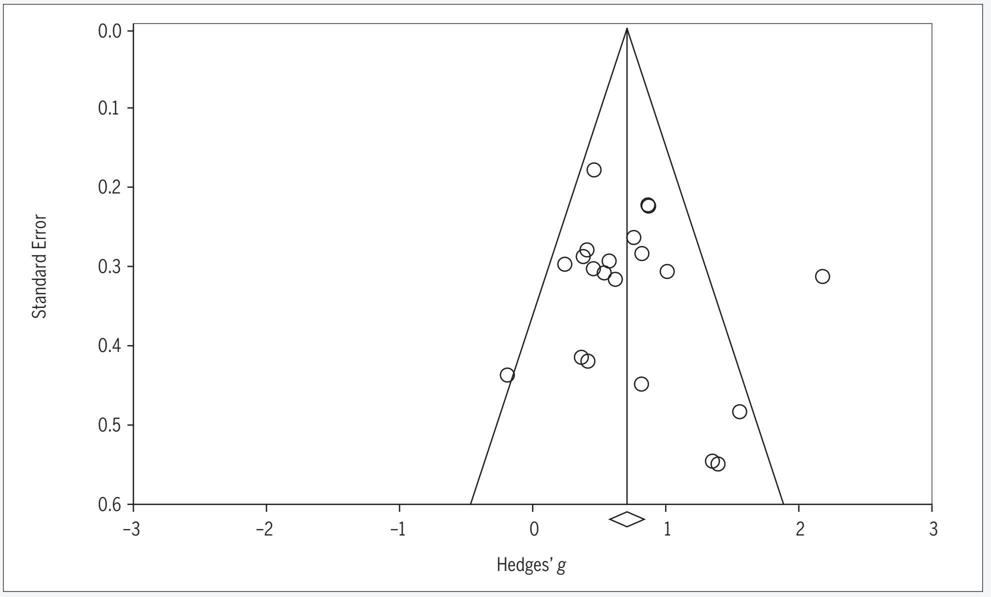
APPENDIX A

PUBMED SEARCH STRATEGY

(((neck pain*[Text Word]) OR cervicobrachial neuralgia[Text Word]) OR cervicobrachial neuralgia[Text Word]) OR cervicogenic headache*[Text Word]) OR post traumatic headache*[Text Word]) OR neck ache*[Text Word]) OR posterior cervical pain*[Text Word]) OR posterior neck pain*[Text Word]) OR anterior cervical pain*[Text Word]) OR anterior neck pain*[Text Word]) OR cervicalgia[Text Word]) OR whiplash[Text Word]) OR cervical pain[Text Word]) OR cervicodynia[Text Word]) OR brachial plexus neuritis[Text Word]) OR neck injur*[Text Word]) OR neck pain[MeSH Terms]) OR radiculopathy[MeSH Terms]) OR radiculopathies[Text Word]) OR cervical radiculopathies[Text Word]) OR radiculitis[Text Word]) OR nerve root compression*[Text Word]) OR posttraumatic headache[MeSH Terms]) OR brachial plexus neuritis[MeSH Terms]) OR neck injuries[MeSH Terms]) OR whiplash injuries[MeSH Terms]) OR spondylolysis[MeSH Terms]) OR torticollis[MeSH Terms]) OR cervical rib syndrome[MeSH Terms]) OR intervertebral disc displacement[MeSH Terms]) OR intervertebral disc displacement*[Text Word]) OR herniated dis*[Text Word]) OR slipped dis*[Text Word]) OR disk prolapse*[Text Word]) OR prolapse disk[Text Word]) OR prolapse disc[Text Word]) OR intervertebral disc degeneration[MeSH Terms]) OR intervertebral disc degeneration*[Text Word]) OR Disk degeneration*[Text Word]) OR disc degeneration*[Text Word]) OR spondylarthritis[MeSH Terms]) OR spinal arthritis[Text Word])) AND (((((((((((((((range of motion, articular[MeSH Terms]) OR joint range of motion[Text Word]) OR joint flexibility[Text Word]) OR range of motion[Text Word]) OR passive range of motion[Text Word]) OR exercise therapy[MeSH Terms]) OR resistance training[MeSH Terms]) OR resistance training[Text Word]) OR strength training[Text Word]) OR exercise movement techniques[MeSH Terms]) OR "physical therapy modalities"[MeSH Terms]) OR exercise therap*[Text Word]) OR strength*[Text Word]) OR rehabilitation exercise[Text Word]) OR movement[MeSH Terms])

APPENDIX B

FUNNEL PLOT FOR OVERALL PAIN INTENSITY



Egger test: intercept, 0.91; standard error, 1.17 (95% confidence interval: −1.53, 3.36); t = 0.79, P = .44.