SHAW BRONNER, PT, PhD, OCS[1] • **EVIE CHODOCK**[2]
IGOR EDWARDO REIS URBANO, PT[3] • **TERESA SMITH,** PT, DPT, OCS, CSCS[4]

Psychometric Properties of the Dance Functional Outcome Survey (DFOS): Reliability, Validity, and Responsiveness

BACKGROUND: There are no outcomes measures that focus on the unique functional requirements of dancers.

OBJECTIVES: To evaluate test-retest reliability, internal consistency, construct validity, sensitivity, and responsiveness of the Dance Functional Outcome Survey (DFOS) in professional and preprofessional adult dancers.

METHODS: This prospective cohort study examined test-retest reliability of the DFOS in 198 healthy and injured dancers over 2 weeks, using intraclass correlation coefficients ($ICC_{2,1}$). In a sample of 725 healthy and injured dancers, the following were examined: (1) construct validity, by comparing the DFOS to the Medical Outcomes Study 36-Item Short-Form Health Survey (SF-36) using Pearson correlations; (2) exploratory factor analysis and internal consistency; and (3) sensitivity, by generating receiver operating characteristic curves and determining area under the curve (AUC). In a subgroup of 47 injured dancers, we determined internal responsiveness across 4 time points using repeated-measures analysis of variance ($P<.05$). Injured dancers' scores were analyzed for floor and ceiling effects.

RESULTS: The DFOS demonstrated high test-retest reliability (ICC≥0.93). Single-factor loading in exploratory factor analysis supported unidimensionality of the scale, with high internal consistency ($\alpha = .96$). The DFOS total score and activities-of-daily-living (ADL) and dance technique subscores had strong construct validity compared with scores on the SF-36 physical component summary ($r \geq 0.77$). This study found excellent sensitivity, with high AUC values (AUC≥0.91). There were significant differences across time for DFOS scores ($P<.001$), demonstrating responsiveness to change. There were no floor or ceiling effects.

CONCLUSION: The DFOS demonstrates acceptable psychometric performance as an outcome and screening measure for dancers. The DFOS is a useful tool to monitor both healthy state and functional limitation following lower extremity or low back injury in adult ballet and modern dancers. *J Orthop Sports Phys Ther 2019;49(2):64-79. Epub 27 Jul 2018. doi:10.2519/jospt.2019.8247*

KEY WORDS: *ballet, health, modern dance, musculoskeletal injury, SF-36, wellness*

Due to the functional impact of injury on dancers, improved prevention, timely triage, and intervention are critical to optimizing their recovery, performance, and well-being. Outcomes measures are important for clinical research and the assessment of patients' physical and emotional function, perceptions about their condition, ability to execute daily activities and tasks, and change in health status over time. Outcome measures may also guide the development of new interventions. Questionnaires with adequate psychometric properties can assist in assessing injury severity, recovery, an individual's well-being and function, performance readiness for athletic activities, planning injury prevention strategies, and guiding the rehabilitation process.[12,20] Outcomes measures in general health care and orthopaedics may be generic, body region specific, or population specific.

Previously, we published a preliminary analysis of a new instrument, the 16-item Dance Functional Outcome Survey (DFOS).[4,43] The DFOS was developed to be a self-report functional-outcome questionnaire for ballet and modern dance populations, applicable to musculoskeletal injuries of the low back and lower extremities, the most commonly affected regions of dance injuries.[36,37] Following item generation of a 20-item questionnaire, we submitted the DFOS to an expert panel of dance medicine health care providers, dance educators, professional dancers, and an outcomes development specialist for face validity assessment. Based on their feedback, we shortened the DFOS to 16 items and tested for internal consistency, test-retest reliability, equivalence reliability (Likert scale versus visual analog

[1]ADAM Center, New York, NY; Alvin Ailey American Dance Theater, New York, NY. [2]School of Kinesiology, University of Michigan, Ann Arbor, MI. [3]Universidade Federal de Minas Gerais, Belo Horizonte, Brazil. [4]Kentucky Orthopedic Rehab Team Physical Therapy, Louisville, KY. This project was approved by the Institutional Review Boards of the following universities: Long Island University (1998-2012), Northeastern University (IRB 13-08-28; 2013-2014), and Boston Children's Hospital (2014-present). The study was funded by the US National Institutes of Health (MBRS grant 2SO6 GM54650) under the title "Effect of Musculoskeletal Injury on Skilled Movement." This funding was not subject to open-access requirements. Study design, data collection, analyses, and writing were supported by the ADAM Center, New York, NY. 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 Shaw Bronner, Alvin Ailey American Dance Theater, 405 West 55th Street, New York, NY 10019. E-mail: shaw.bronner@gmail.com

scale), and concurrent validity compared to 3 established orthopaedic outcomes instruments.

Comparison of the DFOS Likert scale and visual analog scale found high equivalence reliability (r = 0.74), internal consistency (α = .90), and test-retest reliability (r>0.90). Both the Likert scale and visual analog scale demonstrated acceptable construct validity compared to the Cincinnati Knee Rating System, Foot and Ankle Questionnaire, and Oswestry Disability Index. Investigation of individual DFOS items revealed that elimination of 2 items resulted in improved correlations with the 3 orthopaedic instruments. Dancers and administrators preferred the Likert scale due to its ease of use, intelligibility, scoring, and interpretation. Therefore, the DFOS Likert scale was selected for further study.

The revised 14-item DFOS required further reliability and validity testing. The Medical Outcomes Study 36-Item Short-Form Health Survey (SF-36) was selected for validation purposes. The SF-36 is a widely used self-report measure of general health and function in clinical trials. Considered a generic instrument to assess quality of life in the general population, the SF-36 is available in more than 170 languages[32] and has been studied across multiple diagnoses, disease severities, and musculoskeletal injuries. It is frequently used as the principal measure for comparisons with new instruments.[1] The physical component summary (PCS) and related subscales have demonstrated acceptable construct validity compared to lower extremity and back instruments,[1,2,5,16,19,23,24,46] while the mental component summary (MCS) and related subscales have demonstrated divergent evidence.[19,23]

The aims of this study were to investigate test-retest reliability, construct validity, internal consistency, sensitivity, and internal responsiveness of the revised DFOS in adult dancers with and without musculoskeletal injury to the low back or lower extremities.

METHODS

Instruments

The DFOS is a dance-specific, lower extremity and low back functional outcome measure. The 14-question DFOS assesses the dancer's ability in areas of activities of daily living (ADL, 40 points) and dance technique (technique, 50 points) (**APPENDICES A** and **B**). Total points are normalized to a percentage, with 100% representing full function without limitations.

We assessed construct validity of the DFOS with the SF-36, a self-administered questionnaire that measures 8 domains of health-related function, testing both physical and mental health. The domains of the SF-36 include physical functioning, limitations due to physical problems ("role physical"), bodily pain, general health perceptions, vitality, social functioning, limitations due to emotional problems ("role emotional"), and mental health. Each domain is scored as a z score ranging from 0 to 100, with 0 representing severe disability and 100 representing no disability.[48] Standard procedures were used to obtain a PCS score, consisting primarily of the physical functioning, role physical, bodily pain, and general health perceptions domains, and an MCS score, consisting primarily of the vitality, social functioning, role emotional, and mental health domains. Scores for the 2 component summary scores were normalized, such that the mean ± SD value of each composite score is 50 ± 10 for the US general population. Norm-based scoring equates all scores, so scores above 50 are better than the general population average for all scales and summary measures, whereas scores below 50 are worse.[47]

Protocol

Healthy adult dancers from dance companies and preprofessional dance schools and injured dancers from dance medicine physical therapy clinics were recruited to participate in this series of studies. For healthy dancers, inclusion criteria were (1) a minimum of 3 years of dance training, including ballet and/or modern dance; (2) an intermediate to expert skill level; (3) 18 years of age or older; and (4) no low back or lower extremity injury in the previous 3 months. For injured dancers, inclusion criteria were new referral for musculoskeletal injury to the low back or lower extremity. For all dancers, the exclusion criteria were (1) non–English speaking, (2) pregnancy, (3) current active disease processes, and (4) musculoskeletal injury anywhere other than the low back or lower extremity.

All participants gave written consent according to guidelines approved by the Long Island University, Northeastern University (IRB 13-08-28), and Boston Children's Hospital Institutional Review Boards. Upon enrollment, dancers answered a demographics questionnaire, the DFOS, and the SF-36. If they took part in the reliability portion, then they filled out the DFOS a second time within 4 to 9 days.

Subjects

A priori analysis was conducted to determine sample size for test-retest reliability with 1 group (including "healthy" and "injured" dancers), 2 measurements (test-retest), an effect-size change of 0.25, power of 0.95, and a significance level of α = .05,[8] resulting in 54 subjects (**APPENDIX C**). A sample of 198 dancers was used in this analysis. To assess construct validity with 1 group, with an effect-size change of 0.25, power of 0.95, and a significance level of α = .05, a sample of 197 was required. For factor analysis, 10 to 15 participants per item or a minimum of 300 subjects are recommended.[10,33] Therefore, with 14 items, a sample of 140 to 210 subjects was estimated. A sample of 725 dancers was used in the construct validity and factor analyses.

To assess differences between healthy and injured groups in receiver operating characteristic (ROC) analyses, sample-size estimation was conducted using a predetermined level of sensitivity of 80% (alternative hypothesis [H_a] = 0.80, null hypothesis [H_0] = 0.50, α = .05, power

of 0.95).[8,11] A sample of 52 per group or a total sample of 104 was necessary. The same sample of 725 dancers was used in this analysis.

To assess instrument responsiveness to change, a priori analysis for sample size was conducted using a 1-group repeated-measures analysis of variance over 4 time points: healthy ($T_{healthy}$), injured at intake ($T_{injured}$), at discharge ($T_{discharge}$), and at 3-month follow-up (T_{3mo}).[8] With a small effect-size change of 0.25, α = .05, and power of 0.95, a total sample of 36 was required. A sample of 47 dancers who sustained 60 injuries was used in this analysis.

Data Analysis

Demographics, DFOS, and SF-36 data were entered into an Excel 2011 database (Microsoft Corporation, Redmond, WA). Incomplete questionnaires missing more than 2 items were eliminated. For those missing 1 or 2 items (less than 5% of the sample), values were filled in using mean imputation.[40] The DFOS total score and ADL and technique subscores were obtained by summing individual question scores. The SF-36 scores for the 8 domains and composite MCS and PCS scores were obtained using standard procedures.[48] Higher scores for both questionnaires reflected higher function.

Test-retest reliability analysis separately compared combined, healthy, and injured groups for the DFOS total score and ADL and technique subscores using the intraclass correlation coefficient ($ICC_{2,1}$), calculated in SPSS Version 23 (IBM Corporation, Armonk, NY). The ICC values of 0.49 or less were considered low, 0.50 to 0.69 moderate, 0.70 to 0.89 high, and 0.90 to 1.00 very high.[28] For test-retest reliability, we hypothesized high correlations (ICC≥0.70). Absolute reliability, defined as variability of scores from measurement to measurement reflecting measurement accuracy, was measured using the standard error of measurement (SEM).[6,35] The SEM, expressed in the units of original measurement, was calculated from the standard deviation of measurement error, with the assumption that measurement error is normally distributed: $SEM = SD \times \sqrt{1 - r}$, where r is the coefficient alpha.

To determine construct validity of the DFOS versus the SF-36, DFOS total score, subscores, and items were compared to SF-36 PCS, MCS, and domain scores using Pearson correlation coefficients in SPSS (IBM Corporation). Convergent correlations between the SF-36 PCS and DFOS (Pearson r≥0.50) and divergent correlations between the DFOS and SF-36 MCS (Pearson r≤0.49) were hypothesized, with correlation strength interpreted as weak (0.49 or less), moderate (0.50-0.69), or strong (0.70-1.00).[9]

Exploratory factor analysis (EFA), a variable reduction technique, was conducted to identify the number of latent constructs and underlying structure using parallel analysis, eigenvalues, scree plots, suppression of small coefficients, and rotation to determine DFOS factor structure.[7,13,52] Cronbach's alpha was calculated to estimate internal item consistency. The EFA and Cronbach's alpha were conducted in open-source software (JASP Version 0.8.1.2; University of Amsterdam, Amsterdam, the Netherlands). We hypothesized a single-factor model, with item correlations of 0.70 or greater and a Cronbach's alpha of .70 or greater.

To conduct sensitivity analyses in the healthy group and injured group, the researchers conducted a t test for equal variances not assumed, due to unequal sample sizes for each group (healthy, 638; injured, 87) and a significant Levene test (P<.001). Predictive accuracy or sensitivity was measured by generating ROC curve, area under the curve (AUC), and associated 95% confidence interval (CI) for the 3 DFOS scores (total score, ADL and technique subscores) in SPSS. The ROC curves used DFOS scores as test variables, with the binary state or outcome variable coded as 0 (healthy) and 1 (injured). Sensitivity and specificity for cutoff values were determined.

To determine internal responsiveness, we examined differences in DFOS and SF-36 scores in injured dancers across 4 time points using repeated-measures (time) analysis of variance in SPSS. For all analyses, Mauchly's test was used to assess the assumption of sphericity. In the case of significance, the Huynh-Feldt correction was applied to the degrees of freedom and F value if the epsilon value was 0.75 or greater, and the Greenhouse-Geisser correction was used if epsilon was less than 0.75. In these cases, epsilon and corrected values (eg, degrees of freedom and F values) are reported. Pairwise comparisons were conducted where there was a significant main effect. Given the number of dependent variables, a conservative level of significance was set at α = .001. We hypothesized pairwise differences across time points.

Internal responsiveness was defined in 4 ways: SEM, minimal detectable change at the 95% CI (MDC_{95}), standardized response mean (SRM), and effect size, using the following equations: $MDC_{95} = 1.96 \times \sqrt{2} \times SEM$; SRM = mean change in score/SD of change scores; effect size = mean change scores/SD of baseline scores.

The SEM, MDC_{95}, SRM, and effect size were calculated for the DFOS total score and subscores and for the SF-36 PCS and MCS. We anticipated SRM values of 0.80 or greater, demonstrating high responsiveness,[29,39] and large effect sizes (greater than 1.0). Effect-size values between 0.20 and 0.50 were considered small, 0.51 to 0.80 medium, and greater than 0.80 large.[14] Floor and ceiling effects were determined by the percentage of dancers who achieved the lowest and highest DFOS scores, respectively, within the injured group. Floor and ceiling effects of less than 15% of respondents were considered acceptable.[42]

RESULTS

Test-Retest Reliability

One hundred ninety-eight dancers participated in DFOS test-retest reliability analysis (130 female, 68 male; mean ± SD age, 24.56 ± 6.22 years;

range, 18-51 years) (**TABLE 1**). One hundred thirty-seven dancers (69%) were professionals, representing 8 modern and ballet companies, and 61 dancers (31%) were enrolled in preprofessional programs. One hundred six dancers (54%) were categorized as healthy and 92 (46%) as injured.

For combined groups, test-retest reliability values of the DFOS total score and ADL and technique subscores were very high ($ICC_{2,1}$ = 0.99) (**TABLE 2**). Investigation at the item level found that test-retest reliability of all items was high ($ICC_{2,1}$≥0.93). The SEM values were 2.31 (DFOS total), 1.29 (ADL), and 1.86 (technique).

Healthy-group test-retest reliability values of the DFOS total score and ADL and technique subscores were high ($ICC_{2,1}$ = 0.95, 0.89, 0.92, respectively). Item correlations for the DFOS were high (ranging from $ICC_{2,1}$ = 0.70 to 0.93), with the exception of rond de jambe, which was moderate ($ICC_{2,1}$ = 0.67). The SEM ranged from 0.58 to 0.86. Injured-group reliability was high for all DFOS scores and items. The SEM ranged from 1.72 to 3.22.

TABLE 1 Demographics (Test-Retest Reliability)

Demographic/Group	Male	Female	Total
Subjects, n (%)*	68 (34)	130 (66)	198 (100)
Professional dancers	60 (30)	77 (39)	137 (69)
Preprofessional students	8 (4)	53 (27)	61 (31)
Age, y†	24.69 ± 5.26	23.97 ± 6.61	24.56 ± 6.22
Dance training, y†	12.54 ± 5.26	16.17 ± 6.96	14.92 ± 6.55
Dance training, n (%)*			
Ballet	7 (3)	23 (12)	30 (15)
Modern	25 (13)	31 (16)	56 (28)
Modern/ballet	36 (18)	76 (38)	112 (57)
Professional experience, y†	4.73 ± 5.09	4.11 ± 6.19	4.67 ± 5.87
Status, n (%)*			
Healthy	35 (18)	71 (36)	106 (54)
Injured	33 (17)	59 (30)	92 (46)
Ethnicity, n (%)*			
African American	46 (23)	43 (22)	89 (45)
Caucasian	14 (7)	68 (34)	82 (41)
Hispanic	3 (1)	2 (1)	5 (3)
Asian	3 (1)	10 (5)	13 (7)
Other	0 (0)	6 (3)	6 (3)

**All percentages are out of the total n.*
†Values are mean ± SD.

TABLE 2 Test-Retest Reliability of the DFOS*

	Combined		Healthy		Injured	
Measure	ICC	SEM	ICC	SEM	ICC	SEM
DFOS total	0.99 (0.99, 0.99)	2.31	0.95 (0.92, 0.96)	0.86	0.98 (0.94, 0.99)	3.22
ADL	0.99 (0.98, 0.99)	1.29	0.89 (0.84, 0.92)	0.74	0.97 (0.86, 0.99)	1.72
Overall activity	0.97 (0.96, 0.97)		0.72 (0.61, 0.80)		0.92 (0.87, 0.95)	
Movement quality	0.97 (0.96, 0.98)		0.73 (0.63, 0.81)		0.93 (0.89, 0.96)	
Walking	0.93 (0.89, 0.95)		0.77 (0.68, 0.84)		0.90 (0.73, 0.96)	
Stairs	0.96 (0.95, 0.97)		0.71 (0.60, 0.79)		0.94 (0.91, 0.96)	
Stability/symptoms	0.97 (0.97, 0.98)		0.80 (0.72, 0.86)		0.96 (0.94, 0.98)	
Pain	0.93 (0.91, 0.95)		0.70 (0.59, 0.79)		0.89 (0.78, 0.94)	
Technique	0.99 (0.99, 0.99)	1.86	0.92 (0.88, 0.94)	0.58	0.96 (0.94, 0.97)	2.66
Plié	0.95 (0.94, 0.97)		0.93 (0.90, 0.95)		0.90 (0.85, 0.94)	
Développé	0.97 (0.97, 0.98)		0.78 (0.68, 0.85)		0.97 (0.96, 0.98)	
Relevé balance	0.96 (0.95, 0.97)		0.79 (0.71, 0.85)		0.93 (0.89, 0.95)	
Rond de jambe	0.99 (0.99, 0.99)		0.67 (0.51, 0.77)		0.99 (0.98, 0.99)	
Kneeling	0.99 (0.98, 0.99)		0.76 (0.67, 0.83)		0.98 (0.98, 0.99)	
Turning	0.98 (0.98, 0.99)		0.81 (0.73, 0.87)		0.96 (0.94, 0.97)	
Jumping	0.99 (0.99, 0.99)		0.85 (0.79, 0.90)		0.98 (0.98, 0.99)	
Grand allegro	0.94 (0.92, 0.95)		0.85 (0.78, 0.89)		0.72 (0.60, 0.80)	

Abbreviations: ADL, activities of daily living; DFOS, Dance Functional Outcome Survey; ICC, intraclass correlation coefficient; SEM, standard error of measurement.
**Values in parentheses are 95% confidence interval. All correlations were significant (P<.001).*

Construct Validity

A combined subject pool comprising 761 dancers was included in the factor analysis and analyses of internal consistency, construct validity, and sensitivity. Dancers with incomplete questionnaires (more than 2 missing items) were eliminated, resulting in 725 participants (95%) (**TABLE 3**). The group consisted of 45% professional dancers, rehearsal directors, choreographers, and teachers and 55% preprofessional students. Professionals were from 17 dance companies (12 modern and 5 ballet). Students represented 10 preprofessional programs.

Strong Pearson correlations were found between the SF-36 PCS and DFOS total score (r = 0.79) and subscores (ADL, r = 0.79; technique, r = 0.77) (**TABLE 4**). Individual ADL items were compared to PCS domains (physical functioning, role physical, and bodily pain), with correlations ranging from r = 0.50 to 0.69. Individual DFOS items were best correlated to physical functioning, ranging from r = 0.66 to 0.69. In contrast, weak correlations were found between the SF-36 MCS and the DFOS total score (r = 0.26) and subscores (ADL, r = 0.31; technique, r = 0.26). Individual ADL items were compared to MCS domains (vitality, social functioning, mental health, role emotional), with correlations ranging from r = 0.10 to 0.34. Individual DFOS items were best correlated to social functioning but remained weak, ranging from r = 0.32 to 0.34.

Factor Analysis and Internal Consistency

Data from the same group of 725 participants were used in EFA and to determine internal consistency (Cronbach's alpha). Initial parallel analysis produced a root-mean-square error of approximation of 0.80 and a Tucker-Lewis index of 0.962, with 4 factors and item loadings ranging from 0.32 to 0.91. The EFA was rerun, using Kaiser's criterion of eigenvalues greater than 1 point of inflection within a scree plot (**FIGURE 1**), suppression of coefficients less than 0.30, and rotation to determine best fit using oblique oblimin rotation for correlated variables.[52] Interitem correlations loaded from 0.74 to 0.90 and resulted in single-factor loading (**TABLE 5**). The Kaiser-Meyer-Olkin value was 0.96, indicating sampling sufficiency. Single-factor loading accounted for 72% of the common variance, with an eigenvalue of 10.116. There were no coefficients less than 0.80 and no cross-loadings; therefore, no items were eliminated. Cronbach's alpha values were high for all 14 items (α = .96), for the 6 items within the ADL portion (α = .90), and for the 8 items within the technique portion (α = .92).

Sensitivity

Data from the same 725 participants were also used in sensitivity analyses. Significant differences were found between healthy (85.75 ± 5.65) and injured (32.11 ± 24.54) dancers for DFOS total score ($t_{84.16}$ = 19.97, P<.001), ADL subscore (healthy, 37.80 ± 2.75; injured, 16.85 ± 10.91; $t_{84.39}$ = 17.53, P<.001), and technique subscore (healthy, 47.89 ± 3.52; injured, 15.50 ± 14.52; $t_{84.28}$ = 20.360, P<.001). There were also differences between groups for SF-36 PCS scores

TABLE 3 Demographics (Validity, Factor Analysis, and ROC)

	Male	Female	Total
Subjects, n (%)*	221 (30)	504 (70)	725 (100)
Professional dancers	145 (20)	154 (21)	299 (41)
Teachers, choreographers, directors	9 (1)	19 (3)	28 (4)
Preprofessional students	67 (9)	331 (46)	398 (55)
Age, y†	23.99 ± 6.65	21.86 ± 7.09	22.51 ± 7.02
Dance training, y†	11.41 ± 6.14	14.37 ± 7.18	13.48 ± 7.01
Dance training, n (%)*			
Ballet	27 (4)	49 (7)	76 (11)
Modern	33 (4)	64 (9)	97 (13)
Modern/ballet	161 (22)	391 (54)	552 (76)
Professional experience, y†	4.37 ± 6.11	2.45 ± 5.51	3.03 ± 5.76
Status, n (%)*			
Healthy	189 (26)	449 (62)	638 (88)
Injured	32 (4)	55 (8)	87 (12)
Ethnicity, n (%)*			
African American	123 (17)	183 (25)	306 (42)
Caucasian	60 (8)	238 (33)	298 (41)
Hispanic	18 (3)	31 (4)	49 (7)
Asian	13 (2)	37 (5)	50 (7)
Other	7 (1)	15 (2)	22 (3)

Abbreviation: ROC, receiver operating characteristic.
**All percentages are out of the total n.*
†Values are mean ± SD.

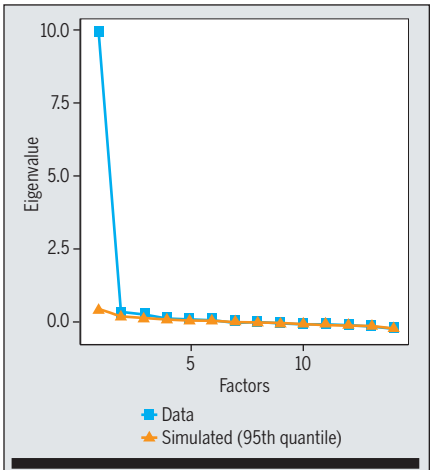


FIGURE 1. Scree plot. Cutoff reflects point of inflection for eigenvalues greater than 1 (1 eigenvalue = 10.116).

(healthy, 65.78 ± 10.04; injured, 25.32 ± 18.58; $t_{79.66}$ = 18.51, P<.001), but not for MCS scores.

The ROC curves resulted in AUC values of 0.94 (95% CI: 0.92, 0.96) for the DFOS total score, 0.91 (95% CI: 0.87, 0.94) for ADL subscore, and 0.94 (95% CI: 0.92, 0.96) for technique subscore, suggesting a high level of accuracy (**FIGURE 2**). Cutoffs were 77.5 for the DFOS total score (based on sensitivity and specificity values of 0.92 and 0.82, respectively), 35.5 for ADL subscore (sensitivity and specificity values of 0.85 and 0.82, respectively), and 43.5 for technique subscore (sensitivity and specificity values of 0.91 and 0.81, respectively).

Internal Responsiveness

Forty-seven dancers (24 female, 23 male; mean ± SD age, 27.60 ± 6.26 years) participated in the internal responsiveness part of this project after sustaining 60 injuries (all were initially screened as healthy dancers). The majority were professional dancers (87%), representing 3 companies. Most injuries were to the foot (32%), followed by the leg (17%) and knee (17%) (**FIGURE 3**).

Mauchly's test of sphericity was not significant; therefore, sphericity was not violated. There were significant differences across time for the DFOS total scores ($F_{1,59}$ = 97.295, P<.001), ADL subscores ($F_{1,59}$ = 102.579, P<.001), and technique subscores ($F_{1,59}$ = 69.815, P<.001)

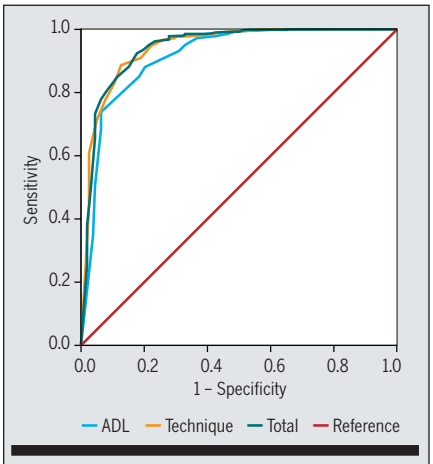


FIGURE 2. Graph of the receiver operating characteristic curve. Abbreviation: ADL, activities of daily living.

TABLE 4 Validity Analysis of DFOS Versus SF-36

Score	PCS	r*	MCS	r*
DFOS total	PCS†	0.79 (0.76, 0.82)	MCS	0.26 (0.19, 0.33)
ADL	PCS†	0.79 (0.76, 0.82)	MCS	0.31 (0.24, 0.38)
Overall activity	PF†	0.69 (0.65, 0.73)	VT†	0.14 (0.07, 0.22)
	RP†	0.54 (0.48, 0.59)	SF†	0.34 (0.27, 0.40)
Movement quality	PF†	0.67 (0.62, 0.71)	MH‡	0.10 (0.02, 0.17)
	RP†	0.52 (0.46, 0.57)	SF†	0.33 (0.26, 0.40)
Walking	PF†	0.66 (0.61, 0.70)	SF†	0.33 (0.26, 0.39)
Stairs	PF†	0.67 (0.62, 0.71)	SF†	0.32 (0.25, 0.39)
Stability	BP†	0.51 (0.45, 0.56)	SF†	0.34 (0.27, 0.40)
Pain	BP†	0.50 (0.45, 0.56)	VT†	0.21 (0.13, 0.28)
			RE†	0.22 (0.14, 0.31)
			SF†	0.33 (0.26, 0.42)
Technique	PCS†	0.77 (0.74, 0.80)	MCS	0.26 (0.19, 0.33)
Plié	PF†	0.65 (0.60, 0.69)	SF†	0.32 (0.25, 0.39)
Développé	PF†	0.55 (0.50, 0.60)	SF†	0.25 (0.18, 0.32)
Relevé balance	PF†	0.57 (0.51, 0.62)	SF†	0.28 (0.21, 0.35)
Rond de jambe	PF†	0.58 (0.53, 0.63)	SF†	0.23 (0.15, 0.30)
Kneeling	PF†	0.59 (0.53, 0.63)	SF†	0.29 (0.21, 0.36)
Turning	PF†	0.63 (0.58, 0.67)	SF†	0.29 (0.22, 0.36)
Jumping	PF†	0.63 (0.58, 0.68)	SF†	0.30 (0.23, 0.37)
Grand allegro	PF†	0.66 (0.62, 0.70)	SF†	0.31 (0.24, 0.38)

Abbreviations: ADL, activities of daily living; BP, bodily pain; DFOS, Dance Functional Outcome Survey; MCS, mental component summary; MH, mental health; PCS, physical component summary; PF, physical functioning; RE, role emotional; RP, role physical; SF, social functioning; SF-36, Medical Outcomes Study 36-Item Short-Form Health Survey; VT, vitality.
**Values in parentheses are 95% confidence interval.*
†PCS correlations and individual MCS subscores significant (P<.001).
‡P = .014.

TABLE 5 Factor Loadings

Item Content	Factor 1
Overall activity	0.90
Movement quality	0.87
Walking	0.87
Stairs	0.85
Stability/symptoms	0.86
Pain	0.77
Plié	0.86
Développé	0.74
Relevé balance	0.81
Rond de jambe	0.79
Kneeling	0.79
Turning	0.88
Jumping	0.88
Grand allegro	0.90

(**FIGURE 4**, **TABLE 6**). Pairwise comparisons were also significant for $T_{healthy}$, $T_{injured}$, and $T_{discharge}$ ($P<.001$) for DFOS total scores and ADL and technique subscores, but not for $T_{healthy}$ versus T_{3mo}. There were significant differences across time for SF-36 PCS ($F_{1,56} = 13.565$, $P<.001$) and MCS ($F_{1,56} = 21.229$, $P<.001$) scores. The PCS pairwise comparisons were significant for $T_{healthy}$, $T_{injured}$, and $T_{discharge}$ ($P<.001$), but not for $T_{healthy}$ versus T_{3mo}. No MCS pairwise comparisons were significant. For DFOS and SF-36 scores over 4 time points, SEM values were highest at $T_{injured}$ and lowest at $T_{healthy}$ and T_{3mo}. The MDC_{95}, SRM, and effect size displayed a pattern of decreasing from large values at $T_{injured}$ to small values at T_{3mo}. All SRM values were greater than 1.0, with the exception of those of the MCS, while all effect sizes exceeded 1.0 at $T_{injured}$ and $T_{discharge}$, with the exception of those of the MCS.

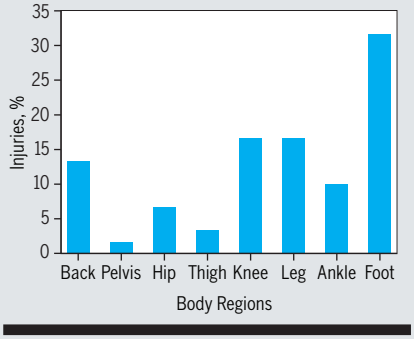


FIGURE 3. Injuries categorized by body region (percent).

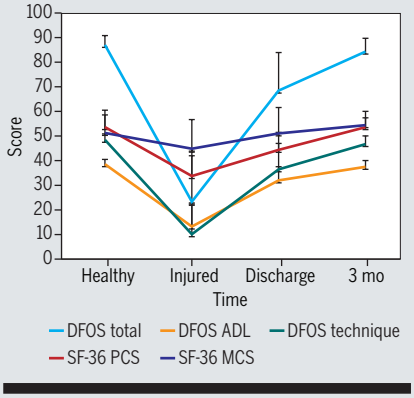


FIGURE 4. Mean ± SD DFOS and SF-36 scores across 4 time points. Abbreviations: ADL, activities of daily living; DFOS, Dance Functional Outcome Survey; MCS, mental component summary; PCS, physical component summary; SF-36, Medical Outcomes Study 36-Item Short-Form Health Survey.

Floor and Ceiling Effects

Within the injured group, DFOS scores were examined for floor and ceiling effects. One percent (1/87) of injured individuals had the minimum DFOS total score, and none had the maximum score. Therefore, no ceiling or floor effects were considered to be present.[42]

DISCUSSION

All study hypotheses were supported. The DFOS scores demonstrated high test-retest reliability, high internal consistency, strong construct validity compared with the SF-36 PCS, high AUC values, high internal responsiveness across time, and no floor or ceiling effects. Exploratory factor analysis determined single-factor loading, supporting unidimensionality of the scale. Each of these findings is discussed below.

Test-Retest Reliability

Reliability correlations were high for DFOS total scores and subscores for combined, healthy, and injured groups, supporting the hypothesis of an ICC≥0.70. Individual item correlations were high as well (ICC≥0.70), with the exception of ronde de jambe in the healthy group. Not all modern dancers take ballet classes regularly; therefore, this question may be difficult for them to answer. However, the item reflects a unique dance movement that reveals knee function. Additionally, Likert scale wording is not always sufficient to describe the subjective perception of a condition.[25]

Within the 1- to 2-week period of test-retest analysis, dancers may qualitatively view fluctuations as changes in their ability to perform these movements. Minor pains may cause these fluctuations, even with no time loss due to injury. Similarly, a dancer experiencing pain might stop or modify jumping; however, the correlation across test-retest for jumping was high.

We calculated the SEM for each group. The injured group reflected higher SEM for all DFOS scores. By their second physical therapy visit, function may improve following treatment and scores may reflect this. These values are comparable to those reported for injured groups using the Foot and Ankle Ability Measure.[19]

Construct Validity

Construct validity was examined by analyzing the strength of correlation of DFOS scores with those of the SF-36. Although the SF-36 is considered a generic outcome instrument, it is used to assess outcomes in musculoskeletal conditions, including back, hip, knee, ankle, and lower extremity conditions.[1,2,15,18,19,23,24,30,34] The SF-36 also includes a psychological component, whereas the DFOS does not, and therefore may provide additional useful information. As an indicator of convergent validity, strong correlations were found with the PCS and its domains (Pearson $r≥0.50$), as hypothesized. In contrast, supporting the hypothesis of divergent validity, we found weak correlations between the DFOS and the MCS and its domains (Pearson $r<0.50$). Within the PCS, the physical functioning domain correlated best to most DFOS ADL and technique items. Similar patterns of convergent and divergent relationships between musculoskeletal questionnaires and the SF-36 PCS and MCS have been reported for the Hip Outcome Score,[21] Hip Sports Activity Scale,[30] Cincinnati Knee Rating System,[21] Foot and Ankle Ability Measure,[19] and Lower Extremity Functional Scale,[1] further supporting the DFOS as a patient-specific outcome tool.

Factor Analysis and Internal Consistency

The EFA found that all 14 items loaded onto 1 factor, indicating that a single dimension was reflected in the DFOS. All factor items were 0.74 or greater, therefore none were considered for elimination. The high Cronbach's alpha (.96) indicated excellent internal consistency. The Cronbach's alpha is grounded in the assumption that each test item measures the same latent trait of the construct.[41] Factor analysis

with 1-factor loading confirms this assumption. Furthermore, high loadings of the factor items on the predicted factor, accounting for 72% of the variance, indicated convergent validity. Although EFA indicated single-factor loading, in order for the clinician to interpret the impact of injury on ADL versus technique and to make clinical decisions in rehabilitation progression, we stress that clinicians should review both ADL and technique subscores in addition to DFOS total scores.

Sensitivity

The ROC analysis is used in clinical epidemiology to quantify how accurately medical diagnostic tests can distinguish between 2 states—in this case, healthy and injured.[38,42] The ROC curve plots sensitivity against 1 minus specificity across the full range of values. The AUC assesses the overall diagnostic accuracy, or discrimination, by summarizing the entire location of the ROC curve rather than depending on a specific operating point. An AUC value of 0.94, found here, is considered excellent.[17]

Both the DFOS total scores and subscores and the SF-36 PCS scores demonstrated discrimination between healthy and injured dancers. No differences were found between groups in SF-36 MCS scores.

Internal Responsiveness

The majority of musculoskeletal injuries in dance are reported to be at the foot and ankle.[3,31,36,37] The injured group sustained the greatest number of injuries at the foot. Injuries ranged from herniated discs, iliopsoas strains, and hip labral tears proximally to fifth metatarsal fractures, plantar fascia ruptures, and plantar plate tears distally. These diagnoses represented typical musculoskeletal injuries seen in this population.

Internal responsiveness was examined using SRM, effect size, and other sensitivity-to-change measures across the spectrum of states from healthy, injured, recovery at discharge, to 3 months post discharge. All DFOS scores declined from $T_{healthy}$ to $T_{injured}$ and improved at $T_{discharge}$ and T_{3mo} post discharge. Scores at T_{3mo} were similar to dancers' healthy baseline scores ($T_{healthy}$).

TABLE 6 Responsiveness of the DFOS and SF-36

Time	DFOS ADL	DFOS Technique	DFOS Total	PCS	MCS
Healthy					
Score*	38.57 ± 1.94	48.47 ± 2.40	87.03 ± 3.80	53.64 ± 4.95	51.28 ± 9.24
SEM	0.41	0.50	0.80	1.04	1.94
MDC_{95}	1.13	1.40	2.21	2.88	5.37
Injured					
Score*	13.23 ± 9.58	10.10 ± 11.77	23.33 ± 20.12	33.75 ± 8.24	44.86 ± 11.85
SEM	2.01	2.47	4.22	1.73	2.48
MDC_{95}	5.57	6.84	11.7	4.79	6.89
Change (healthy – injured)	25.33	38.37	63.7	19.89	6.42
SRM	2.57	3.19	3.07	2.35	0.44
ES	13.04	15.96	16.76	4.02	0.69
Discharge					
Score*	31.98 ± 5.56	36.48 ± 10.52	68.47 ± 15.47	44.39 ± 7.10	51.75 ± 10.52
SEM	1.17	2.21	3.25	1.49	2.21
MDC_{95}	3.24	6.12	8.99	4.13	6.12
Change (injured – discharge)	18.75	26.38	45.13	19.87	6.22
SRM	1.93	1.93	2.12	2.26	0.50
ES	3.39	4.98	4.89	1.87	0.022
3 mo					
Score*	37.48 ± 2.57	46.75 ± 3.31	84.23 ± 5.41	53.62 ± 3.80	54.43 ± 5.60
SEM	0.54	0.69	1.13	0.80	1.18
MDC_{95}	1.49	1.93	3.15	2.21	3.26
Change (discharge – 3 mo)	5.5	10.27	15.85	9.23	3.36
SRM	1.17	1.02	1.13	1.39	0.37
ES	0.56	0.71	0.71	<0.01	−0.34

Abbreviations: ADL, activities of daily living; DFOS, Dance Functional Outcome Survey; ES, effect size; MCS, mental component summary; MDC_{95}, minimal detectable change at 95% confidence interval; PCS, physical component summary; SEM, standard error of measurement; SF-36, Medical Outcomes Study 36-Item Short-Form Health Survey; SRM, standardized response mean.
**Values are mean ± SD.*

The DFOS SEMs in healthy and injured dancers were comparable, and the MDC_{95} was similar to, or smaller than, those of other instruments, such as the Lower Extremity Functional Scale,[1,49,50] Foot and Ankle Ability Measure,[19] Lysholm knee score,[2] Hip Outcome Score,[22] and Oswestry Disability Index.[44] The SEM and MDC_{95} values increased from $T_{healthy}$ to $T_{injured}$, decreased at $T_{discharge}$, and returned close to baseline at T_{3mo}.

This pattern has been reported in other instruments,[50] demonstrating the importance of considering the time frame when selecting an MDC. The change from $T_{injured}$ at the start of treatment to $T_{discharge}$ is likely the most important to consider, and was approximately 3 points for the DFOS total scores. Again, this is comparable to that reported for the Lower Extremity Functional Scale[51] and lower than the Foot and Ankle Ability Measure.[19] The SRM values demonstrated high responsiveness or large effect, as hypothesized, throughout the 4 measurements. We also calculated a modified Cohen threshold, which still resulted in a large effect, substantiating this finding.[26] The SRM exceeded those reported for the Cincinnati Knee Rating System,[24] Lysholm knee score and Tegner activity scale,[2] Hip Sports Activity Scale,[30] Hip Outcome Score,[29] and International Hip Outcome Tool.[27]

Effect size was large for all DFOS and PCS scores for $T_{injured}$ and $T_{discharge}$ change scores. We anticipated a minimal effect size when comparing $T_{discharge}$ to T_{3mo} scores. In contrast, effect size was insubstantial for the MCS at all time points. In general, effect sizes for DFOS scores were 4 times those of the PCS, and larger than effect sizes reported for other musculoskeletal questionnaires.[2,29,30,45] This suggests that the DFOS may be most responsive to reflecting the functional status of dancers.

Floor and Ceiling Effects

Floor and ceiling effects exist when patients score at extremes, either minimum or maximum scores. Extreme scores permit no measurement of change, whether to reflect improvement or worsening function. A floor or ceiling effect of greater than 15% is considered unacceptable and indicates limited content quality.[42] The DFOS total score within the injured group did not reflect floor or ceiling effects, supporting the clinical usefulness of the DFOS in assessing change over time following injury.

Limitations

The DFOS has only been assessed for adult dancers aged 18 years or older. Future studies will examine adolescent dancers, as they comprise a large proportion of injured dancers seen in dance medicine clinics. Although paper questionnaires were checked for completeness, responses were occasionally missing. Currently, we use online questionnaires that do not permit the participant to progress to the next question without answering. This has eliminated missing data and data-entry error, as the data are downloaded directly into Excel.

CONCLUSION

The DFOS demonstrates acceptable psychometric performance as an outcome and screening measure for dancers. The DFOS is a useful tool to monitor health state and functional limitation following lower extremity or low back injury in adult ballet and modern dancers. ◉

KEY POINTS

FINDINGS: The Dance Functional Outcome Survey (DFOS) demonstrates acceptable psychometric performance as an outcome and screening measure for adult ballet and modern dancers. The DFOS demonstrated excellent reliability, sensitivity, internal responsiveness, and validity when compared to the Medical Outcomes Study 36-Item Short-Form Health Survey.

IMPLICATIONS: The DFOS is a useful tool to monitor both healthy state and functional limitation following lower extremity or low back injury in adult ballet and modern dancers. The DFOS focuses on dance-specific movements that are unaddressed in other sport or generic questionnaires and provides an important tool for investigating clinical efficacy.

CAUTION: The analyses were performed on professional and elite-level preprofessional dancers and may not apply to a broader population of dancers.

ACKNOWLEDGMENTS : *We thank Robert Turner, who assisted in developing the preliminary 16-item DFOS; the experts who participated in face validity assessment of the preliminary 16-item DFOS; the dancers who participated as subjects in this study; the organizations that allowed us access to these dancers; and the individuals who assisted in collecting data and conducting preliminary analyses, including Robert Turner, Sheyi Ojofeitimi, Dana Hash-Campbell, Alyson Filipa, Mark Paterno, and Kevin Brown.*

REFERENCES

1. Binkley JM, Stratford PW, Lott SA, Riddle DL. The Lower Extremity Functional Scale (LEFS): scale development, measurement properties, and clinical application. *Phys Ther*. 1999;79:371-383. https://doi.org/10.1093/ptj/79.4.371
2. Briggs KK, Kocher MS, Rodkey WG, Steadman JR. Reliability, validity, and responsiveness of the Lysholm knee score and Tegner activity scale for patients with meniscal injury of the knee. *J Bone Joint Surg Am*. 2006;88:698-705.
3. Bronner S, McBride C, Gill A. Musculoskeletal injuries in professional modern dancers: a prospective cohort study of 15 years. *J Sports Sci*. 2018;36:1880-1888. https://doi.org/10.1080/02640414.2018.1423860
4. Bronner S, Urbano IR. Dance Functional Outcome Survey: development and preliminary analyses. *Sports Med Int Open*. 2018;2:E191-E199. https://doi.org/10.1055/a-0729-3000
5. Busse JW, Bhandari M, Guyatt GH, et al. Use of both Short Musculoskeletal Function Assessment questionnaire and Short Form-36 among tibial-fracture patients was redundant. *J Clin Epidemiol*. 2009;62:1210-1217. https://doi.org/10.1016/j.jclinepi.2009.01.014
6. Domholdt E. *Physical Therapy Research: Principles and Applications*. 2nd ed. Philadelphia, PA: Saunders; 2000.
7. Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ. Evaluating the use of exploratory

factor analysis in psychological research. *Psychol Methods*. 1999;4:272-299. https://doi.org/10.1037/1082-989X.4.3.272

8. Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behav Res Methods*. 2009;41:1149-1160. https://doi.org/10.3758/BRM.41.4.1149
9. Fowler J, Jarvis P, Chevannes M. *Practical Statistics for Nursing and Health Care*. West Sussex, UK: John Wiley & Sons; 2002.
10. Garson GD. *Factor Analysis*. Asheboro, NC: Statistical Associates; 2013.
11. Hajian-Tilaki K. Sample size estimation in diagnostic test studies of biomedical informatics. *J Biomed Inform*. 2014;48:193-204. https://doi.org/10.1016/j.jbi.2014.02.013
12. Hincapié CA, Morton EJ, Cassidy JD. Musculoskeletal injuries and pain in dancers: a systematic review. *Arch Phys Med Rehabil*. 2008;89:1819-1829. https://doi.org/10.1016/j.apmr.2008.02.020
13. Hinkin TR. A brief tutorial on the development of measures for use in survey questionnaires. *Org Res Methods*. 1998;1:104-121. https://doi.org/10.1177/109442819800100106
14. Kane RL. *Understanding Health Care Outcomes Research*. Gaithersburg, MD: Aspen; 1997.
15. Ko S, Chae S. Correlations between the SF-36, the Oswestry-Disability Index and Rolland-Morris [sic] Disability Questionnaire in patients undergoing lumbar decompression according to types of spine origin pain. *Clin Spine Surg*. 2017;30:E804-E808. https://doi.org/10.1097/BSD.0000000000000438
16. Madeley NJ, Wing KJ, Topliss C, Penner MJ, Glazebrook MA, Younger AS. Responsiveness and validity of the SF-36, Ankle Osteoarthritis Scale, AOFAS Ankle Hindfoot Score, and Foot Function Index in end stage ankle arthritis. *Foot Ankle Int*. 2012;33:57-63. https://doi.org/10.3113/FAI.2012.0057
17. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol*. 2010;5:1315-1316. https://doi.org/10.1097/JTO.0b013e3181ec173d
18. Martin DP, Engelberg R, Agel J, Swiontkowski MF. Comparison of the Musculoskeletal Function Assessment questionnaire with the Short Form-36, the Western Ontario and McMaster Universities Osteoarthritis Index, and the Sickness Impact Profile health-status measures. *J Bone Joint Surg Am*. 1997;79:1323-1335.
19. Martin RL, Irrgang JJ, Burdett RG, Conti SF, Van Swearingen JM. Evidence of validity for the Foot and Ankle Ability Measure (FAAM). *Foot Ankle Int*. 2005;26:968-983. https://doi.org/10.1177/107110070502601113
20. Martin RL, Irrgang JJ, Lalonde KA, Conti S. Current concepts review: foot and ankle outcome instruments. *Foot Ankle Int*. 2006;27:383-390. https://doi.org/10.1177/107110070602700514
21. Martin RL, Kelly BT, Philippon MJ. Evidence of validity for the Hip Outcome Score. *Arthroscopy*. 2006;22:1304-1311. https://doi.org/10.1016/j.arthro.2006.07.027
22. Martin RL, Philippon MJ. Evidence of reliability and responsiveness for the Hip Outcome Score. *Arthroscopy*. 2008;24:676-682. https://doi.org/10.1016/j.arthro.2007.12.011
23. Martin RL, Philippon MJ. Evidence of validity for the Hip Outcome Score in hip arthroscopy. *Arthroscopy*. 2007;23:822-826. https://doi.org/10.1016/j.arthro.2007.02.004
24. Marx RG, Jones EC, Allen AA, et al. Reliability, validity, and responsiveness of four knee outcome scales for athletic patients. *J Bone Joint Surg Am*. 2001;83-A:1459-1469.
25. McCormack HM, Horne DJ, Sheather S. Clinical applications of visual analogue scales: a critical review. *Psychol Med*. 1988;18:1007-1019. https://doi.org/10.1017/S0033291700009934
26. Middel B, van Sonderen E. Statistical significant change versus relevant or important change in (quasi) experimental design: some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research. *Int J Integr Care*. 2002;2:e15.
27. Mohtadi NG, Griffin DR, Pedersen ME, et al. The development and validation of a self-administered quality-of-life outcome measure for young, active patients with symptomatic hip disease: the International Hip Outcome Tool (iHOT-33). *Arthroscopy*. 2012;28:595-605. https://doi.org/10.1016/j.arthro.2012.03.013
28. Munro BH. *Statistical Methods for Health Care Research*. 3rd ed. Philadelphia, PA: Lippincott; 1997.
29. Naal FD, Impellizzeri FM, von Eisenhart-Rothe R, Mannion AF, Leunig M. Reproducibility, validity, and responsiveness of the Hip Outcome Score in patients with end-stage hip osteoarthritis. *Arthritis Care Res (Hoboken)*. 2012;64:1770-1775. https://doi.org/10.1002/acr.21746
30. Naal FD, Miozzari HH, Kelly BT, Magennis EM, Leunig M, Noetzli HP. The Hip Sports Activity Scale (HSAS) for patients with femoroacetabular impingement. *Hip Int*. 2013;23:204-211. https://doi.org/10.5301/hipint.5000006
31. Nilsson C, Leanderson J, Wykman A, Strender LE. The injury panorama in a Swedish professional ballet company. *Knee Surg Sports Traumatol Arthrosc*. 2001;9:242-246. https://doi.org/10.1007/s001670100195
32. Optum. SF Health Surveys. Available at: https://campaign.optum.com/content/optum/en/optum-outcomes/what-we-do/health-surveys.html. Accessed June 18, 2018.
33. Osborne JW, Costello AB. Sample size and subject to item ratio in principal components analysis. *Pract Assess Res Eval*. 2004;9:11.
34. Park KB, Shin JS, Lee J, et al. Minimum clinically important difference and substantial clinical benefit in pain, functional, and quality of life scales in failed back surgery syndrome patients. *Spine (Phila Pa 1976)*. 2017;42:E474-E481. https://doi.org/10.1097/BRS.0000000000001950
35. Portney LG, Watkins MP. Reliability. In: *Foundations of Clinical Research: Applications to Practice*. Norwalk, CT: Appleton & Lange; 1993:53-67.
36. Ramkumar PN, Farber J, Arnouk J, Varner KE, McCulloch PC. Injuries in a professional ballet dance company: a 10-year retrospective study. *J Dance Med Sci*. 2016;20:30-37. https://doi.org/10.12678/1089-313X.20.1.30
37. Smith PJ, Gerrie BJ, Varner KE, McCulloch PC, Lintner DM, Harris JD. Incidence and prevalence of musculoskeletal injury in ballet: a systematic review. *Orthop J Sports Med*. 2015;3:2325967115592621. https://doi.org/10.1177/2325967115592621
38. Stratford PW, Binkley JM, Riddle DL. Health status measures: strategies and analytic methods for assessing change scores. *Phys Ther*. 1996;76:1109-1123. https://doi.org/10.1093/ptj/76.10.1109
39. Stratford PW, Spadoni G, Kennedy D, Westaway MD, Alcock GK. Seven points to consider when investigating a measure's ability to detect change. *Physiother Can*. 2002;54:16-24.
40. Tabachnick BG, Fidell LS. *Using Multivariate Statistics*. 5th ed. Boston, MA: Pearson/Allyn & Bacon; 2007.
41. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ*. 2011;2:53-55. https://doi.org/10.5116/ijme.4dfb.8dfd
42. 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
43. Urbano I, Bronner S. Reliability and validity of a dance outcomes instrument. Part I. Comparison of the 16 question DFOS to knee, foot and ankle outcomes tools. Performing Arts Medical Association 2016 International Symposium; July 7-10, 2016; New York, NY.
44. Vianin M. Psychometric properties and clinical usefulness of the Oswestry Disability Index. *J Chiropr Med*. 2008;7:161-163. https://doi.org/10.1016/j.jcm.2008.07.001
45. Vincent JI, MacDermid JC, King GJ, Grewal R. Validity and sensitivity to change of patient-reported pain and disability measures for elbow pathologies. *J Orthop Sports Phys Ther*. 2013;43:263-274. https://doi.org/10.2519/jospt.2013.4029
46. Walsh TL, Hanscom B, Lurie JD, Weinstein JN. Is a condition-specific instrument for patients with low back pain/leg symptoms really necessary? The responsiveness of the Oswestry Disability Index, MODEMS, and the SF-36. *Spine (Phila Pa 1976)*. 2003;28:607-615. https://doi.org/10.1097/01.BRS.0000050654.97387.DF
47. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*. 1992;30:473-483.
48. Ware JE, Snow K, Kosinski M, Gandek B. *SF-36 Health Survey: Manual and Interpretation Guide*.

Boston, MA: The Health Institute, New England Medical Center; 1993.
49. Watson CJ, Propps M, Ratner J, Zeigler DL, Horton P, Smith SS. Reliability and responsiveness of the Lower Extremity Functional Scale and the Anterior Knee Pain Scale in patients with anterior knee pain. *J Orthop Sports Phys Ther*. 2005;35:136-146. https://doi.org/10.2519/jospt.2005.35.3.136
50. Williams VJ, Piva SR, Irrgang JJ, Crossley C, Fitzgerald GK. Comparison of reliability and responsiveness of patient-reported clinical outcome measures in knee osteoarthritis rehabilitation. *J Orthop Sports Phys Ther*. 2012;42:716-723. https://doi.org/10.2519/jospt.2012.4038
51. Yeung TS, Wessel J, Stratford P, MacDermid J. Reliability, validity, and responsiveness of the Lower Extremity Functional Scale for inpatients of an orthopaedic rehabilitation ward. *J Orthop Sports Phys Ther*. 2009;39:468-477. https://doi.org/10.2519/jospt.2009.2971
52. Yong AG, Pearce S. A beginner's guide to factor analysis: focusing on exploratory factor analysis. *Tutor Quant Methods Psychol*. 2013;9:79-94.

APPENDIX A

DANCE FUNCTIONAL OUTCOME SURVEY

Name: _____ Today's date _____

Please answer every section, and mark in each section the one statement which most applies to you. We realize that two statements in any one section may relate to you, but just mark the one which most closely describes your level now. These questions are based only on what ***you*** can do at this time. Do not compare yourself to other dancers. If a section is not applicable, please skip it.

ACTIVITIES OF DAILY LIVING

1. Overall Activity Level

___ I have no limitations. I am able to do everything, including strenuous dancing and exercise.
___ I can dance, but at a lower level. I must guard myself and limit the amount of heavy dancing.
___ Light dancing is possible with occasional problems. I must avoid certain movements.
___ No dancing is possible. Daily activities are possible with occasional problems.
___ Daily activities cause moderate problems.
___ Daily activities cause severe problems.

2. Movement Quality

___ I feel confident that I can perform at the same level and quality as prior to my injury. I am able to articulate my limbs with 100% certainty or clarity.
___ I feel confident that I am almost at the same level and quality of performance as prior to my injury. I am able to articulate my limbs with 80% certainty or clarity.
___ I am improving but have a ways to go before I am back to the level and quality I was prior to my injury. I am able to articulate my limbs with 60% certainty or clarity.
___ I am improving but can only control my movement quality some of the time. I am able to articulate my limbs with 40% certainty or clarity.
___ I am improving but only beginning to focus on movement quality. I am able to articulate my limbs with 20% certainty or clarity.
___ I am improving but am working on basics and not able to focus on quality at this time.

3. Walking

___ Normal and unlimited, including hills.
___ Slight problems, relatively unlimited distances.
___ Mild problems, most surfaces, up to half a mile or 10 blocks.
___ Moderate problems, flat surfaces, no more than 1/4 mile or 5 blocks.
___ Severe problems, only 1/8 mile or 2-3 blocks.
___ Severe problems, need cane or crutches.

4. Stairs

___ Normal, unlimited up and down stairs.
___ Slight problems, need to be careful, particularly (circle one) up/down stairs.
___ Mild problems, have to go slowly, particularly (circle one) up/down stairs.
___ Moderate problems, only 10-15 steps possible, particularly (circle one) up/down stairs.
___ Severe problems, require a banister for support, particularly (circle one) up/down stairs.
___ Severe problems, only 0-5 steps with support, especially (circle one) up/down stairs.

5. Stability and Symptoms

___ I can do everything without symptoms of: giving out, locking, catching, grinding, or feeling weak.
___ I only have symptoms (of giving out, locking, catching, grinding, or feeling weak) with strenuous dancing or exercise.
___ I only have symptoms (of giving out, locking, catching, grinding, or feeling weak) with moderate dancing; it limits my vigorous activities.
___ Because I have symptoms (of giving out, locking, catching, grinding, or feeling weak) with light dancing, it limits almost all of my dancing. I occasionally have symptoms with walking or light household work.
___ I have symptoms frequently with simple activities such as walking. I must guard my injury at all times.
___ I have severe problems with symptoms (of giving out, locking, catching, grinding, or feeling weak). I can't do much of anything without having symptoms.

APPENDIX A

6. Pain

___ I have no pain.
___ I have occasional pain with strenuous dance or exercise. I don't think that things are entirely back to normal. Limitations are mild and tolerable, if I am careful.
___ There is occasional pain with moderate dancing or light exercise.
___ I have pain with any dancing, exercise, or light recreational activities. Occasional pain is brought on by daily activities.
___ Pain is a significant problem with activities as simple as walking. The pain is relieved by rest. I can't participate in dancing or exercise.
___ I have pain at all times, even during walking, standing, or light household work.

TECHNIQUE

7. Plié

___ Able to fully perform grand plié in all positions, including fourth and fifth.
___ Able to perform grand plié in first and second only.
___ Able to perform grand plié in second position only.
___ Cannot grand plié, but can demi-plié in all positions.
___ Have some difficulty with demi-plié.
___ Cannot demi-plié.

8. Développé

___ I am able to fully perform all parts of développé to the front or side without a problem.
___ I have slight problems performing développé to the front or side.
___ I have mild problems fully extending my leg in développé to the front or side, and must développé at a lower height.
___ I have moderate problems fully extending my leg in développé to the front or side and must mark it, but I can fully passé.
___ I do not développé to the front or side at all, but can do a full passé.
___ I cannot perform a full passé.

9. Relevé Balance *(if you do pointe work, indicate whether you can perform the indicated level on pointe)*

___ Able to attain and maintain my balance in relevé/pointe on the involved side without a problem.
___ Able to attain and maintain my balance in relevé/pointe on the involved side with only slight problems.
___ Able to attain and maintain my balance in relevé/pointe on the involved side with moderate difficulty.
___ Able to relevé but can't maintain the balance on the involved side without barre assistance.
___ Able to maintain my balance on flat foot, but cannot balance in relevé.
___ Cannot relevé or maintain my balance on the involved side on flat foot.

10. Rond de Jambe

___ Able to fully perform as much and as often as required, at 90°: grand rond de jambe en l'aire à la seconde (rotational movements of the leg in the air).
___ Able to perform at reduced speed: rond de jambe en l'aire à la seconde (rotational movements of the leg in the air).
___ Able to perform with mild problems such as reduced number and speed: rond de jambe en l'aire à la seconde (rotational movements of the leg in the air).
___ Able to perform with moderate problems such as reduced number, speed, and height (at 45°): rond de jambe en l'aire à la seconde (rotational movements of the leg in the air).
___ I mark or avoid all rond de jambe en l'aire type movements (rotational movements of the leg in the air).
___ I am unable to perform rond de jambe en l'aire à la seconde (rotational movements of the leg in the air) at all.

11. Kneeling/Floorwork

___ Able to fully perform floorwork or kneeling activities, without limitations.
___ Able to perform floorwork or kneeling activities, with mild limitations.
___ Able to perform floorwork or kneeling activities, with moderate limitations.
___ Able to perform floorwork or kneeling activities, with more moderate limitations: may require less repetitions or slight modification.
___ Severe problems, require support or modification.
___ Severe problems, unable to do.

APPENDIX A

12. Turning

___ Able to fully perform unlimited multiple turns of all kinds, on either leg (to the extent you were able prior to your injury).
___ Able to perform, but not quite fully, turns of all kinds, on either leg (to the extent you were able prior to your injury).
___ Able to perform, with slight problems, turns of most kinds, on either leg. I have to be careful about placement.
___ I have moderate problems with turning. I am able to do single inside and outside turns on the involved side.
___ Severe problems, no turning. I only do turn preparation and balance in relevé on the involved side.
___ Severe problems, unable to balance on the involved side.

13. Jumping

___ Able to fully perform everything: all grand and petit allegro (big and small jumping) combinations, including beats (to the extent you were able prior to your injury). Take-off power is normal and unlimited. Able to maintain my balance when landing from a jump or hop.
___ Able to perform, but not quite fully, grand and petit allegro (big and small jumping) combinations (to the extent you were able prior to your injury). Take-off power and ability to maintain my balance when landing is pretty good.
___ Able to perform with slight problems and some guarding: grand and petit allegro, and balance when landing from jumps or hops. I avoid most difficult jumps. Unable to do repeated jumps.
___ I have moderate problems with jumping. I am only doing simple jumps in the center.
___ Severe problems, affects all jumping in center floor. Can do simple jumps at the barre.
___ Severe problems, no jumping activity possible.

14. Grand Allegro/Across the Floor/Traveling/Running

___ Able to fully perform all traveling combinations (change of direction, pivots, quick stops and starts, or run) at full speed.
___ Able to perform, but not quite fully, all traveling combinations (change of direction, pivots, quick stops and starts, or run).
___ Able to perform, with slight problems, traveling combinations (change of direction, pivots, quick stops and starts, or run) at reduced speed.
___ I have moderate problems, and must move slowly and carefully in traveling combinations (change of direction, pivots, quick stops and starts, or run).
___ I have severe problems, and must avoid most traveling combinations. I stick to barre and adagio (or center floor).
___ I avoid all traveling combinations.

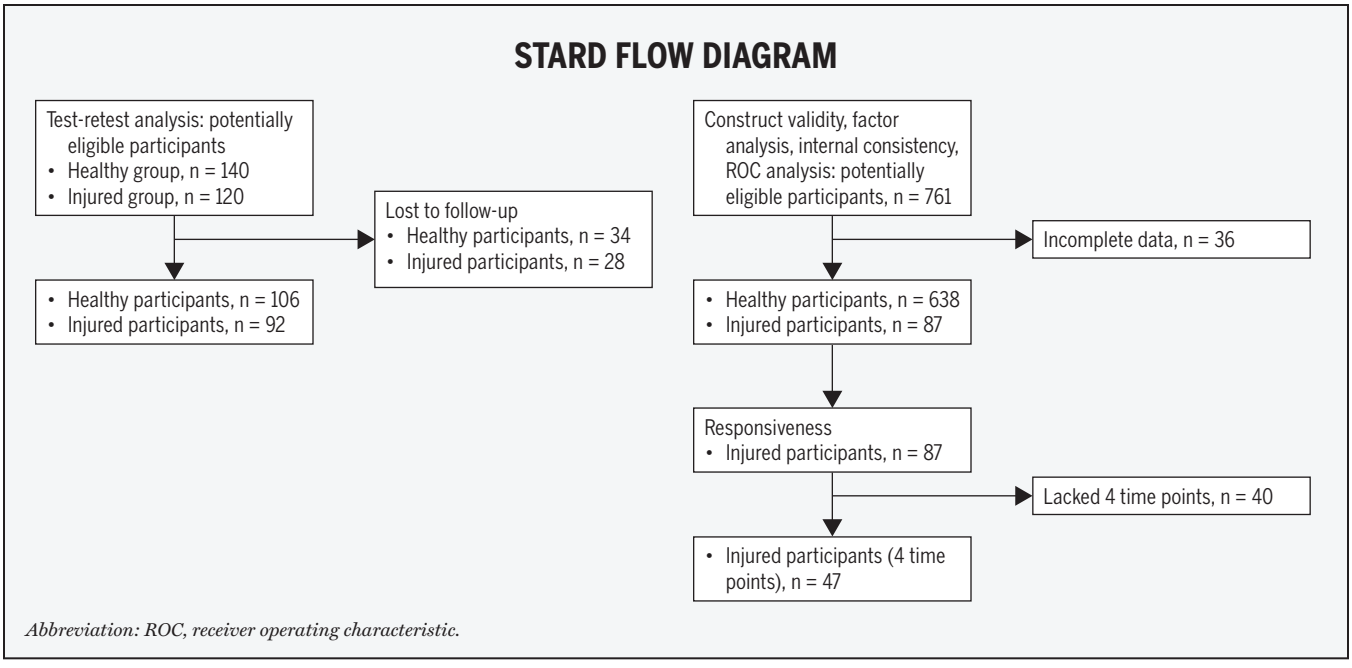
APPENDIX B

DANCE FUNCTIONAL OUTCOME SURVEY SCORING

A. Activities of daily living	
1. Overall Activity	10
2. Movement Quality	10
3. Walking	5
4. Stairs	5
5. Stability and Symptoms	5
6. Pain	5
Subtotal	**40**
B. Technique	
7. Plié	5
8. Développé	5
9. Relevé Balance	5
10. Rond de jambe	5
11. Kneeling/Floorwork	5
12. Turning	5
13. Jumping	10
14. Grand Allegro/Across the Floor/Traveling/Running	10
Subtotal	**50**
Total	**90**

Total score = % (eg, 90/90 = 100%).
If a question is unanswered, take a ratio of the answered questions.
Last question: As a healthy dancer, or compared to before my injury, if I had to give my dancing performance a grade from 0 to 100, with 0 being the worst and 100 being the best, I would give myself a _____.
Because this question is out of 100, it is a percentage.

APPENDIX C



Abbreviation: ROC, receiver operating characteristic.

ALESSANDRO CHIAROTTO, PT, PhD[1,2]

Patient-Reported Outcome Measures: Best Is the Enemy of Good (But What if Good Is Not Good Enough?)

J Orthop Sports Phys Ther 2019;49(2):39-42. doi:10.2519/jospt.2019.0602

Patient-reported outcome measures (PROMs) are increasingly important in research and clinical practice and to monitoring the efficiency of health care services.[2] The selection process of a PROM is fundamental to ensure that what matters to patients is captured in a valid, reliable, responsive, and feasible manner.[4] However, selecting a fit-for-purpose PROM is not always an easy task, as many clinimetric and sociological factors can play a role. In this Viewpoint, 2 different perspectives on PROM selection are presented and debated, and a few key suggestions are provided to improve PROM development and assessment. The measurement of physical functioning in patients with low back pain (LBP) is used as a recurring example.

Guidance on How to Select a PROM

Prinsen et al[16] proposed consensus-based guidance in selecting an outcome measure in the context of a core outcome measurement set (**FIGURE**). Such guidance is applicable to PROMs and can be used in other measurement contexts (eg, clinical trials, clinical practice). Three primary steps are involved: (1) making conceptual considerations, (2) identifying existing outcome measures, and (3) assessing the quality of the measures.

The first step is to determine which outcome domains to measure and to identify the population of interest,[16] which should be undertaken before searching for measures. A domain is an aspect of health that can be measured, and lists of potentially measurable domains can be found in existing health frameworks, such as the International Classification of Functioning, Disability and Health and the Wilson and Cleary[22] model of health-related quality of life. When choosing a domain, it is fundamental to establish whether it is relevant to be measured as an outcome. This could be determined based on personal clinical experience or, more appropriately, by looking at studies investigating relevant stakeholders' opinions on outcome domains.

For example, there is consensus among researchers, clinicians, and patients that physical functioning is the most important outcome domain to be measured in patients with nonspecific LBP.[8] It is also important to reflect whether the domain is unidimensional (ie, measures a single aspect) or multidimensional (ie, includes various aspects). A domain like physical functioning includes various subdomains (eg, mobility, dexterity, axial function, and the ability to carry out daily activities[6]); therefore, one may wish to measure it with a single PROM covering the whole domain or with multiple PROMs for each subdomain. The population of interest should be defined primarily in terms of sociodemographic (eg, age, sex) and disease (eg, diagnostic criteria, duration) characteristics.

Having established what to measure, the second step is to retrieve existing

[1]Department of Epidemiology and Biostatistics, Amsterdam Movement Sciences Research Institute, Amsterdam UMC (VUmc), Amsterdam, the Netherlands. [2]Department of General Practice, Erasmus MC, University Medical Center, Rotterdam, the Netherlands. The author certifies that he has 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 Alessandro Chiarotto, Department of Epidemiology and Biostatistics, Amsterdam Movement Sciences Research Institute, Amsterdam UMC (VUmc), de Boelelaan 1089a, Medical Faculty F-wing, 1081HV Amsterdam, the Netherlands. E-mail: a.chiarotto@vumc.nl

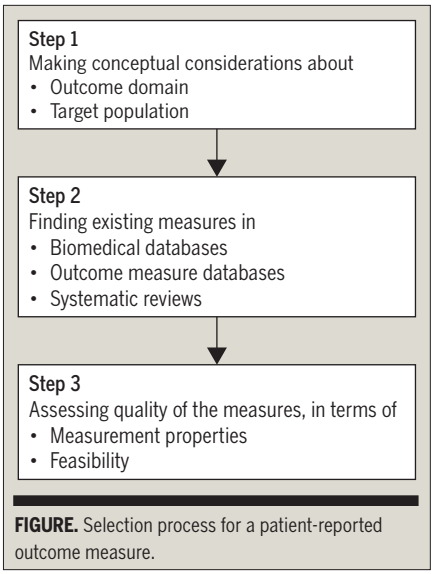


FIGURE. Selection process for a patient-reported outcome measure.

measures. Ideally, all existing PROMs for a given domain should be identified.[16] Biomedical databases and PROM-specific databases (eg, Rehabilitation Measures Database) can be systematically searched for this purpose. At this stage, it is important to establish whether the focus should be only on disease-specific measures or also on generic ones. Grotle et al[14] retrieved 36 disease-specific PROMs to measure functioning in LBP, a number that would substantially increase were generic (and newer) measures to be added.

The third step consists of assessing the quality of the potentially eligible PROMs in the target population, in terms of measurement properties and feasibility.[16] Considering the large variety of measurement properties and definitions, the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) initiative reached consensus on definitions for 9 measurement properties.[15] Among these properties, content validity should be the first measurement property assessed, as it allows one to make a clear link between the content of a domain and that of the PROM.[16] More specifically, a PROM with sufficient content validity would be expected to include aspects that are relevant, comprehensive, and comprehensible for the targeted (sub) domain(s) and population.[21]

The second measurement property to evaluate is structural validity, which determines whether the dimensionality of a PROM is aligned with that of the domain.[16] For example, the total score of a PROM is expected to be unidimensional if the domain it purports to measure is unidimensional as well. Assessment of other measurement properties should be performed only if an outcome measure meets these first 2 criteria.[16]

Feasibility of a PROM should be determined for practical characteristics like interpretability, ease of administration, length, completion time, ease of standardization, costs, copyright, and ease of score calculation.[16] Based on measurement properties and feasibility, one can select the best available PROM for a given (sub)domain and population.

The Best PROM Is the Enemy of a Good PROM

"Best is the enemy of good" is the English translation of an ancient Italian quote ("*Il meglio è nemico del bene*") included in a book on Italian proverbs from 1603. The French philosopher Voltaire subsequently popularized this quote, and it can also be found in a few books of Shakespeare. One of the most common interpretations of this quote is that trying to do something the best possible way can represent a barrier for doing it well, because trying to reach perfection is very often impossible and counterproductive. Applying this statement to the world of PROMs may mean that expecting a "perfect" PROM is simply unrealistic. Nevertheless, evidence on the best available PROM to use may come from head-to-head comparison studies, in which different measures are administered to the same patients.[19]

The Oswestry Disability Index and the Roland-Morris Disability Questionnaire (RMDQ) are the most frequently used measures to assess physical functioning in LBP.[13] A systematic review summarized the findings of head-to-head comparisons of these 2 PROMs and found that it was not possible to determine which one had better measurement properties.[9] Moreover, this and other reviews have identified that these and other PROMs for physical functioning in LBP are far from "perfect."[9,10,18] The Oswestry Disability Index and the RMDQ (published in the early 1980s) have been criticized by some to be outdated, with the suggestion that other physical functioning PROMs should be used or that new measures should be developed.[7] While it is true that these measures were not developed through qualitative methods involving patients and using advanced psychometric methods,[10] it is also true that there is no consistent evidence showing that other measures perform better than these 2.

Thus, the evidence generated so far on PROMs for physical functioning in LBP seems to support the statement that a best measure is not available. Recent reviews on PROMs for other domains and/or musculoskeletal conditions seem to suggest the same.[3,11] Consistent with these results, the Outcome Measures in Rheumatology (OMERACT) initiative suggests that "good" measures are sufficient for a core outcome measurement set, because the development of a core set of measures can continue despite the lack of "perfect" measures.[5]

When Is a Good PROM Good Enough?

The quote "Best is the enemy of good" is based on the simple assumption that the good is good enough. This means that, to be considered good, a PROM should not display major pitfalls, such as high-quality evidence for insufficient measurement properties, as suggested by the OMERACT initiative.[5] Moreover, it was proposed that a PROM should be selected only if it displays high-quality evidence of sufficient content validity and sufficient internal consistency.[16] In contrast with this guidance, recent high-quality evidence has demonstrated that the 24-item RMDQ has insufficient comprehensiveness of its content and insufficient unidimensionality of its total score.[10] These issues clearly indicate that content and structural validity of this measure are

suboptimal (ie, not good enough?), and, therefore, according to current recommendations on PROM selection,[5,16] that this long-standing measure should probably not be used anymore.

This perspective of discarding PROMs that are not good enough is consistent with the philosophy of the proponents of item response theory (IRT) Rasch modeling.[20] Concisely, this is a mathematical model that satisfies some basic measurement requirements (eg, the theory of conjoint measurement) that can be applied to a data set of PROM responses. According to these proponents, a PROM is "good" only when it fits the Rasch model, as this is the only model that may be able to provide interval-level measurement data, similar to those used in the exact sciences.[20] This is a valuable perspective; however, it has been criticized because other IRT models can also provide data that approximate interval-level measurement, and because Rasch modeling may pose a threat to a PROM's content validity.[12] Additionally, the results of Rasch analysis often lead to different versions of the same PROM,[10] which may be difficult to pool in cross-cultural comparisons.

Despite being aware of the aforementioned measurement limitations, a majority of researchers and clinicians recently decided to endorse the 24-item RMDQ as a core outcome measure for physical functioning in LBP clinical trials.[7] This decision clearly shows that the selection of a PROM goes beyond a confined and close look at measurement aspects. A possible explanation is that the measure was considered good enough in the absence of a "perfect" PROM. Another consideration is that an old, widely used measure should be discarded only when another (newer?) measure clearly proves to be better. Such comparative evidence against the 24-item RMDQ is not currently available.[9,10] Therefore, only new head-to-head comparison studies will be able to inform on whether this and other frequently used measures are, at least, good enough.

What Is Next in the Field of PROMs?

Because content validity is the first measurement property to consider in PROM selection,[16] and because it is understudied,[10,11,18] a major effort should be made to assess it. High-quality evidence on structural validity of several PROMs is also needed, and various psychometric methods are available (eg, [bi-]factor analysis, parametric and nonparametric IRT analysis) to assess this property. Both properties are crucial in PROM development[21] and for existing tools. Head-to-head comparison studies will help to establish if there is a tool that best measures the (sub)domain of interest, consistent with its content and dimensionality. For instance, such comparison would help to establish whether the 24-item RMDQ limitations are really worse than those of other physical functioning PROMs for LBP.

Another measurement property that is almost never assessed is cross-cultural validity, which evaluates whether the performance of a PROM's items is similar in different samples, cultures, and languages[15]; this property is essential to determine whether data on the same PROM can be pooled from different samples. High-quality comparative evidence on all the other measurement properties should also be generated.

Future head-to-head comparisons should include PROMs other than "the usual suspects," and possibly also generic measures. Generic PROMs can facilitate comparisons across health conditions, and, considering the substantial volume of comorbidities,[1] they may be better for those patients for whom it is difficult to attribute their complaints to a single disease. For example, the Patient-Reported Outcomes Measurement Information System (PROMIS) initiative has developed generic, patient-reported, domain-specific item banks and short forms that can be used with all health conditions.[6] Item banks are large sets of items developed to measure an ample range of "levels" of a domain. For physical functioning, this range may go from a person not able to get out of bed to an Olympic athlete. However, before widespread use, PROMIS tools should demonstrate measurement properties at least similar to those of the most frequently used disease-specific PROMs.

A broader use of computerized adaptive testing (CAT) may help to improve PROM assessment. Computerized adaptive testing is based on an item bank and IRT models.[17] Generally speaking, IRT analysis calibrates the difficulty and discrimination level of every item. Patients completing a CAT survey are administered items based on their responses to previous items, and a score of the patient's level on the domain is generated. Computerized adaptive tests are less time consuming than standard PROMs because fewer items are required to obtain the same measurement precision. A potential disadvantage of a CAT survey is that it requires the use of a computer and specific software, potentially reducing feasibility. The PROMIS item banks can be administered as CAT surveys as well.[6] To date, there have been few studies assessing whether a generic, domain-specific CAT instrument can outperform and replace standard PROMs.

Summary

Because a best or perfect PROM is realistically never available, PROMs that are good in terms of measurement properties and feasibility should be used. However, existing evidence suggests that several widely used PROMs for conditions like LBP and neck pain may not meet minimum good requirements. These potential limitations cannot be ignored, and a major effort by the scientific community is necessary to develop and to find PROMs that have at least sufficient content and structural validity for the target domain and population.

Key Points

- Selecting the best available PROM for a given domain may be a daunting task, as comparative evidence of competing PROMs is very often not available.

- Selecting a "good" PROM for a given domain means finding a tool with (at least) high-quality evidence of sufficient content validity and no high-quality evidence against any property.
- Some recent summaries of the evidence on widely used PROMs in the musculoskeletal field found a lack of content validity assessment and high-quality evidence against some properties.
- Content validity (ie, qualitative research with patients) and structural validity (ie, quantitative psychometric assessment) are the measurement properties that require priority in PROM development and assessment.
- Future high-quality head-to-head comparison studies may help to determine whether a best PROM is available for a given domain, and whether limitations of some PROMs currently in use should prevent their future recommendation. ◉

ACKNOWLEDGMENTS: *The author acknowledges Prof Raymond Ostelo, Dr Caroline Terwee, and Prof Maarten Boers, as the content of this manuscript is partly inspired by face-to-face discussions with them. Additionally, the author acknowledges Prof Martin Underwood, who taught him the expression "Best is the enemy of good."*

REFERENCES

1. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *Lancet*. 2012;380:37-43. https://doi.org/10.1016/S0140-6736(12)60240-2
2. Black N. Patient reported outcome measures could help transform healthcare. *BMJ*. 2013;346:f167. https://doi.org/10.1136/bmj.f167
3. Bobos P, MacDermid JC, Walton DM, Gross A, Santaguida PL. Patient-reported outcome measures used for neck disorders: an overview of systematic reviews. *J Orthop Sports Phys Ther*. 2018;48:775-788. https://doi.org/10.2519/jospt.2018.8131
4. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for Outcome Measures in Rheumatology. *J Rheumatol*. 1998;25:198-199.
5. Boers M, Kirwan JR, Tugwell P, et al. *The OMERACT Handbook*. Outcome Measures in Rheumatology; 2018.
6. Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. *J Clin Epidemiol*. 2010;63:1179-1194. https://doi.org/10.1016/j.jclinepi.2010.04.011
7. Chiarotto A, Boers M, Deyo RA, et al. Core outcome measurement instruments for clinical trials in nonspecific low back pain. *Pain*. 2018;159:481-495. https://doi.org/10.1097/j.pain.0000000000001117
8. Chiarotto A, Deyo RA, Terwee CB, et al. Core outcome domains for clinical trials in non-specific low back pain. *Eur Spine J*. 2015;24:1127-1142. https://doi.org/10.1007/s00586-015-3892-3
9. Chiarotto A, Maxwell LJ, Terwee CB, Wells GA, Tugwell P, Ostelo RW. Roland-Morris Disability Questionnaire and Oswestry Disability Index: which has better measurement properties for measuring physical functioning in nonspecific low back pain? Systematic review and meta-analysis. *Phys Ther*. 2016;96:1620-1637. https://doi.org/10.2522/ptj.20150420
10. Chiarotto A, Ostelo RW, Boers M, Terwee CB. A systematic review highlights the need to investigate the content validity of patient-reported outcome measures for physical functioning in patients with low back pain. *J Clin Epidemiol*. 2018;95:73-93. https://doi.org/10.1016/j.jclinepi.2017.11.005
11. Chiarotto A, Terwee CB, Kamper SJ, Boers M, Ostelo RW. Evidence on the measurement properties of health-related quality of life instruments is largely missing in patients with low back pain: a systematic review. *J Clin Epidemiol*. 2018;102:23-37. https://doi.org/10.1016/j.jclinepi.2018.05.006
12. Cook KF, Monahan PO, McHorney CA. Delicate balance between theory and practice: health status assessment and item response theory. *Med Care*. 2003;41:571-574. https://doi.org/10.1097/01.MLR.0000064780.30399.A4
13. Froud R, Patel S, Rajendran D, et al. A systematic review of outcome measures use, analytical approaches, reporting methods, and publication volume by year in low back pain trials published between 1980 and 2012: *respice, adspice, et prospice*. *PLoS One*. 2016;11:e0164573. https://doi.org/10.1371/journal.pone.0164573
14. Grotle M, Brox JI, Vøllestad NK. Functional status and disability questionnaires: what do they assess? A systematic review of back-specific outcome questionnaires. *Spine (Phila Pa 1976)*. 2005;30:130-140. https://doi.org/10.1097/01.brs.0000149184.16509.73
15. 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
16. Prinsen CA, Vohra S, Rose MR, et al. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" – a practical guideline. *Trials*. 2016;17:449. https://doi.org/10.1186/s13063-016-1555-2
17. Revicki DA, Cella DF. Health status assessment for the twenty-first century: item response theory, item banking and computer adaptive testing. *Qual Life Res*. 1997;6:595-600. https://doi.org/10.1023/A:1018420418455
18. Speksnijder CM, Koppenaal T, Knottnerus JA, Spigt M, Staal JB, Terwee CB. Measurement properties of the Quebec Back Pain Disability Scale in patients with nonspecific low back pain: systematic review. *Phys Ther*. 2016;96:1816-1831. https://doi.org/10.2522/ptj.20140478
19. Stratford PW, Binkley J, Solomon P, Gill C, Finch E. Assessing change over time in patients with low back pain. *Phys Ther*. 1994;74:528-533. https://doi.org/10.1093/ptj/74.6.528
20. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Rheum*. 2007;57:1358-1362. https://doi.org/10.1002/art.23108
21. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res*. 2018;27:1159-1170. https://doi.org/10.1007/s11136-018-1829-0
22. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *JAMA*. 1995;273:59-65. https://doi.org/10.1001/jama.1995.03520250075037

JESSE C. CHRISTENSEN, DPT, PhD, SCS[1,2,3] • **K. BO FOREMAN,** PT, PhD[1,4] • **PAUL C. LASTAYO,** PT, PhD[1,4]
ROBIN L. MARCUS, PT, PhD[4] • **CHRISTOPHER E. PELT,** MD[1] • **RYAN L. MIZNER,** PT, PhD[5]

Comparison of 2 Forms of Kinetic Biofeedback on the Immediate Correction of Knee Extensor Moment Asymmetry Following Total Knee Arthroplasty During Decline Walking

BACKGROUND: Individuals with total knee arthroplasty (TKA) display interlimb knee extensor moment (KEM) asymmetry during level walking that is exacerbated as task demands are increased. Studies using biofeedback to correct interlimb KEM asymmetry following TKA have reported mixed results.

OBJECTIVE: To compare the immediate effect of 2 forms of real-time kinetic biofeedback—vertical ground reaction force (vGRF) or KEM—on improving interlimb peak KEM symmetry during the weight-acceptance phase of decline walking in persons who have undergone TKA.

METHODS: In this cross-sectional, controlled laboratory study, 30 participants (17 men; mean ± SD age, 61.9 ± 8.5 years; body mass index, 28.4 ± 3.7 kg/m^2) were allocated to either a vGRF or KEM real-time biofeedback group. Peak KEM interlimb asymmetry was obtained during both nonbiofeedback and biofeedback decline walking trials 3 months following TKA.

RESULTS: Significant interlimb asymmetry in peak KEM was observed in both groups during the nonbiofeedback condition (KEM, P = .02; vGRF, P<.01). The KEM biofeedback group demonstrated an immediate improvement in peak KEM asymmetry (P = .42). No change in peak KEM asymmetry was observed in the vGRF biofeedback group (P = .01).

CONCLUSION: Knee extensor moment biofeedback has an immediate effect on improving peak KEM asymmetry 3 months post TKA. *J Orthop Sports Phys Ther 2019;49(2):105-111. Epub 20 Aug 2018. doi:10.2519/jospt.2019.7800*

KEY WORDS: *biomechanics, GRF, KEM, knee, TKA, walking*

Individuals recovering from unilateral total knee arthroplasty (TKA) exhibit interlimb asymmetries characterized by higher dynamic knee stiffness, decreased limb loading, and reduced knee extensor moments (KEMs).[12,29,41,43] These asymmetries persist despite improvements in perceived functional performance and knee pain,[3] and are evident for years following a successful postoperative recovery.[20,26,43] Chronic knee interlimb asymmetry has been shown to lead to muscle disuse in the surgical limb and abnormal overloading of the nonsurgical limb.[1,22,34] Interlimb asymmetries also have been shown to predispose patients to pain in other joints, functional limitations, and arthritic changes over time.[1,32]

Knee interlimb asymmetry in persons who have undergone TKA has been reported during sit-to-stand,[6,10,44] level walking,[1,6,44] and stair climbing.[17,19] However, the magnitude of knee interlimb asymmetry varies, as more physically demanding activities (eg, sit-to-stand) result in greater compensatory strategies compared to lower-demand activities (eg, level walking).[29] Several clinical factors (eg, pain, swelling, muscle weakness, etc) limit the restoration of functional mobility,[2,28,36] and addressing these modifiable risk factors is important to normalizing knee joint mechanics.[11,24,29]

Eccentrically biased mobility tasks have been shown to be the most physically demanding and commonly reported functional limitation following TKA.[11] However, the degree of knee interlimb asymmetry during more physically demanding tasks, such as decline walking, has been understudied in this population. Decline walking is a commonly performed gait task that requires greater KEMs compared to other activities of daily

[1]Department of Orthopaedics, University of Utah, Salt Lake City, UT. [2]Department of Physical Medicine and Rehabilitation, University of Colorado, Aurora, CO. [3]Eastern Colorado Geriatric Research Education and Clinical Center, Aurora, CO. [4]Department of Physical Therapy and Athletic Training, University of Utah, Salt Lake City, UT. [5]School of Physical Therapy and Rehabilitation Science, University of Montana, Missoula, MT. The University of Utah Institutional Review Board (Salt Lake City, UT) approved this study. This project was partially funded by the L.S. Peery Discovery Program in Musculoskeletal Restoration. 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 Jesse C. Christensen, University of Colorado School of Medicine, Department of Physical Medicine and Rehabilitation, 13001 E. 17th Pl, Aurora, CO, 80045. E-mail: jesse.christensen@ucdenver.edu

living.[14,15,30] As such, an understanding of the degree of knee interlimb asymmetry during this task and how biofeedback can assist in correcting this asymmetry during the early recovery period following TKA is needed.

Real-time biofeedback using vertical ground reaction force (vGRF) has been studied as a means of correcting gait asymmetry following TKA.[6,18,44] Studies have reported that vGRF biofeedback is effective in correcting knee interlimb asymmetry during level walking. For activities requiring a larger knee extensor demand, such as sit-to-stand, however, results are conflicting, as the use of vGRF biofeedback to correct KEM asymmetry has been reported to be successful[44] and not successful.[6] It is possible that vGRF as a biofeedback variable is suboptimal during high-demand activities, as it is unable to provide knee-specific kinetic information.

An alternative biofeedback option is use of the KEM, as this variable can provide knee-specific kinetic information to the patient. This form of biofeedback may be more effective in correcting knee moment asymmetry compared to vGRF biofeedback, which is not knee specific. The purpose of the current study was to compare the immediate effects of 2 forms of kinetic biofeedback (vGRF versus KEM) on improving interlimb peak knee moment symmetry during decline walking following unilateral TKA. The authors hypothesized that KEM biofeedback would be superior to vGRF biofeedback in reducing peak knee moment asymmetry 3 months following surgery.

METHODS

Participants

Patients who had previously undergone primary unilateral TKA (n = 30) participated in this study (17 men; mean ± SD age, 61.9 ± 8.5 years; body mass index [BMI], 28.4 ± 3.7 kg/m^2). All participants underwent a primary unilateral TKA and met the following inclusion criteria: between 45 and 75 years of age; BMI of less than 40 kg/m^2; University of California, Los Angeles (UCLA) activity scale score of greater than or equal to 3; nonsurgical knee pain of less than or equal to 4/10 on a visual analog scale; no comorbidities that would influence balance or walking ability; no current diagnosis or treatment of neurological conditions; no prior knee joint replacement procedure to either limb; and no plans of undergoing a TKA on the contralateral limb within 12 months after the initial procedure. All surgical procedures were performed by 1 of 3 orthopaedic surgeons who were recruited from a single academic medical center (Salt Lake City, UT). The study was approved by the University of Utah Institutional Review Board, and all subjects consented to participation prior to enrollment.

Each participant was assigned to 1 of 2 biofeedback groups (**TABLE 1**). Fifteen participants underwent a single session of gait symmetry training using vGRF biofeedback, and were compared to an age- and BMI-matched TKA group of 15 participants who underwent a single session of gait symmetry training using KEM biofeedback. Nonrandomized matched assignment was conducted, with the first 15 participants being enrolled in the vGRF group. Fifteen matched participants were then enrolled in the KEM group. The matching criterion was defined as less than a 10% participant difference based on age and BMI.

An a priori sample-size calculation conducted based on an effect size of 1.1[6] indicated a minimum of 15 participants to detect between-group differences (80% power with a 2-sided alpha of .05). For within-group differences, 15 participants were required to achieve 80% power and detect an effect size of 0.78.

Procedures

All testing was completed at the Motion Capture Core Facility at the University of Utah Department of Physical Therapy and Athletic Training. Motion analysis was performed using a 10-camera motion-analysis system sampling at 200 Hz (Vicon; Oxford Metrics, Oxford, UK). Kinetic data were obtained using a dual-belt instrumented treadmill (Bertec Corporation, Columbus, OH,) sampling at 1000 Hz. Kinetic and kinematic data were recorded and synchronized using Nexus 2.1.1 software (Oxford Metrics).

Each participant was fitted with compressive clothing and a safety harness and instrumented with 50 retroreflective markers (14 mm), allowing for the tracking of 8 body segments. Prior to data collection, the motion-analysis system was calibrated, and a standing calibration trial was obtained to determine joint centers and to create a segment coordinate system. The modified Plug-in Gait marker set (Oxford Metrics) defined 1 combined head, arms, and trunk segment, 1 pelvis segment, 2 thigh segments, 2 shank segments, and 2 foot segments. Marker locations were used for attributing coordinate systems for each segment and were positioned on the seventh cervical spinous process, the manubrium

TABLE 1 Descriptive Characteristics of Patients by Group*

Variable	vGRF (n = 15)	KEM (n = 15)	*P* Value
Age, y	61.6 ± 8.9	62.1 ± 8.2	.90
Sex (male), n (%)	8 (53.3)	9 (60.0)	.14
Mass, kg	87.5 ± 19.1	81.5 ± 12.9	.46
Height, m	1.73 ± 0.1	1.71 ± 0.1	.67
BMI (kg/m^2)	29.2 ± 3.6	27.9 ± 3.8	.54

Abbreviations: BMI, body mass index; KEM, knee extensor moment; vGRF, vertical ground reaction force.
**Values are mean ± SD unless otherwise indicated.*

of the sternum, the inferior body of the sternum, bilaterally on the anterior and posterior superior iliac spines, the right spine of the scapula, iliac crests, greater trochanters, acromions, medial and lateral epicondyles of the femurs, medial and lateral malleoli, first and fifth heads of the metatarsals, dorsum of the feet, and calcaneal tuberosities. One rigid cluster with 4 noncollinear markers was placed at the base of the lumbar spine, and 2 nonrigid clusters with 4 noncollinear markers were placed at the lateral side of each thigh and shank.[4]

Gait Symmetry Training

Participants underwent a single session of biofeedback training, based on their group assignment, 3.3 ± 0.5 months following TKA. All participants walked shod on a 10° decline slope at a constrained velocity of 0.8 m/s. The constrained velocity was used to control the task demands across conditions (nonbiofeedback and biofeedback). A decline angle of 10° has been shown to require greater knee joint demand than level walking and is a common slope encountered within the community.[14,35] Participant gait analysis was conducted under 2 conditions: (1) nonbiofeedback trials, in which participants were instructed to walk "as normal as possible, as if walking downhill," without exposure to any form of visual biofeedback, and (2) biofeedback trials, in which participants were instructed to use the visual kinetic biofeedback provided to assist in correcting knee interlimb asymmetry.

Technical instructions were provided to all participants prior to biofeedback data collection using the following script: "For the next downhill walking series, you will see 2 signals on the monitor. One will represent the signal of your surgical limb and the other your nonsurgical limb. Your goal is to attempt to make the 2 signals you see on the screen as equal and symmetrical as possible." Depending on the assigned biofeedback group, each participant was educated on the kinetic signal variable he or she would be receiving, so that he or she could understand the context of the visual representation.

Participants assigned to the vGRF group received biofeedback via real-time tracing of both lower-limb signals through commercial software (Oxford Metrics). Participants assigned to the KEM group received biofeedback via real-time kinetic computation of the sagittal plane KEM signal through Visual3D software (C-Motion, Inc, Germantown, MD) (**FIGURE 1**). Visual biofeedback for both groups was displayed on a 101.6-cm monitor positioned approximately 1.0 m anterior to the treadmill (**FIGURE 2**).

Initially, a 3- to 5-minute warm-up period was provided at the constrained speed and slope angle to allow the participants to become comfortable walking on the instrumented treadmill. Once participants confirmed they felt comfortable with the task, they were asked to walk at the constrained speed as the nonbiofeedback trials were collected (approximately

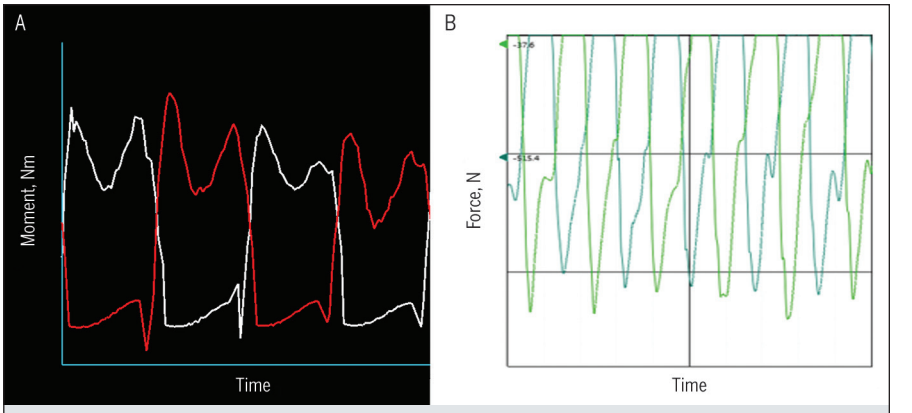


FIGURE 1. Visual illustration of each biofeedback variable. (A) Knee extensor moment signal (white signal, surgical knee; red signal, nonsurgical knee). (B) Vertical ground reaction force (blue signal, surgical limb; green signal, nonsurgical limb).

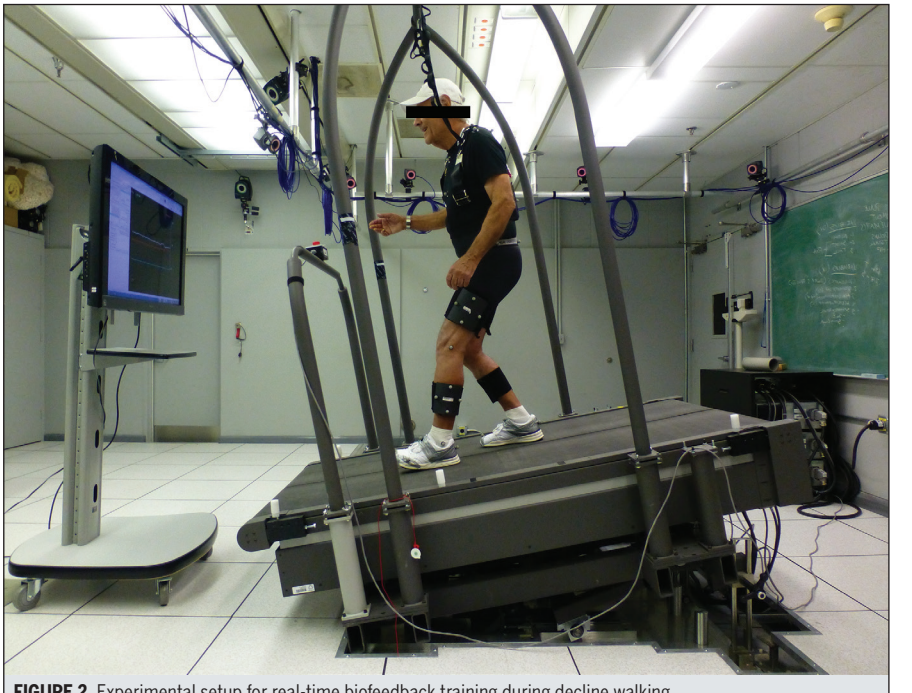


FIGURE 2. Experimental setup for real-time biofeedback training during decline walking.

3-5 minutes of data collection). Participants were provided a 5- to 10-minute rest period prior to beginning the biofeedback trials. As the participants began the biofeedback training, they were instructed to maintain symmetry between the surgical and nonsurgical limbs by using the kinetic biofeedback provided on the monitor.

A trial, defined as 10 successful steps, was considered acceptable if all markers were visible and the participant's foot landed successfully on the instrumented treadmill force platforms without any disturbance to gait. Trials in which participants lost their balance, used their upper limbs for support on the surrounding bars, or stepped onto the adjacent force platform were excluded. For each walking condition, 10 successful steps were averaged and used for statistical analysis.

Data Processing

Postprocessing and extraction of biomechanical variables were acquired using Visual3D Version 6.00.27 (C-Motion, Inc). The raw marker and force data were filtered using a fourth-order, low-pass Butterworth digital filter at a cutoff frequency of 6 Hz (trajectory) and 25 Hz (analog). The cutoff frequency was determined by residual analysis and visual inspection.[39]

Data Analysis

Knee extensor moments were computed using inverse dynamics equations in Visual3D (C-Motion, Inc) and normalized to body mass (kilograms). The peak KEM was identified within the weight-acceptance phase of the gait cycle. This phase of gait was evaluated because it has been shown to be more mechanically demanding for the knee and to be appropriate for identifying knee joint asymmetry.[14,37,40] For both the nonbiofeedback and biofeedback trials, the difference in the peak sagittal plane KEM of each limb (surgical versus nonsurgical) was calculated. A value equal to zero signified perfect symmetry, while values greater or less than zero indicated asymmetry.

Statistical Analysis

Participant demographics were evaluated using descriptive statistics. To compare between-limb differences in the peak KEM within each group prior to exposure to biofeedback, paired-samples *t* tests were used. To compare between-limb differences in the peak KEM within each group after exposure to biofeedback, an analysis of covariance (ANCOVA) was conducted, controlling for baseline between-limb differences. To compare between-group differences in correcting interlimb peak KEM differences, a separate ANCOVA was conducted, controlling for the baseline asymmetry. The ANCOVA model was initially fit, and marginal estimation was used to obtain adjusted means and mean differences.[25] We did not examine normality or homogeneity of variance, as *t* tests and ANCOVAs are known to be robust to those assumptions, even with sample sizes as small as 4.[9,16]

Effect sizes were computed as partial correlations (Cohen's f^2)[33]; for the adjusted model, Cohen's f^2 equal to or greater than 0.02 represents a small effect, a value equal to or greater than 0.15 represents a medium effect, and a value equal to or greater than 0.35 represents a strong effect.[7,33] Data were analyzed using Stata Version 14.1 (StataCorp LLC, College Station, TX).

RESULTS

Descriptive statistics revealed that the groups were comparable with respect to age, sex, and BMI scores (**TABLE 1**). The within-group analysis revealed significant interlimb differences in peak KEM in both groups during the nonbiofeedback baseline condition (**TABLE 2**). Following the biofeedback trials, significant interlimb differences in the peak KEM ($P = .01$; Cohen $f^2 = 0.24$) remained in the vGRF biofeedback group (**TABLE 2**). In contrast, no significant interlimb differences in the peak KEM were observed in the KEM biofeedback group following the biofeedback condition ($P = .42$) (**TABLE 2**). The between-group analysis revealed that the between-limb difference in peak KEM post biofeedback was larger in the vGRF group compared with the KEM group ($P = .01$; Cohen $f^2 = 0.24$) (**TABLE 3**).

DISCUSSION

The purpose of this study was to compare the immediate effects of 2 forms of kinetic biofeedback (vGRF or KEM) on improving interlimb KEM symmetry during decline walking following TKA. The principal findings of this study were that (1) significant interlimb

TABLE 2 Unadjusted and Adjusted Within-Group Comparison of the Peak KEM During Both the Nonbiofeedback and Biofeedback Conditions for Each Group*

	vGRF Group (n = 15)			KEM Group (n = 15)		
Condition/Variable	Surgical Knee	Nonsurgical Knee	*P* Value	Surgical Knee	Nonsurgical Knee	*P* Value
Nonbiofeedback						
Peak KEM, Nm/kg†	0.61 ± 0.06	0.79 ± 0.06	<.01‡	0.52 ± 0.05	0.72 ± 0.06	.02‡
Biofeedback						
Peak KEM, Nm/kg§	0.70 ± 0.07	0.87 ± 0.07	.01‡	0.81 ± 0.09	0.89 ± 0.09	.42

Abbreviations: KEM, knee extensor moment; vGRF, vertical ground reaction force.
**Values are mean ± standard error between the surgical knee and nonsurgical knee.*
†Unadjusted values based on the t test model.
‡Significant between-limb difference (P<.05).
§Adjusted values based on the analysis of covariance model.

asymmetry in peak KEM was present in both biofeedback groups prior to exposure to either form of gait symmetry training, and (2) KEM biofeedback resulted in improvement in peak KEM symmetry, while vGRF biofeedback did not improve peak KEM symmetry, during high-demand decline walking.

Correcting knee interlimb asymmetry 3 months after unilateral TKA is important to address in postoperative rehabilitation, as chronic asymmetry can accelerate arthritic changes, lead to muscle weakness, and lower functional performance.[6,22,29,34,44] Patients with TKA demonstrate reduced speed, single-leg stance time, knee flexion excursion, and weight-acceptance loading on the surgical limb compared to the dominant limb of healthy adults during decline walking at time points greater than 12 months following surgery.[30,37] Moreover, knee interlimb asymmetry does not appear to resolve over time, as residual deficits continue to persist even during less physically demanding tasks such as level walking.[1,8,20,21,43]

Gait symmetry training, as it pertains to interlimb KEM symmetry, could be an important addition to post-TKA rehabilitation protocols. However, studies using gait symmetry training with visual, auditory, and tactile biofeedback have reported mixed results in correcting knee interlimb asymmetry and improving functional performance.[6,18,44] One potential explanation for these inconsistent findings is that vGRF or equivalent variables of biofeedback do not provide knee-specific kinetic information that could more precisely assist in correcting asymmetry, especially during tasks that require larger KEM demands (ie, decline walking, descending stairs, stand-to-sit).

The findings of the present study are clinically relevant in that simply providing vGRF biofeedback does not appear to result in immediate attenuation of knee interlimb asymmetry, particularly during tasks that require higher knee demands.[6,44] Knee extensor moment biofeedback has not been studied as a component of motor training, possibly due to the complexity of computation of the real-time moment signal. To date, achieving accurate KEM biofeedback requires a sophisticated gait laboratory, robust marker-set model, synced communication between software, and patient comprehension, which can be challenging for most rehabilitation settings. However, as many as 80% of patients with TKA exhibit sagittal plane knee moment asymmetry following surgery compared to healthy peers.[20]

Additionally, knee interlimb asymmetry has been linked to quadriceps weakness, degradation of the contralateral limb, and poorer functional performance.[5,27,29,41,42] In the current study, KEM biofeedback training resulted in an immediate change in knee moment asymmetry during a higher-demand task 3 months following surgery, compared to vGRF biofeedback training.

Most studies investigating knee interlimb asymmetry have investigated mobility tasks that require fairly low mechanical demand at the knee (ie, level walking).[20,23,29,43] Investigating more physically challenging mobility tasks with TKA patients is necessary to detect potential compensation strategies that may not be detectable during lower-demand tasks.[15] As the number of joint arthroplasty procedures continues to increase in younger and more active individuals,[13,31,38] investigating more physically demanding mobility tasks is needed to provide valuable information on movement behaviors and potential asymmetries that could be mitigated through postoperative rehabilitation. However, a more pragmatic way to provide KEM biofeedback is needed in the clinic, and further research is required to assist in developing this technology within the rehabilitation setting.

The finding of peak KEM asymmetry prior to biofeedback training supports the premise that compensatory strategies remain after successful recovery following TKA. Although immediate improvements in knee interlimb symmetry were observed using KEM biofeedback, these changes were a result of motor adaptations during a single treatment session, and further investigation is needed to determine whether these results are retained over time. Further, a longitudinal cohort study is also needed to determine whether motor training using KEM biofeedback can be effective at long-term retention and ultimately lead to improved functional performance.

This study has several limitations that should be noted. No long-term follow-up

TABLE 3 Adjusted Between-Group Comparison of the Effect of Biofeedback on Interlimb Peak KEM Asymmetry After Controlling for the Nonbiofeedback Condition

Variable	vGRF Group (n = 15)*	KEM Group (n = 15)*	Mean Difference†	Effect Size, Cohen f^2‡	*P* Value
Post biofeedback interlimb difference in peak KEM, Nm/kg	0.17 ± 0.02	0.08 ± 0.02	0.09 ± 0.03 (0.01, 0.19)	0.24	.01

Abbreviations: KEM, knee extensor moment; vGRF, vertical ground reaction force.
**Values are adjusted postbiofeedback mean ± standard error between the surgical knee and the nonsurgical knee. Mean values equal to zero signified perfect symmetry and values greater than zero signified higher asymmetry.*
†Values are mean difference ± standard error (95% confidence interval) from the analysis of covariance model and marginal estimation.
‡Effect-size categories: 0.02, small; 0.15, medium; 0.35, large.

measures were obtained following biofeedback training. Although the authors were able to draw conclusions regarding immediate correction in peak KEM asymmetry in the short term, future studies should assess knee interlimb asymmetry in joint mechanics with larger sample sizes and longer follow-up. Despite the increases in peak knee interlimb symmetry seen using KEM biofeedback, these findings were observed during single treatment sessions of training, and further research is needed to determine whether longitudinal training can lead to long-term retention.

Although the authors studied the effectiveness of 2 forms of biofeedback in correcting peak knee interlimb asymmetry during a more physically demanding mobility task, there are many factors (eg, surgeon, implant design, strength, knee motion, etc) that could influence KEM asymmetry. Furthermore, no randomization of group assignment or inclusion of a control group might have led to bias in the results. Data from this study cannot determine the cause of the knee interlimb asymmetry, and it is important to note that causes of asymmetry can be multifactorial in nature. Last, the clinical relevance of using a laboratory-based biofeedback option is a concern, as most rehabilitation clinics do not have access to this type of technology. However, determining the influence of a joint-specific kinetic form of biofeedback is a necessary first step before more practical variables of training can be implemented.

CONCLUSION

Knee extensor moment biofeedback training reduced interlimb peak KEM asymmetry 3 months post TKA. In contrast, vGRF biofeedback training did not change peak interlimb KEM asymmetry. These findings indicate that patients 3 months post TKA can immediately mitigate interlimb KEM asymmetry during a higher-demand mobility task; however, this is dependent on the type of biofeedback provided. ◉

KEY POINTS

FINDINGS: Knee extensor moment biofeedback was effective at correcting interlimb peak knee moment asymmetry 3 months post total knee arthroplasty when compared to vertical ground reaction force biofeedback.

IMPLICATIONS: Correcting interlimb knee extensor moment asymmetry could result in improved overall recovery of the surgical limb, leading to improved longevity of independent living and recreational opportunities.

CAUTION: Only the immediate effects of biofeedback training were evaluated. Further research to explore the retention of prolonged training beyond this study's single treatment session design is warranted.

REFERENCES

1. Alnahdi AH, Zeni JA, Snyder-Mackler L. Gait after unilateral total knee arthroplasty: frontal plane analysis. *J Orthop Res*. 2011;29:647-652. https://doi.org/10.1002/jor.21323
2. Bade MJ, Kohrt WM, Stevens-Lapsley JE. Outcomes before and after total knee arthroplasty compared to healthy adults. *J Orthop Sports Phys Ther*. 2010;40:559-567. https://doi.org/10.2519/jospt.2010.3317
3. Baker PN, van der Meulen JH, Lewsey J, Gregg PJ. The role of pain and function in determining patient satisfaction after total knee replacement. Data from the National Joint Registry for England and Wales. *J Bone Joint Surg Br*. 2007;89:893-900. https://doi.org/10.1302/0301-620X.89B7.19091
4. Christensen JC, LaStayo PC, Mizner RL, et al. Joint mechanical asymmetries during low- and high-demand mobility tasks: comparison between total knee arthroplasty and healthy-matched peers. *Gait Posture*. 2018;60:104-110. https://doi.org/10.1016/j.gaitpost.2017.11.017
5. Christensen JC, Mizner RL, Foreman KB, Marcus RL, Pelt CE, LaStayo PC. Quadriceps weakness preferentially predicts detrimental gait compensations among common impairments after total knee arthroplasty. *J Orthop Res*. 2018;36:2355-2363. https://doi.org/10.1002/jor.23894
6. Christiansen CL, Bade MJ, Davidson BS, Dayton MR, Stevens-Lapsley JE. Effects of weight-bearing biofeedback training on functional movement patterns following total knee arthroplasty: a randomized controlled trial. *J Orthop Sports Phys Ther*. 2015;45:647-655. https://doi.org/10.2519/jospt.2015.5593
7. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
8. Debbi EM, Bernfeld B, Herman A, et al. Frontal plane biomechanics of the operated and non-operated knees before and after unilateral total knee arthroplasty. *Clin Biomech (Bristol, Avon)*. 2015;30:889-894. https://doi.org/10.1016/j.clinbiomech.2015.07.013
9. de Winter JCF. Using the Student's *t*-test with extremely small sample sizes. *Pract Assess Res Eval*. 2013;18:1-12.
10. Farquhar SJ, Reisman DS, Snyder-Mackler L. Persistence of altered movement patterns during a sit-to-stand task 1 year following unilateral total knee arthroplasty. *Phys Ther*. 2008;88:567-579. https://doi.org/10.2522/ptj.20070045
11. Gaffney BM, Harris MD, Davidson BS, Stevens-Lapsley JE, Christiansen CL, Shelburne KB. Multi-joint compensatory effects of unilateral total knee arthroplasty during high-demand tasks. *Ann Biomed Eng*. 2016;44:2529-2541. https://doi.org/10.1007%10439-015-1524-z
12. Hatfield GL, Hubley-Kozey CL, Astephen Wilson JL, Dunbar MJ. The effect of total knee arthroplasty on knee joint kinematics and kinetics during gait. *J Arthroplasty*. 2011;26:309-318. https://doi.org/10.1016/j.arth.2010.03.021
13. Hawker GA. Who, when, and why total joint replacement surgery? The patient's perspective. *Curr Opin Rheumatol*. 2006;18:526-530. https://doi.org/10.1097/01.bor.0000240367.62583.51
14. Hong SW, Wang TM, Lu TW, Li JD, Leu TH, Ho WP. Redistribution of intra- and inter-limb support moments during downhill walking on different slopes. *J Biomech*. 2014;47:709-715. https://doi.org/10.1016/j.jbiomech.2013.11.028
15. Komnik I, Weiss S, Fantini Pagani CH, Potthast W. Motion analysis of patients after knee arthroplasty during activities of daily living – a systematic review. *Gait Posture*. 2015;41:370-377. https://doi.org/10.1016/j.gaitpost.2015.01.019
16. Lumley T, Diehr P, Emerson S, Chen L. The importance of the normality assumption in large public health data sets. *Annu Rev Public Health*. 2002;23:151-169. https://doi.org/10.1146/annurev.publhealth.23.100901.140546
17. Mandeville D, Osternig LR, Chou LS. The effect of total knee replacement on dynamic support of the body during walking and stair ascent. *Clin Biomech (Bristol, Avon)*. 2007;22:787-794. https://doi.org/10.1016/j.clinbiomech.2007.04.002
18. McClelland J, Zeni J, Jr., Haley RM, Snyder-Mackler L. Functional and biomechanical outcomes after using biofeedback for retraining symmetrical movement patterns after total knee arthroplasty: a case report. *J Orthop Sports Phys Ther*. 2012;42:135-144. https://doi.org/10.2519/jospt.2012.3773
19. McClelland JA, Feller JA, Menz HB, Webster KE. Patterns in the knee flexion-extension moment profile during stair ascent and descent in patients with total knee arthroplasty. *J Biomech*.

2014;47:1816-1821. https://doi.org/10.1016/j.jbiomech.2014.03.026
20. McClelland JA, Webster KE, Feller JA. Gait analysis of patients following total knee replacement: a systematic review. *Knee*. 2007;14:253-263. https://doi.org/10.1016/j.knee.2007.04.003
21. McGinnis K, Snyder-Mackler L, Flowers P, Zeni J. Dynamic joint stiffness and co-contraction in subjects after total knee arthroplasty. *Clin Biomech (Bristol, Avon)*. 2013;28:205-210. https://doi.org/10.1016/j.clinbiomech.2012.11.008
22. McMahon M, Block JA. The risk of contralateral total knee arthroplasty after knee replacement for osteoarthritis. *J Rheumatol*. 2003;30:1822-1824.
23. Milner CE. Interlimb asymmetry during walking following unilateral total knee arthroplasty. *Gait Posture*. 2008;28:69-73. https://doi.org/10.1016/j.gaitpost.2007.10.002
24. Milner CE. Is gait normal after total knee arthroplasty? Systematic review of the literature. *J Orthop Sci*. 2009;14:114-120. https://doi.org/10.1007/s00776-008-1285-8
25. Mitchell MN. *Interpreting and Visualizing Regression Models Using Stata*. College Station, TX: Stata Press; 2012.
26. Mizner RL, Petterson SC, Clements KE, Zeni JA, Jr., Irrgang JJ, Snyder-Mackler L. Measuring functional improvement after total knee arthroplasty requires both performance-based and patient-report assessments: a longitudinal analysis of outcomes. *J Arthroplasty*. 2011;26:728-737. https://doi.org/10.1016/j.arth.2010.06.004
27. Mizner RL, Petterson SC, Snyder-Mackler L. Quadriceps strength and the time course of functional recovery after total knee arthroplasty. *J Orthop Sports Phys Ther*. 2005;35:424-436. https://doi.org/10.2519/jospt.2005.35.7.424
28. Mizner RL, Petterson SC, Stevens JE, Vandenborne K, Snyder-Mackler L. Early quadriceps strength loss after total knee arthroplasty: the contributions of muscle atrophy and failure of voluntary muscle activation. *J Bone Joint Surg Am*. 2005;87:1047-1053. https://doi.org/10.2106/JBJS.D.01992
29. Mizner RL, Snyder-Mackler L. Altered loading during walking and sit-to-stand is affected by quadriceps weakness after total knee arthroplasty. *J Orthop Res*. 2005;23:1083-1090. https://doi.org/10.1016/j.orthres.2005.01.021
30. Myles CM, Rowe PJ, Walker CR, Nutton RW. Knee joint functional range of movement prior to and following total knee arthroplasty measured using flexible electrogoniometry. *Gait Posture*. 2002;16:46-54. https://doi.org/10.1016/S0966-6362(01)00198-9
31. National Institutes of Health. Consensus Development Conference Statement on Total Knee Replacement. Bethesda, MD: National Institutes of Health; 2003.
32. Ritter MA, Carr KD, Keating EM, Faris PM. Long-term outcomes of contralateral knees after unilateral total knee arthroplasty for osteoarthritis. *J Arthroplasty*. 1994;9:347-349. https://doi.org/10.1016/0883-5403(94)90043-4
33. Selya AS, Rose JS, Dierker LC, Hedeker D, Mermelstein RJ. A practical guide to calculating Cohen's f^2, a measure of local effect size, from PROC MIXED. *Front Psychol*. 2012;3:111. https://doi.org/10.3389/fpsyg.2012.00111
34. Shakoor N, Block JA, Shott S, Case JP. Nonrandom evolution of end-stage osteoarthritis of the lower limbs. *Arthritis Rheum*. 2002;46:3185-3189. https://doi.org/10.1002/art.10649
35. Sheehan RC, Gottschall JS. At similar angles, slope walking has a greater fall risk than stair walking. *Appl Ergon*. 2012;43:473-478. https://doi.org/10.1016/j.apergo.2011.07.004
36. Stevens-Lapsley JE, Balter JE, Kohrt WM, Eckhoff DG. Quadriceps and hamstrings muscle dysfunction after total knee arthroplasty. *Clin Orthop Relat Res*. 2010;468:2460-2468. https://doi.org/10.1007/s11999-009-1219-6
37. Wiik AV, Aqil A, Tankard S, Amis AA, Cobb JP. Downhill walking gait pattern discriminates between types of knee arthroplasty: improved physiological knee functionality in UKA versus TKA. *Knee Surg Sports Traumatol Arthrosc*. 2015;23:1748-1755. https://doi.org/10.1007/s00167-014-3240-x
38. Williams DP, Price AJ, Beard DJ, et al. The effects of age on patient-reported outcome measures in total knee replacements. *Bone Joint J*. 2013;95-B:38-44. https://doi.org/10.1302/0301-620X.95B1.28061
39. Winter DA. *Biomechanics and Motor Control of Human Movement*. 3rd ed. Hoboken, NJ: Wiley; 2005.
40. Winter DA. Overall principle of lower limb support during stance phase of gait. *J Biomech*. 1980;13:923-927. https://doi.org/10.1016/0021-9290(80)90162-1
41. Worsley P, Stokes M, Barrett D, Taylor M. Joint loading asymmetries in knee replacement patients observed both pre- and six months post-operation. *Clin Biomech (Bristol, Avon)*. 2013;28:892-897. https://doi.org/10.1016/j.clinbiomech.2013.07.014
42. Yoshida Y, Mizner RL, Ramsey DK, Snyder-Mackler L. Examining outcomes from total knee arthroplasty and the relationship between quadriceps strength and knee function over time. *Clin Biomech (Bristol, Avon)*. 2008;23:320-328. https://doi.org/10.1016/j.clinbiomech.2007.10.008
43. Yoshida Y, Zeni J, Jr., Snyder-Mackler L. Do patients achieve normal gait patterns 3 years after total knee arthroplasty? *J Orthop Sports Phys Ther*. 2012;42:1039-1049. https://doi.org/10.2519/jospt.2012.3763
44. Zeni J, Jr., Abujaber S, Flowers P, Pozzi F, Snyder-Mackler L. Biofeedback to promote movement symmetry after total knee arthroplasty: a feasibility study. *J Orthop Sports Phys Ther*. 2013;43:715-726. https://doi.org/10.2519/jospt.2013.4657

CÉSAR FERNÁNDEZ-DE-LAS-PEÑAS, PT, PhD, DMSc[1,2] • **RICARDO ORTEGA-SANTIAGO,** PT, PhD[1,2]
HOMID FAHANDEZH-SADDI DÍAZ, MD, PhD[3] • **JAIME SALOM-MORENO,** PT, PhD[1,2] • **JOSHUA A. CLELAND,** PT, PhD[4-6]
JUAN A. PAREJA, MD, PhD[7] • **JOSÉ L. ARIAS-BURÍA,** PT, MSc, PhD[1,2]

Cost-Effectiveness Evaluation of Manual Physical Therapy Versus Surgery for Carpal Tunnel Syndrome: Evidence From a Randomized Clinical Trial

BACKGROUND: Carpal tunnel syndrome (CTS) results in substantial societal costs and can be treated either by nonsurgical or surgical approaches.

OBJECTIVE: To evaluate differences in cost-effectiveness of manual physical therapy versus surgery in women with CTS.

METHODS: In this randomized clinical trial, 120 women with a clinical and an electromyographic diagnosis of CTS were randomized through concealed allocation to either manual physical therapy or surgery. Interventions consisted of 3 sessions of manual physical therapy, including desensitization maneuvers of the central nervous system, or decompression/release of the carpal tunnel. Societal costs and health-related quality of life (estimated by the European Quality of Life-5 Dimensions [EQ-5D] scale) over 1 year were used to generate incremental cost per quality-adjusted life year ratios for each treatment.

RESULTS: The analysis was possible for 118 patients (98%). Incremental quality-adjusted life years showed greater cost-effectiveness in favor of manual physical therapy (difference, 0.135; 95% confidence interval: 0.134, 0.136). Manual therapy was significantly less costly than surgery (mean difference in cost per patient, €2576; $P<.001$). Patients in the surgical group received a greater number of other treatments and made more visits to medical doctors than those receiving manual physical therapy ($P = .02$). Absenteeism from paid work was significantly higher in the surgery group ($P<.001$). The major contributors to societal costs were the treatment protocol (surgery versus manual therapy mean difference, €106 980) and absenteeism from paid work (surgery versus manual physical therapy mean difference, €42 224).

CONCLUSION: Manual physical therapy, including desensitization maneuvers of the central nervous system, has been found to be equally effective but less costly (ie, more cost-effective) than surgery for women with CTS. From a cost-benefit perspective, the proposed CTS manual physical therapy intervention can be considered.

LEVEL OF EVIDENCE: Economic and decision analyses, level 1b. *J Orthop Sports Phys Ther 2019;49(2):55-63. Epub 30 Nov 2018. doi:10.2519/jospt.2019.8483*

KEY WORDS: *carpal tunnel syndrome, cost-effectiveness, physical therapy, surgery*

Carpal tunnel syndrome (CTS) is considered the most common entrapment neuropathy of the upper extremity, with a prevalence ranging from 6.3% to 11.7%.[40] Because CTS usually affects middle-aged active workers,[23] it is associated with substantial health care costs and economic burden, including loss of work productivity. For instance, the income loss per patient over a period of 6 years has been estimated to be $45 000 to $89 000.[10] The overall cost associated with CTS in the United States exceeds $2 billion annually.[38] Additionally, a recent study has shown that CTS-associated burden extends beyond direct costs to include adverse financial impacts and household disruption.[9]

Management of CTS includes both nonsurgical and surgical approaches, but there is no consensus on which therapeutic

[1]Department of Physical Therapy, Occupational Therapy, Rehabilitation and Physical Medicine, Universidad Rey Juan Carlos, Alcorcón, Spain. [2]Cátedra de Investigación y Docencia en Fisioterapia, Terapia Manual y Punción Seca, Universidad Rey Juan Carlos, Alcorcón, Spain. [3]Department of Traumatology and Orthopaedic Surgery, Hospital Universitario Fundación Alcorcón, Madrid, Spain. [4]Department of Physical Therapy, Franklin Pierce University, Manchester, NH. [5]Rehabilitation Services, Concord Hospital, Concord, NH. [6]Manual Therapy Fellowship Program, Regis University, Denver, CO. [7]Department of Neurology and Neurophysiology, Hospital Universitario Fundación Alcorcón, Madrid, Spain. The study was approved by the Hospital Universitario Fundación Alcorcón Institutional Review Board (PI01223-HUFA12/14), and the original clinical trial was registered with ClinicalTrials.gov (NCT01789645). A research project grant (FIS PI14/ 00364) from the Health Institute Carlos III (PN I+D+I 2014-2017; Spanish Government) funded this work. The sponsor had no role in the design, conduct, or analysis of the study other than the usual peer-review process. The corresponding author was responsible for the decision to submit the manuscript for publication, and the sponsor did not participate in this decision. All authors had full access to the data. The corresponding author takes responsibility for the integrity of the data and the accuracy of the data analysis. 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 César Fernández-de-las-Peñas, Facultad de Ciencias de la Salud, Universidad Rey Juan Carlos, Avenida de Atenas s/n, 28922 Alcorcón, Madrid, Spain. E-mail: cesar.fernandez@urjc.es

strategy is more beneficial.[13] Because there is limited evidence supporting the use of nonsurgical treatments (eg, exercise and mobilization techniques,[26] splinting,[25] or ultrasound[27]), surgery continues to be the intervention most often recommended,[1] as differences with nonsurgical treatment are sometimes smaller than expected.[37,43] Additionally, the vast majority of patients with CTS usually attempt to avoid surgery.[17] Because carpal tunnel release shows the highest use rate of surgical procedures performed in the United States for the upper extremity,[16] an analysis of the economic impact of the different interventions for CTS is needed.

Several studies have reported slight differences in costs among the various types of surgery (ie, open or endoscopic[20,36,41]); however, only 2 articles have compared the cost-effectiveness of nonsurgical interventions versus surgery for CTS. Korthals-de Bos et al[19] found similar health costs for individuals with CTS receiving surgery (mean cost per patient, €2126) or receiving splint therapy (cost per patient, €2111). Similarly, in a retrospective study, Pomerance et al[31] also observed similar costs between individuals receiving surgery (mean ± SD cost, $3068 ± $983) and those receiving nonsurgical treatment consisting of splint and/or local corticoid injections (mean ± SD cost, $3335 ± $2097). Because both studies also reported outcomes that were superior for the surgery group, the incremental cost-utility ratio was slightly favorable for surgery.[19,31]

A recent study investigating the effects of manual therapies, including desensitization maneuvers of the central nervous system, versus surgery in a sample of women with CTS found that 3 sessions of manual physical therapy resulted in better short-term outcomes (1 and 3 months) and similar long-term effects (6 and 12 months) on pain intensity and function compared to surgery.[8] This trial provides promising results for the nonsurgical management of CTS; however, there is currently no published study comparing health care costs between manual physical therapy and surgery, as previous trials only examined local treatments such as splints or local injections.[19,31] Therefore, the purpose of the current analysis was to evaluate the cost-effectiveness of manual therapy compared with that of surgery in women with CTS, undertaken alongside the aforementioned clinical trial.[8]

METHODS

Study Design

An economic evaluation was performed alongside a randomized clinical trial[8] in a general hospital in Madrid, Spain, to evaluate the cost-effectiveness of manual physical therapy, including desensitization maneuvers of the central nervous system, versus surgery for women with CTS. Full details of the trial, participants, interventions, and results of the clinical outcomes are reported elsewhere.[8] Differences in the analysis and reporting of clinical efficacy and economic evaluation reflect the different research objectives of these efforts. This study was approved by the Hospital Universitario Fundación Alcorcón Institutional Review Board (PI01223-HUFA12/14), and the original clinical trial was registered with ClinicalTrials.gov (NCT01789645).

Participants

Women who were diagnosed with CTS, according to clinical and electrophysiological findings from a local regional hospital (Madrid, Spain), were consecutively screened for eligibility criteria. To be included in the analysis, patients had to exhibit the following clinical signs: pain and paresthesia in the median nerve distribution, increasing symptoms during the night, positive Tinel sign, and positive Phalen sign. Symptoms had to have persisted for at least 12 months. Further, the electrodiagnostic examination had to reveal deficits of sensory and motor median nerve conduction, according to the guidelines of the American Association of Electrodiagnostic Medicine, the American Academy of Neurology, and the American Academy of Physical Medicine and Rehabilitation.[15] Patients were excluded if they exhibited any of the following: (1) sensory/motor deficit in the ulnar or radial nerves, (2) older than 65 years of age, (3) previous hand surgery or steroid injection treatment, (4) multiple diagnoses on the upper extremity, (5) cervical, shoulder, and/or upper extremity trauma, (6) any systemic disease causing CTS (eg, diabetes mellitus, thyroid disease), (7) comorbid musculoskeletal medical conditions (eg, rheumatoid arthritis and/or fibromyalgia), (8) pregnancy, or (9) presence of depressive symptoms (Beck Depression Inventory-II score greater than 8 points). All participants signed an informed-consent form prior to their inclusion in the study.

Interventions

Patients were randomly assigned to receive manual physical therapy or a surgical procedure, as previously described.[20] Patients allocated to the manual physical therapy group received 3 treatment sessions of manual therapy, including desensitization maneuvers of the central nervous system of 30-minute duration, once a week. Briefly, the desensitization maneuvers consisted of soft tissue mobilization, including manual techniques directed at anatomical sites of potential entrapment of the median nerve. Additionally, lateral glides were applied to the cervical spine, and tendon- and nerve-gliding exercise interventions were also included. Finally, patients received an educational teaching session on performing the tendon- and nerve-gliding exercises at home. Full description of this intervention can be found elsewhere.[8]

Patients randomly allocated to the surgery group underwent open or endoscopic release of the carpal tunnel. For pragmatic reasons, and because there is no evidence supporting one particular surgical procedure, surgery was based on both surgeon and patient preference.[45] Patients allocated to this group also received the same educational session for performing tendon- and nerve-gliding exercises that the manual physical therapy group received.[8]

Economic Evaluation

A societal perspective, including direct health care costs, direct non–health care costs, and indirect costs due to CTS, was used as the basis for the economic evaluation.[5] Direct health care costs included the costs of each treatment (ie, number of sessions, number of visits to manual physical therapists), additional visits to health care providers (ie, medical specialist or other health care professional), additional received treatments, prescribed medications, and professional home care. Direct non–health care costs included only costs of over-the-counter medications, time spent visiting a health care provider, and travel expenses. Indirect costs of lost productivity due to CTS-related absence from work were also included in the main analysis.

Data were collected by patients in a diary, where they registered the number of visits to medical doctors, medication intake, other treatments received, and any other circumstance or action taken related to their CTS during the follow-up period.[11] This cost diary was returned to a research assistant for analysis. To more accurately represent the health care costs of the entire country, direct health care and non–health care costs were estimated by averaging official costs of 5 representative regions of Spain (Comunidad de Madrid,[33] Cataluña,[29] País Vasco,[39] Andalucía,[32] and Castilla-León[34]). All health care costs used in this study were those officially published for each geographical region. Indirect costs of lost productivity for paid work were also calculated based on the current employment status of the participant at the time of the trial.[14] Finally, national prices of other daily costs were also consulted (see **TABLE 1**).

Outcomes

Health-related quality of life was measured at baseline and at each follow-up period (1, 3, 6, and 12 months after treatment), using the 5-level version of the European Quality of Life-5 Dimensions scale (EQ-5D-5L)[12,18] in written form, distributed by an examiner blinded to the treatment allocation of the participants. Responses were converted to an overall

TABLE 1 Health Care and Non–Health Care Costs

		Manual Therapy Group		Surgery Group			
Cost Type/Modality	**Unit Cost**	**n**	**Total Cost**	**n**	**Total Cost**	**Difference**	***P* Value**
Per-protocol treatment							
Manual therapy[29,32-34,39]	€10 ± €6	180 sessions (60 × 3 sessions)	€1800	0	€0		
Surgery[29,32-34,39]	€1813 ± €76	0	€0	60	€108780		
Total			€1800		€108780	€–106980	<.001
Direct health care							
Other treatments							.13
Yes		15 (26%)		10 (17%)			
No		43 (74%)		50 (83%)			
Type							
Wrist band*	€31 ± €5	6	€186	1	€31	€155	
Extra physical therapy[29,32-34,39]	€10 ± €6	41 sessions	€410	234 sessions	€2340	€–1930	
Local infiltration[29,32-34,39]	€2	6	€12	0	€0	€12	
Total			€608		€2371	€–1763	.001
Medical consultation							.15
Yes		17 (29%)		59 (98%)			
No		41 (71%)		1 (2%)			
Type							
Primary care[29,32-34,39]	€44 ± €7	5	€220	4	€176	€44	
Traumatologist + NL + EMG[29,32-34,39]	€63 ± €18	32	€2016	70	€4410	€–2394	
Urgent care[29,32-34,39]	€193 ± €72	2	€386	0	€0	€386	
Total			€2622		€4586	€–1964	.02
Direct non–health care							
Neobrufen 600 mg	€0.05						
Number of tablets†		196 (35%)	€10	358 (65%)	€20		
Total			€10		€20	€–10	.25

Table continues on page 58.

TABLE 1 Health Care and Non–Health Care Costs (continued)

		Manual Therapy Group		Surgery Group			
Cost Type/Modality	**Unit Cost**	**n**	**Total Cost**	**n**	**Total Cost**	**Difference**	***P* Value**
Travel via public transport (EMT)	€1.50						
Number of trips‡		148 (32%)	€222	316 (68%)	€474		
Total			€222		€474	€−252	.08
Indirect							
Absence from paid work	€13						.001
Yes		4 (7%)		52 (87%)			
No		54 (93%)		8 (13%)			
Absence from paid work, d[14]		112	€1456	3360	€43680		.001
Total			€1456		€43680	€−42224	<.001
Subsequent surgery[29,32-34,39]	€1813 ± €76						
Yes		3 (5%)		4 (7%)			
No		55 (95%)		56 (93%)			
Number of surgeries[29,32-34,39]		3	€5439	4 reoperated	€7252		
Total			€5439		€7252	€−1813	.32
Societal							
Excluding work absence							
Yes		0 (0%)		0 (0%)			
No		58 (100%)		60 (100%)			
Total			€0		€0	€0	
Total costs			€12147		€167143	€−154996	<.001

Abbreviations: EMG, electromyography; EMT, emergency medical technician; NL, neurologist.
**(1) http://tienda.fisaude.com; (2) http://www.efisioterapia.net/tienda/; (3) http://www.lacasadelfisio.com/#; (4) http://www.cramersportsmed.com; (5) http://www.lacasadelmasajista.com; (6) http://www.quirumed.com/es/fisioterapia-y-masaje?gclid=CjwKEAiA6YDBBRDwtpTQnYzx5lASJAC57ObMBoOnTJKoE0cc8Skds-_mZMf_-0yiUiyZ8DGph-Cl5BoCuBPw_wcB.*
†Por el que se regula el sistema de precios de referencia y de agrupaciones homogéneas de medicamentos en el Sistema Nacional de Salud y determinados sistemas de información en materia de financiación y precios de los medicamentos y productos sanitarios, Real Decreto 177/2014, 21 de marzo (2014).
‡http://www.emtmadrid.es/ViajarenBus/Titulosytarifas.

utility score by applying cross-walk index values for Spain.[42] Quality-adjusted life years (QALYs) were estimated for each participant using area-under-the-curve analysis, with linear interpolation between observations resulting from the fitting-the-curve exercise.[22]

The QALY combines length and quality of life into a single index number between 0 and 1, where 0 corresponds to a health state judged to be equivalent to death and 1 corresponds to optimal health.[22] A graphical analysis of the utility results measured in QALYs was conducted throughout the study follow-up periods. The scatter plot was fitted to a curve for each group, by comparing the trends of both groups and creating a curve to fit the measurement points. This process is more realistic than using a linear evolution of the quality-of-life data (**FIGURE 1**).

Subsequently, an incremental cost-effectiveness analysis was undertaken using the EQ-5D-5L, to calculate the cost per additional QALY gained over the treatment period.

Statistical Analysis

Sample size was based on changes in the intensity of hand pain at 1-year follow-up.[8] An incremental cost-effectiveness analysis was hence performed.[5] To compare health costs between both groups, the authors calculated a deterministic cost-utility value and a probabilistic one using bootstrapping techniques and computing confidence intervals (CIs). The 95% CIs were obtained by bias-corrected and accelerated bootstrapping, choosing 1000 iterations.

The primary outcome was the incremental cost-effectiveness ratio, calculated by dividing the incremental costs by the incremental QALYs. Uncertainty was explored by graphical display of cost-effectiveness planes and acceptability curves.[6]

A 1-way sensitivity analysis, calculating the results of the evaluation using a value above and below that used in the base case, was used to explore the uncertainty around each parameter by examining the changes in the results in the range of parameter values.[4] Statistical analysis was performed using Stata Version 13.1 (StataCorp LLC, College Station, TX). Cost-effectiveness planes and acceptability

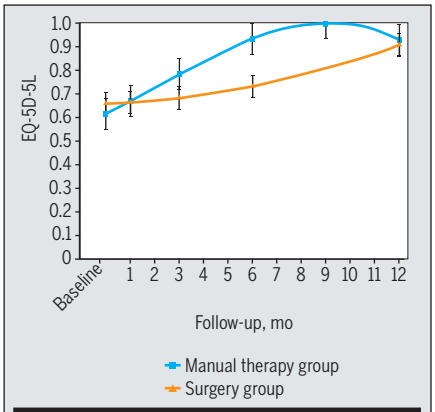


FIGURE 1. Progression in quality of life in both groups. Abbreviation: EQ-5D-5L, 5-level version of the European Quality of Life-5 Dimensions scale.

curves were generated using Excel Version 16.0 (Microsoft Corporation, Redmond, WA).

RESULTS

Of 120 patients initially included in the trial and randomly allocated to the manual therapy group (n = 60) or the surgery group (n = 60), 113 (94%) were included in the final clinical analysis[8] and 118 (98%) were included in the economic analysis. Within patients allocated to the manual physical therapy group, 3 individuals were excluded from the clinical analysis because they received surgery to the study hand. Within patients allocated to the surgery group, 4 received surgery to the contralateral hand but were included in the economic analysis. The flow of participants leading to the economic analysis is illustrated in **FIGURE 2**. Demographic features did not differ between groups at baseline (**TABLE 2**).

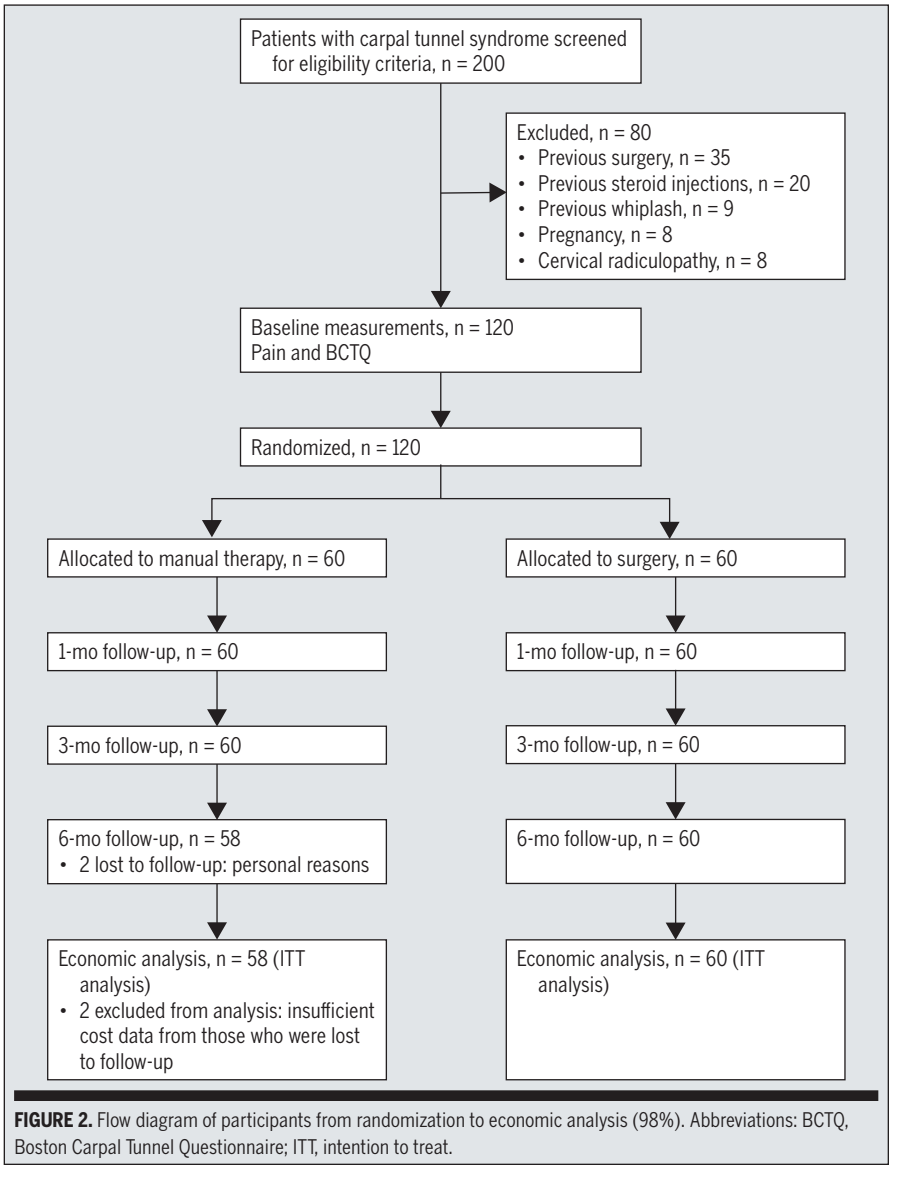


FIGURE 2. Flow diagram of participants from randomization to economic analysis (98%). Abbreviations: BCTQ, Boston Carpal Tunnel Questionnaire; ITT, intention to treat.

Costs

TABLE 1 summarizes and compares the direct health care and non–health care costs of each group. The researchers found a significant between-group difference for per-protocol costs: the surgery group was significantly more expensive (P<.001) than the manual physical therapy group. Additionally, patients in the surgery group also received a greater number of other treatments, mostly complementary manual physical therapy (P = .001), and also made more visits to their orthopaedic surgeon and/or neurologist (P = .02) than those in the manual physical therapy group.

Total indirect costs consisted mainly of lost productivity (work absenteeism) within the surgery group, and surgery or second surgery within the manual therapy and surgery groups, respectively (see **TABLE 1**). Absenteeism from paid work was significantly higher within the surgery group than in the manual physical therapy group, both in terms of the number of individuals missing work (n = 52 [86.7%] versus n = 4 [6.9%]; P = .001) and the number of days off from work (mean, 65 versus 28 days; total, 3360 versus 112 days; P = .001). The major contributors to societal costs were treatment protocol (surgery versus manual physical therapy; mean difference, €106 980; 68% of societal costs overall) and absence from paid work (surgery versus manual therapy; mean difference, €42 224; 28% of societal costs overall).

Mean cost (including work absence) was €12 147 for manual physical therapy

TABLE 2 Baseline Characteristics at the Beginning of the Trial*

	Manual Therapy (n = 60)	Surgery (n = 60)
Age, y	47 ± 10	46 ± 9
Years with pain	3.1 ± 2.7	3.5 ± 3.1
Occupation, n (%)		
Work at home	33 (55)	35 (58)
Secretary/office	27 (45)	25 (42)
Symptom distribution, n (%)		
Unilateral, right side	10 (17)	15 (25)
Unilateral, left side	3 (5)	5 (8)
Bilateral	47 (78)	40 (67)
CTS severity, n (%)		
Minimal	16 (27)	17 (28)
Moderate	23 (38)	20 (33)
Severe	21 (35)	23 (38)
Pain intensity (NPRS, 0-10)		
Average	4.8 ± 1.5	4.9 ± 2.2
Worst pain in the last week	6.6 ± 1.7	7.0 ± 2.0
BCTQ (1-5)		
Functional status subscale	2.3 ± 0.5	2.4 ± 0.6
Severity status subscale	2.5 ± 0.7	2.7 ± 0.6
BDI-II (0-21)	4.2 ± 2.9	3.8 ± 2.7
EQ-5D-5L (0-1)	0.61 ± 0.09	0.66 ± 0.05

Abbreviations: BCTQ, Boston Carpal Tunnel Questionnaire; BDI-II, Beck Depression Inventory; CTS, carpal tunnel syndrome; EQ-5D-5L, 5-level version of the European Quality of Life-5 Dimensions scale; NPRS, numeric pain-rating scale.
**Values are mean ± SD unless otherwise indicated.*

and €167 143 for surgery. Similarly, mean cost per participant (including work absence) was statistically higher in the surgery group than in the manual physical therapy group (€2785 versus €209, $P<.001$) (**TABLE 3**). Therefore, significant incremental costs were observed for the surgery group ($P<.001$).

Quality of Life

Utilities estimated over the follow-up were 50.15 QALYs for the manual physical therapy group and 44.3 QALYs for the surgery group in the deterministic set of results (**TABLE 3**). Incremental QALYs showed significantly greater benefit for the manual physical therapy group (mean difference, 0.135; 95% CI: 0.134, 0.136). Baseline EQ-5D-5L score was a significant independent predictor of 1-year QALYs (mean difference, 0.05; $P = .92$).

Cost-Effectiveness

The deterministic incremental cost-effectiveness ratio revealed a dominant position of the manual physical therapy group; that is, it was less costly (€–154 996) and more effective (5.844 QALYs) than the surgery group. The probabilistic result using bootstrapping was similar to the result of the deterministic analysis: a dominant position of manual physical therapy as less costly (mean cost difference, €–137 378; 95% CI: €–146 531, €–128 225) and more effective (mean QALY difference, 8.13; 95% CI: 7.1, 9.16) than surgery. The cost-effectiveness plane is graphically represented in **FIGURE 3**. Bootstrapped cost-utility pairs fell within the southeast quadrant, indicating cost savings and increased effectiveness of manual physical therapy compared to surgery. Thus, manual physical therapy was more likely to be cost-effective than surgery, with 100% of the iterations falling within the dominant area.

The constructed model was robust for all analyses of univariate sensitivity where there were no significant changes in the direction of the results, because manual physical therapy proved more effective but cheaper than surgery (incremental cost-effectiveness ratio). The results of the univariate sensitivity analysis showed the parameter of "urgency" to be the most uncertain parameter, with an incremental cost-effectiveness ratio ranging from €26 606 to €26 782, and the parameter of "medical specialist" to be the second most uncertain parameter, ranging from €26 666 to €26 714.

DISCUSSION

The current study is the first cost-effectiveness analysis to compare manual physical therapy to surgery in women with CTS. The results revealed that manual physical therapy was more cost-effective than surgery, with incremental QALYs showing greater benefit from manual physical therapy. Additionally, the direct health care costs and absenteeism from employment within the surgery group were significantly greater than in the manual physical therapy group. A previous publication found that the same protocol of manual physical therapy results in better clinical outcomes for pain and function in the short term, but similar clinical effects in the long term, compared to surgery.[8] The results from the current paper further complement the results from that study, as the authors found that manual physical therapy, including desensitization maneuvers of the central nervous system, is also more cost-effective than surgery in women with CTS.

Two studies have examined the cost-effectiveness of nonsurgical interventions compared to surgical intervention in patients with CTS,[19,31] the findings of which differ considerably from the current cost-

TABLE 3 Mean Costs Per Individual (Including Work Absence) Between Groups

	Total			Per Patient		
	Manual Therapy	Surgery	Difference	Manual Therapy	Surgery	Difference
Deterministic analysis						
Costs	€12147	€167143	€−154996	€209	€2785	€−2576
QALYs	50.148	44.304	5.844	0.87	0.74	0.13
ICER			−26684.81			−19944.61
Probabilistic analysis						
Costs	€11002	€148380	€−137378	€190	€2473	€−2283
QALYs	52.35	44.22	8.13	0.90	0.73	0.17
ICER			−16899.64			−16307.14

Abbreviations: ICER, incremental cost-effectiveness ratio; QALYs, quality-adjusted life years.

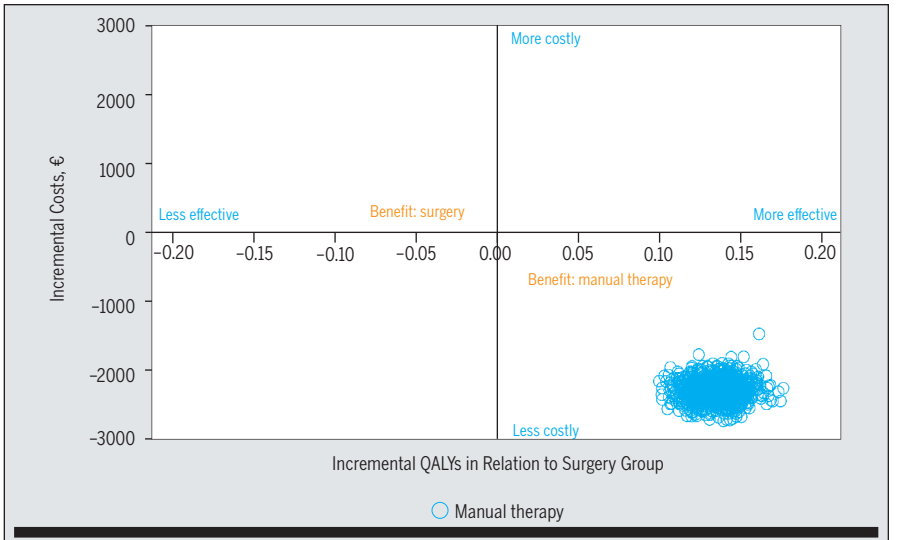


FIGURE 3. Bootstrapped cost-effectiveness pairs fell in the southeast quadrant, indicating cost savings and increased effectiveness of manual physical therapy compared to surgery. Abbreviation: QALYs, quality-adjusted life years.

effectiveness analysis. Pomerance et al[31] found that the direct costs of surgical care were no greater than those of splinting and exercises, and that the incremental cost-effectiveness ratio was in favor of surgery. The study by Korthals-de Bos et al[19] found no between-group differences for direct costs and reported a 90% probability that surgery was more cost-effective than splinting.

Several explanations may account for the differences between these 2 studies and the current research. One possible explanation is that the Pomerance et al[31] study was retrospective and, by nature, inherently had numerous limitations, one of the most crucial being lack of internal validity. Another difference is that interventions between previous studies and the current study were different. Pomerance et al[31] and Korthals-de Bos et al[19] compared surgery to a nonsurgical intervention that was directed solely at the wrist and hand.

In the current study, the manual physical therapy approach included soft tissue mobilization and nerve/tendon-gliding techniques directed at the entire involved upper extremity, according to current nociceptive pain theories on CTS.[8] In fact, the majority of published trials examining the effectiveness of physical therapy versus surgical interventions for CTS have used interventions that solely target the hand, which have demonstrated minimal benefits.[1,17,25-27,37,43] Traditionally, CTS has been considered to be a pathology associated with a peripheral nerve lesion, but there is evidence suggesting that it is a more complex disorder with potential central sensitization processes.[7,44] It has been suggested that manual therapy techniques used in the current trial may have an impact on sensitization mechanisms.[21,24] It is possible that manual physical therapy interventions have the potential to decrease sensitization, which could result in an overall improved quality of life, contributing to greater incremental QALYs and cost-effectiveness.

The authors of the current study found that manual physical therapy had the greatest cost savings associated with work absenteeism, which resulted in a difference of approximately €468 per patient (total difference, €42 224). A pooled analysis reported an incidence rate of CTS of 2.3 cases per 100 persons and a prevalence of 7.8% in a working population.[3] Atroshi et al[2] found that individuals with CTS experienced significantly more missed days from work than people without CTS. Therefore, it seems imperative not only to identify the most effective therapeutic options for the management of CTS, but also to identify treatment options that can reduce the amount of work absenteeism.

Previous data suggest that the duration of sickness absence from work in women with CTS after surgery ranges from 30 to 60 days.[2,28,35] In this study, the surgery group showed a mean of 65 days of work absenteeism, whereas the manual physical therapy group showed a mean of 28 days ($P<.01$). Although return to work after CTS surgery has been associated with different social, economic, psychological, and occupational factors that were not controlled for in the current study,[30] the most remarkable difference was the number of participants who were off work in the surgery group compared to the manual physical therapy group (52 versus 4, respectively). Therefore, reduction in work absenteeism costs was mostly related to the number of participants rather than the number of days.

Although this is the first cost-effectiveness analysis comparing manual physical therapy versus surgery in a population with CTS, there are a few limitations that should be considered when generalizing the results. First, the data were collected in Spain, and it is uncertain how these results may be applied to different countries with different health care systems and costs for manual physical therapy and surgical treatments. Second, the authors did not include a nonintervention control group and could not, therefore, compare manual physical therapy or surgery to the natural course of the condition in terms of QALYs and cost-effectiveness. Last, this economic evaluation was based on outcomes at 12 months after therapy and, therefore, did not investigate short- or medium-term costs.

CONCLUSION

The present study is the first to investigate the cost-effectiveness of manual physical therapy compared to surgical intervention in women with CTS. The results demonstrated that manual physical therapy resulted in lower direct and indirect health care costs. Manual physical therapy also resulted in significantly less work time loss. Additionally, incremental QALYs showed greater benefit in favor of manual physical therapy. Manual physical therapy can be considered a first treatment option for CTS, as it is both clinically effective and cost-effective. ◉

KEY POINTS

FINDINGS: Manual physical therapy, including desensitization maneuvers of the central nervous system, was as effective but less costly (ie, more cost-effective) for women with carpal tunnel syndrome compared to surgery, as it resulted in lower direct and indirect health care costs and less work time loss.

IMPLICATIONS: Manual physical therapy, including desensitization maneuvers of the central nervous system, may be an intervention option for patients with carpal tunnel syndrome as a first line of management prior to, or instead of, surgery.

CAUTION: The generalizability of the results may be limited, as only women from a single hospital were included and the study was conducted in a particular health system. Further, there are no available data on the proper dosage for the manual therapy protocol applied.

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. Atroshi I, Zhou C, Jöud A, Petersson IF, Englund M. Sickness absence from work among persons with new physician-diagnosed carpal tunnel syndrome: a population-based matched-cohort study. *PLoS One*. 2015;10:e0119795. https://doi.org/10.1371/journal.pone.0119795
3. Dale AM, Harris-Adamson C, Rempel D, et al. Prevalence and incidence of carpal tunnel syndrome in US working populations: pooled analysis of six prospective studies. *Scand J Work Environ Health*. 2013;39:495-505. https://doi.org/10.5271/sjweh.3351
4. Dilla T, Alexiou D, Chatzitheofilou I, Ayyub R, Lowin J, Norrbacka K. The cost-effectiveness of dulaglutide versus liraglutide for the treatment of type 2 diabetes mellitus in Spain in patients with BMI ≥30 kg/m^2. *J Med Econ*. 2017;20:443-452. https://doi.org/10.1080/13696998.2016.1275651
5. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. *Methods for the Economic Evaluation of Health Care Programmes*. 3rd ed. New York, NY: Oxford University Press; 2005.
6. Fenwick E, Claxton K, Sculpher M. Representing uncertainty: the role of cost-effectiveness acceptability curves. *Health Econ*. 2001;10:779-787. https://doi.org/10.1002/hec.635
7. Fernández-de-las-Peñas C, de la Llave-Rincón AI, Fernández-Carnero J, Cuadrado ML, Arendt-Nielsen L, Pareja JA. Bilateral widespread mechanical pain sensitivity in carpal tunnel syndrome: evidence of central processing in unilateral neuropathy. *Brain*. 2009;132:1472-1479. https://doi.org/10.1093/brain/awp050
8. Fernández-de-las Peñas C, Ortega-Santiago R, de la Llave-Rincón AI, et al. Manual physical therapy versus surgery for carpal tunnel syndrome: a randomized parallel-group trial. *J Pain*. 2015;16:1087-1094. https://doi.org/10.1016/j.jpain.2015.07.012
9. Foley M, Silverstein B. The long-term burden of work-related carpal tunnel syndrome relative to upper-extremity fractures and dermatitis in Washington State. *Am J Ind Med*. 2015;58:1255-1269. https://doi.org/10.1002/ajim.22540
10. Foley M, Silverstein B, Polissar N. The economic burden of carpal tunnel syndrome: long-term earnings of CTS claimants in Washington State. *Am J Ind Med*. 2007;50:155-172. https://doi.org/10.1002/ajim.20430
11. Goossens ME, Rutten-van Mölken MP, Vlaeyen JW, van der Linden SM. The cost diary: a method to measure direct and indirect costs in cost-effectiveness research. *J Clin Epidemiol*. 2000;53:688-695. https://doi.org/10.1016/S0895-4356(99)00177-8
12. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20:1727-1736. https://doi.org/10.1007/s11136-011-9903-x
13. Huisstede BM, Fridén J, Coert JH, Hoogvliet P, European HANDGUIDE Group. Carpal tunnel syndrome: hand surgeons, hand therapists, and physical medicine and rehabilitation physicians agree on a multidisciplinary treatment guideline—results from the European HANDGUIDE Study. *Arch Phys Med Rehabil*. 2014;95:2253-2263. https://doi.org/10.1016/j.apmr.2014.06.022
14. Instituto Nacional de Estadística. Resultados nacionales (desde el trimestre 1/2008). Available at: http://www.ine.es/jaxiT3/Tabla.htm?t=6030&L=0. Accessed December 4, 2016.
15. Jablecki CK, Andary MT, Floeter MK, et al. Practice parameter: electrodiagnostic studies in carpal tunnel syndrome. Report of the American Association of Electrodiagnostic Medicine, American Academy of Neurology, and the American Academy of Physical Medicine and Rehabilitation. *Neurology*. 2002;58:1589-1592. https://doi.org/10.1212/WNL.58.11.1589

16. Jain NB, Higgins LD, Losina E, Collins J, Blazar PE, Katz JN. Epidemiology of musculoskeletal upper extremity ambulatory surgery in the United States. *BMC Musculoskelet Disord*. 2014;15:4. https://doi.org/10.1186/1471-2474-15-4
17. Jarvik JG, Comstock BA, Kliot M, et al. Surgery versus non-surgical therapy for carpal tunnel syndrome: a randomised parallel-group trial. *Lancet*. 2009;374:1074-1081. https://doi.org/10.1016/S0140-6736(09)61517-8
18. Keeley T, Al-Janabi H, Lorgelly P, Coast J. A qualitative assessment of the content validity of the ICECAP-A and EQ-5D-5L and their appropriateness for use in health research. *PLoS One*. 2013;8:e85287. https://doi.org/10.1371/journal.pone.0085287
19. Korthals-de Bos IB, Gerritsen AA, van Tulder MW, et al. Surgery is more cost-effective than splinting for carpal tunnel syndrome in the Netherlands: results of an economic evaluation alongside a randomized controlled trial. *BMC Musculoskelet Disord*. 2006;7:86. https://doi.org/10.1186/1471-2474-7-86
20. Lorgelly PK, Dias JJ, Bradley MJ, Burke FD. Carpal tunnel syndrome, the search for a cost-effective surgical intervention: a randomised controlled trial. *Ann R Coll Surg Engl*. 2005;87:36-40.
21. Louw A, Nijs J, Puentedura EJ. A clinical perspective on a pain neuroscience education approach to manual therapy. *J Man Manip Ther*. 2017;25:160-168. https://doi.org/10.1080/10669817.2017.1323699
22. Manca A, Hawkins N, Sculpher MJ. Estimating mean QALYs in trial-based cost-effectiveness analysis: the importance of controlling for baseline utility. *Health Econ*. 2005;14:487-496. https://doi.org/10.1002/hec.944
23. Newington L, Harris EC, Walker-Bone K. Carpal tunnel syndrome and work. *Best Pract Res Clin Rheumatol*. 2015;29:440-453. https://doi.org/10.1016/j.berh.2015.04.026
24. Nijs J, Van Houdenhove B, Oostendorp RA. Recognition of central sensitization in patients with musculoskeletal pain: application of pain neurophysiology in manual therapy practice. *Man Ther*. 2010;15:135-141. https://doi.org/10.1016/j.math.2009.12.001
25. Page MJ, Massy-Westropp N, O'Connor D, Pitt V. Splinting for carpal tunnel syndrome. *Cochrane Database Syst Rev*. 2012:CD010003. https://doi.org/10.1002/14651858.CD010003
26. Page MJ, O'Connor D, Pitt V, Massy-Westropp N. Exercise and mobilisation interventions for carpal tunnel syndrome. *Cochrane Database Syst Rev*. 2012:CD009899. https://doi.org/10.1002/14651858.CD009899
27. Page MJ, O'Connor D, Pitt V, Massy-Westropp N. Therapeutic ultrasound for carpal tunnel syndrome. *Cochrane Database Syst Rev*. 2013:CD009601. https://doi.org/10.1002/14651858.CD009601.pub2
28. Parot-Schinkel E, Roquelaure Y, Ha C, et al. Factors affecting return to work after carpal tunnel syndrome surgery in a large French cohort. *Arch Phys Med Rehabil*. 2011;92:1863-1869. https://doi.org/10.1016/j.apmr.2011.06.001
29. Per la qual es regulen els supòsits i conceptes facturables i s'aproven els preus públics corresponents als serveis que presta l'Institut Català de la Salut, SLT/42/2012, 24 de febrer (2012).
30. Peters S, Johnston V, Hines S, Ross M, Coppieters M. Prognostic factors for return-to-work following surgery for carpal tunnel syndrome: a systematic review. *JBI Database System Rev Implement Rep*. 2016;14:135-216. https://doi.org/10.11124/JBISRIR-2016-003099
31. Pomerance J, Zurakowski D, Fine I. The cost-effectiveness of nonsurgical versus surgical treatment for carpal tunnel syndrome. *J Hand Surg Am*. 2009;34:1193-1200. https://doi.org/10.1016/j.jhsa.2009.04.034
32. Por la que se fijan los precios públicos de los servicios sanitarios prestados por Centros dependientes del Sistema Sanitario Público de Andalucía, Orden, 14 de octubre (2005).
33. Por la que se fijan los precios públicos por la prestación de los servicios y actividades de naturaleza sanitaria de la Red de Centros de la Comunidad de Madrid, Orden 731/2013, 6 de septiembre del Consejero de Sanidad (2013).
34. Revisión de las condiciones económicas aplicables en el año 2006, a la prestación de servicios de asistencia sanitaria concertada en el ámbito de la Gerencia Regional de Salud (Salud de Castilla y León), Resolución, 6 de septiembre del Presidente de la Gerencia Regional de Salud (2006).
35. Sanati KA, Mansouri M, Macdonald D, Ghafghazi S, Macdonald E, Yadegarfar G. Surgical techniques and return to work following carpal tunnel release: a systematic review and meta-analysis. *J Occup Rehabil*. 2011;21:474-481. https://doi.org/10.1007/s10926-011-9310-8
36. Saw NL, Jones S, Shepstone L, Meyer M, Chapman PG, Logan AM. Early outcome and cost-effectiveness of endoscopic versus open carpal tunnel release: a randomized prospective trial. *J Hand Surg Br*. 2003;28:444-449. https://doi.org/10.1016/S0266-7681(03)00097-4
37. Shi Q, MacDermid JC. Is surgical intervention more effective than non-surgical treatment for carpal tunnel syndrome? A systematic review. *J Orthop Surg Res*. 2011;6:17. https://doi.org/10.1186/1749-799X-6-17
38. Stapleton MJ. Occupation and carpal tunnel syndrome. *ANZ J Surg*. 2006;76:494-496. https://doi.org/10.1111/j.1445-2197.2006.03770.x
39. Tarifas para la facturación de servicios sanitarios y docentes de Osakidetza para el año 2014 (2014).
40. Thiese MS, Gerr F, Hegmann KT, et al. Effects of varying case definition on carpal tunnel syndrome prevalence estimates in a pooled cohort. *Arch Phys Med Rehabil*. 2014;95:2320-2326. https://doi.org/10.1016/j.apmr.2014.08.004
41. Thoma A, Wong VH, Sprague S, Duku E. A cost-utility analysis of open and endoscopic carpal tunnel release. *Can J Plast Surg*. 2006;14:15-20.
42. van Hout B, Janssen MF, Feng YS, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value Health*. 2012;15:708-715. https://doi.org/10.1016/j.jval.2012.02.008
43. Verdugo RJ, Salinas RA, Castillo JL, Cea JG. Surgical versus non-surgical treatment for carpal tunnel syndrome. *Cochrane Database Syst Rev*. 2008:CD001552. https://doi.org/10.1002/14651858.CD001552.pub2
44. Zanette G, Cacciatori C, Tamburin S. Central sensitization in carpal tunnel syndrome with extraterritorial spread of sensory symptoms. *Pain*. 2010;148:227-236. https://doi.org/10.1016/j.pain.2009.10.025
45. Zuo D, Zhou Z, Wang H, et al. Endoscopic versus open carpal tunnel release for idiopathic carpal tunnel syndrome: a meta-analysis of randomized controlled trials. *J Orthop Surg Res*. 2015;10:12. https://doi.org/10.1186/s13018-014-0148-6

VIEW Videos on *JOSPT*'s Website

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

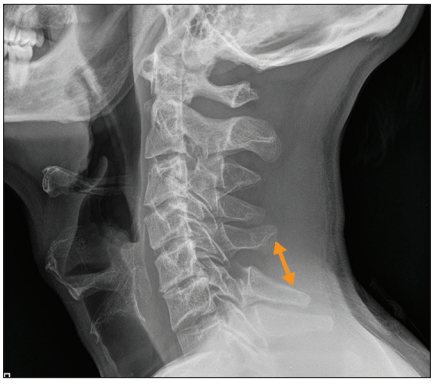


FIGURE 1. Lateral radiograph demonstrating C5-6 interspinous fanning (arrow), suggesting posterior longitudinal ligamentous disruption with probable fracture.

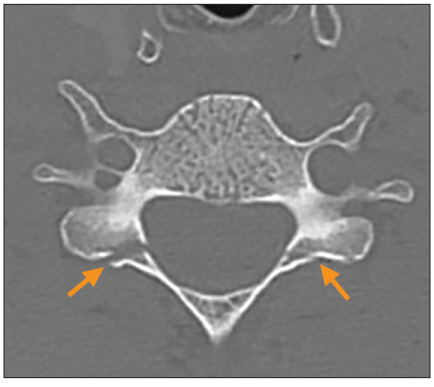


FIGURE 2. Axial computed tomography image demonstrating bilateral C6 lamina fractures.

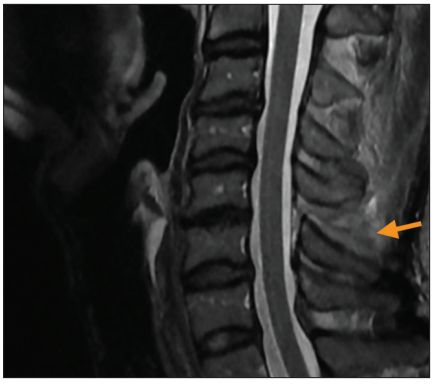


FIGURE 3. Sagittal, T2-weighted magnetic resonance image demonstrating complete disruption of the interspinous and supraspinous ligaments at C5-6.

Cervical Fracture With Posterior Ligamentous Injury While Skydiving

WARREN FLAUTT, PT, DPT, SCS, CSCS, *US Army, Fort Eustis, VA.*
ROBERT ROWLAND, PT, DPT, OCS, RMSK, CSCS, *EXOS, Phoenix, AZ.*
RICHARD B. WESTRICK, PT, DPT, DSc, OCS, SCS, *MGH Institute of Health Professions, Boston, MA; US Army Research Institute of Environmental Medicine, Natick, MA.*

A 46-year-old male soldier reported to an emergency department with severe acute neck pain immediately following a hyperflexion injury from an unusually rapid parachute opening during recreational skydiving. He was evaluated in the emergency department, including radiographs, and released with a diagnosis of "acute neck strain." Six days after the injury, he followed up with a primary care provider, presenting with continued neck pain (rated as 10 on the numeric pain-rating scale), and was sent for a same-day consultation with a physical therapist. The physical therapist evaluation found exquisite tenderness with light palpation over the C5-6 spinous process and inability to actively rotate the neck more than 45° bilaterally. The patient denied neurologic symptoms, and the neurological exam was unremarkable.

Although neck pain is a relatively common complaint following parachute opening shock in skydivers,[2] the patient's levels of severity and irritability were inconsistent with those of typical skydiving or military airborne patients. Due to lack of access to the original radiographs, the dangerous rapid hyperflexion mechanism of injury, midline tenderness, and limited active cervical rotation, the physical therapist ordered cervical spine radiographs,[1,3] which revealed findings consistent with cervical fracture and ligamentous disruption (**FIGURE 1**). The radiologist contacted the physical therapist with the results and recommended computed tomography, magnetic resonance imaging, and a neurosurgery consultation. The computed tomography and magnetic resonance imaging confirmed bilateral C6 lamina fractures and multiple ligamentous disruption at C5-6 (**FIGURES 2** and **3**). Sixteen days following the injury, the patient underwent open reduction internal fixation surgery, with posterior tension band and supraspinous ligament repair (**FIGURE 4**, available at www.jospt.org). One month following surgery, he began physical therapy and a modified strength and conditioning protocol, while wearing a rigid cervical collar for 12 weeks. Eight months following surgery, he returned to full military duties and deployed to combat operations.

J Orthop Sports Phys Ther 2019;49(2):113. doi:10.2519/jospt.2019.8360

References
1. Belot M, Hoens AM, Kennedy C, Li LC. Does every patient require imaging after cervical spine trauma? A knowledge translation project to support evidence-informed practice for physiotherapists. *Physiother Can*. 2017;69:280-289. https://doi.org/10.3138/ptc.2016-32
2. Nilsson J, Fridén C, Burén V, Westman A, Lindholm P, Äng BO. Musculoskeletal pain and related risks in skydivers: a population-based survey. *Aviat Space Environ Med*. 2013;84:1034-1040. https://doi.org/10.3357/ASEM.3570.2013
3. Stiell IG, Wells GA, Vandemheen KL, et al. The Canadian C-spine rule for radiography in alert and stable trauma patients. *JAMA*. 2001;286:1841-1848. https://doi.org/10.1001/jama.286.15.1841

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.

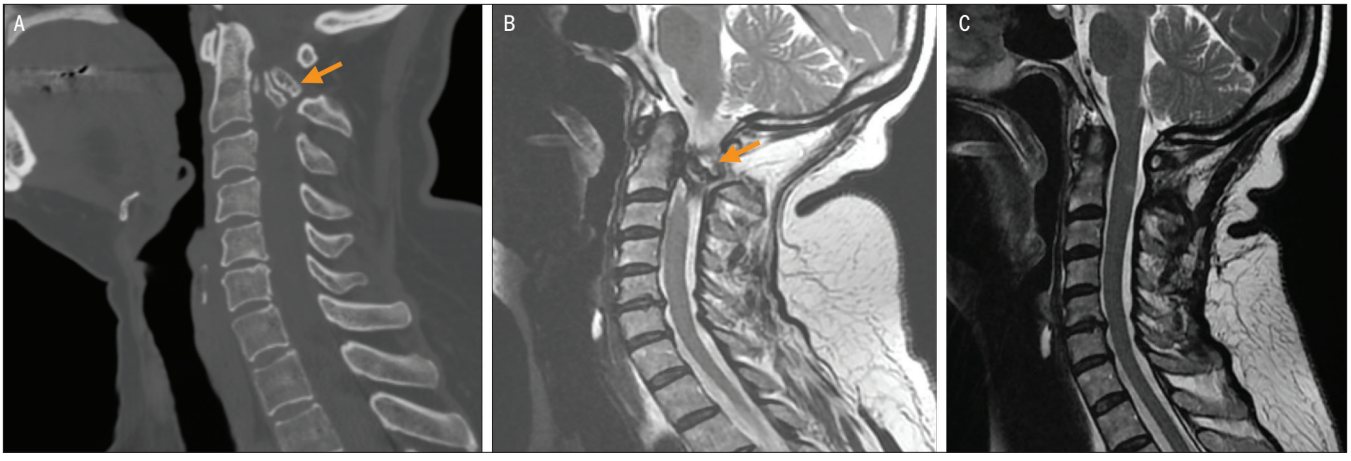


FIGURE 1. Sagittal plane images. (A) Cervical computed tomography showing intracanal stenosis by exostosis at the C1-2 level (arrow). The exostosis came off C2. (B) T2-weighted magnetic resonance imaging showing spinal cord compression. (C) Postoperative, T2-weighted magnetic resonance imaging 6 months after surgery showing small foci of residual high T2 signal in the spinal cord.

Cervical Cord Compression by Exostosis

MICHAEL GRELAT, MD, *Department of Neurosurgery, Dijon University Hospital, Dijon, France.*
CARLOS GIMENEZ, MD, *Department of Neurosurgery, Dijon University Hospital, Dijon, France.*
RACHID MADKOURI, MD, *Department of Neurosurgery, Dijon University Hospital, Dijon, France.*

A 51-YEAR-OLD, LEFT HAND–DOMINANT woman presented to physical therapy with complaints of weakness in her left arm, progressive numbness in both hands, and mild progressive neck pain radiating into the left upper arm. She reported that her condition had started after playing in an amateur tennis tournament 4 weeks prior and progressed to inability to play tennis.

Examination findings were an absence of proprioception in the left arm (wrist and elbow), impaired light touch in all the fingers of her left hand, and inability to discriminate pain and temperature in the right upper extremity and lower extremity consistent with Brown-Sequard syndrome. The patient had no bowel or bladder problems. The patient's pain increased with active bilateral cervical spine rotation. No significant upper extremity strength deficits were found. She reported no personal or familial history of neurological disorders.

The patient was referred to her physician, who ordered magnetic resonance imaging of the spine, which showed a bony exostosis at C1-2 with myelopathy (**FIGURE 1**; **FIGURE 2**, available at www.jospt.org). The patient was referred to a neurosurgical center, where electrophysiological testing indicated normal nerve conduction velocity but detected a reduced evoked potential amplitude in both legs, with no abnormalities of evoked potential amplitudes in her upper extremities. Two weeks after her physical therapy evaluation, the patient underwent surgical excision of the exostosis via a posterior-approach C1-2 laminectomy without fusion.

One week post surgery, the patient reported improved cervical and left upper arm pain. At 3 months post surgery, this patient had a complete recovery from her sensory loss and her weakness. These rare spinal exostoses have been documented in patients with a history of hereditary multiple exostosis.[1] Immediate referral to initiate imaging, in the presence of an abnormal neurologic examination, promoted timely intervention and minimized the potential adverse consequences of a compressive myelopathy.[2] *J Orthop Sports Phys Ther 2019;49(2):112. doi:10.2519/jospt.2019.7942*

References
1. Kamiya Y, Horii E, Sakai Y, Inoue H. Cervical cord compression in pediatrics with hereditary multiple exostoses: a report of two cases and review of the literature. *J Pediatr Orthop B*. 2016;25:267-270. https://doi.org/10.1097/BPB.0000000000000220
2. Lotfinia I, Vahedi P, Tubbs RS, Ghavame M, Meshkini A. Neurological manifestations, imaging characteristics, and surgical outcome of intraspinal osteochondroma. *J Neurosurg Spine*. 2010;12:474-489. https://doi.org/10.3171/2009.11.SPINE0980

STEVEN J. KAMPER, PhD[1]

Fundamentals of Measurement: Linking Evidence to Practice

J Orthop Sports Phys Ther 2019;49(2):114-115. doi:10.2519/jospt.2019.0701

Measurement is fundamental to science, which means that physical therapists must engage with measurement if the profession wishes to call itself scientific. However, while it is simple enough to agree that measurement is a good idea, there is more to measurement than meets the eye.

From the researcher's point of view, the reason for measuring various demographic, personal, and clinical factors is self-evident: scores on the measures answer the research question. In the clinic, physical therapists collect measurements (often informally) from the beginning of the clinical encounter, whether by asking a question like, "What is bothering you today?" or by observing the way the patient gets out of a chair and walks. This information guides further assessment and management. Clinicians and researchers both face similar issues when it comes to interpreting the measures they collect. To do this accurately, understanding some basic concepts about measurement is necessary.

Constructs and Measures

A *construct* is what you are interested in measuring. A *measure* (sometimes called a tool or an instrument) is how the construct is measured. For example, you may be interested in the construct of "disability" in a patient with shoulder pain, and choose to use the Disabilities of the Arm, Shoulder and Hand questionnaire as the measure. There may be several different measures for the same construct; for example, the Shoulder Pain and Disability Index and the Shoulder Disability Questionnaire are also measures of disability in people with shoulder pain (**FIGURE**).

Purpose

There are many reasons to measure something, and any number of available measures. *Screening* measures are designed to estimate how likely it is that a healthy person will have a certain condition in the future, and whether further investigations (eg, screening for cardiovascular risk) should be pursued. *Diagnostic* measures (tests) are designed to determine whether someone does or does not have a certain condition, an example being the Lachman test for anterior cruciate ligament rupture. *Prognostic* tools are designed to help predict whether or when a patient will recover, such as the Örebro Musculoskeletal Pain Questionnaire. *Treatment-based classification* tools are designed to direct a patient toward a certain type of treatment, such as the STarT Back tool. *Outcome* measures are designed to track the level or presence of a symptom, function, or disease marker, for example, the Patient-Specific Functional Scale. Some measures might serve several purposes; for example, pain intensity measured on a numeric rating scale may form part of a diagnostic test, inform likely prognosis, and be tracked over time as an outcome measure.

Subjective and Objective Measures

A common mistake is to call measures rated by the patient "subjective" and those rated by an observer "objective." *Subjectivity* refers to the extent of personal judgment involved in taking a measure, and the personal judgment could be on behalf of the patient or the observer; for example, a physical therapist's rating of the amount of swelling is also subjective. Further, it is better to consider measures as more or less subjective—along a hypothetical continuum—rather than wholly subjective or objective. Another

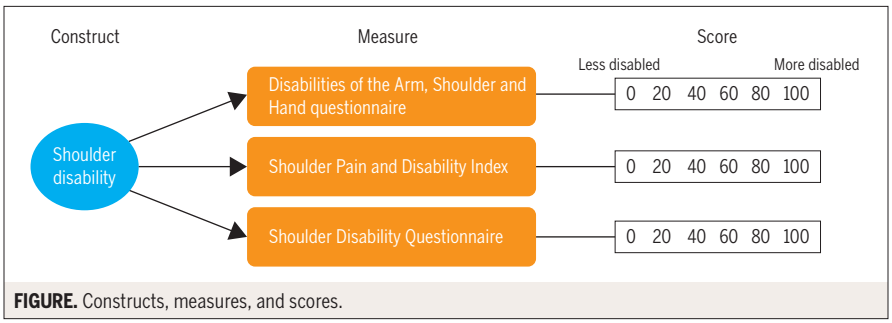


FIGURE. Constructs, measures, and scores.

[1]School of Public Health, University of Sydney, Camperdown, Australia; Centre for Pain, Health and Lifestyle, Australia.

mistake is to automatically consider more objective measures as being superior to more subjective measures, based on the assumption that objective measures are more reliable and valid. Research has shown that reliability and validity do not depend on how objective or subjective a measure is. Finally, it is important to recognize that some constructs are inherently subjective, such as a person's pain experience, and in such cases a more subjective measure will be the most valid.

Methods of Measurement

There are many ways of collecting measurements, and some of the most common types are listed below:

- *Patient-reported measures*. As the name implies, these are ratings provided by the patient. This might involve filling out a written questionnaire or answering questions verbally. Patient-reported measures are commonly used to rate symptom severity, the impacts of a condition, or to measure psychological constructs like quality of life, depression, and self-efficacy.
- *Observer-rated measures*. These measures involve observations made by the clinician. They may include physical capacity measures like strength or range of motion, movement quality, or the ability to perform particular tasks. They may also include observations such as the existence of scoliosis, muscle activation levels, gait characteristics, and the results of clinical tests. Note that many of these measures assess not only physical performance, but also the motivation of the patient. Note that these meausres also involve subjective judgment of the clinician.
- *Scans, images, tests,* and *monitoring devices*. These may be used for screening, diagnosis, or measurement of constructs like habitual physical activity. Even though these types of measures are typically toward the more objective end of the spectrum, their fit for purpose, reliability, and validity should not be automatically assumed. For example, interpretation of scans and images is highly subjective, and findings may not have important functional consequences.
- *Administrative data*. These are most commonly used in research, and include metrics such as hospital attendance, work absence, insurance claim data, and death.

Patient Relevance

Another aspect of interpreting measures is determining patient relevance. This involves judgment as to how important the outcome construct is to your patient. For example, a study might conclude that an intervention is effective because it improves strength, but strength itself may not matter too much to a patient unless it translates to ability to perform important tasks.

Conclusion

The options for measurement are limited only by the imagination, but using poor or inappropriate measures has critical consequences for both clinical practice and research. When reading research, it is important to check that the construct matches the study question. From a clinical perspective, being able to clearly articulate why you want to measure a certain construct is key. Having decided on the appropriate construct, you need to select the best measure. This involves consideration of reliability and validity, which will be the subject of the next Evidence in Practice article.

JUSTIN M. LOSCIALE, DPT, CSCS[1] • **RACHAEL M. ZDEB,** DPT, CSCS, USAW-L1SP[2] • **LEILA LEDBETTER,** MLIS[3]
MICHAEL P. REIMAN, PT, PhD, ATC[4] • **TIMOTHY C. SELL,** PT, PhD[4,5]

The Association Between Passing Return-to-Sport Criteria and Second Anterior Cruciate Ligament Injury Risk: A Systematic Review With Meta-analysis

BACKGROUND: There is no consensus on the components of return-to-sport (RTS) testing following anterior cruciate ligament (ACL) reconstruction or whether passing RTS criteria can reduce a patient's risk of reinjury.

OBJECTIVES: To determine whether impartial, criteria-based RTS decisions are associated with less risk of a second ACL injury (either graft failure or contralateral ACL injury).

METHODS: In this systematic review with meta-analysis, the authors conducted an electronic literature search in PubMed/MEDLINE, Embase, CINAHL, SPORTDiscus, and ProQuest Dissertations and Theses Global using database-specific vocabulary related to ACL reconstruction and return to sport. Individual study quality was assessed using the modified Downs and Black checklist, and overall quality of evidence was determined with the Grading of Recommendations Assessment, Development and Evaluation scale. Pooled risk difference (passed versus failed RTS criteria), injury incidence proportion, and the diagnostic accuracy of each RTS criterion were calculated.

RESULTS: Four studies met the selection criteria. Overall, 42.7% (95% confidence interval [CI]: 18%, 69%) of patients passed RTS criteria, and 14.4% (95% CI: 8%, 21%) of those who passed experienced a second ACL injury (graft rupture or contralateral ACL injury). There was a nonsignificant 3% reduced risk of a second ACL injury after passing RTS criteria (risk difference, −3%; 95% CI: −16%, 10%; I^2 = 74%, P = .610). The evidence rating of the Grading of Recommendations Assessment, Development and Evaluation scale was "very low quality," due to imprecision and heterogeneity of the pooled risk difference estimate.

CONCLUSION: Passing RTS criteria did not show a statistically significant association with risk of a second ACL injury. The quality-of-evidence rating prevents a definitive conclusion on this question and indicates an opportunity for future research.

LEVEL OF EVIDENCE: Prognosis, Level 2a−.
J Orthop Sports Phys Ther 2019;49(2):43-54. Epub 30 Nov 2018. doi:10.2519/jospt.2019.8190

KEY WORDS: *functional testing, knee, reinjury risk, return to play*

Anterior cruciate ligament (ACL) injury is a common knee injury sustained by athletes.[6,31,46,60,83] Patients seeking return to activity commonly undergo anterior cruciate ligament reconstruction (ACLR) to re-establish mechanical knee stability.[18] The impact of ACL injury includes time away from activity; lifelong financial, socioeconomic, and emotional burdens; psychological stress; disability; and the development of osteoarthritis.[1,6,10-12,40,57,83]

The risk of a second ACL injury (either ACL graft failure or a contralateral ACL injury) is a significant concern for those who return to sport.[23,33-36,39,51,66,67,69,83] The incidence of second ACL injury ranges from 3% to 37% and depends on such factors as age, physical demands of the sport, and competition level.[23,33-36,39,51,66,67,69,83] Wiggins et al[83] determined that second ACL injury incidence is 23% in individuals younger than 25 years of age; however, they did not require studies to use objective return-to-sport (RTS) criteria.[83] Patients younger than 20 years of age have a second ACL injury incidence of approximately 1 in 3,[14,42,81] with ACL graft reinjury and native contralateral ACL injuries showing similar incidences.[83]

A second ACL injury tends to occur within the first 6 months to 2 years following return to sport.[23,35,37] Available literature demonstrates that there are deficits in strength, landing kinematics, proprioception, psychological readiness, and perception of knee function that persist at 2 years following ACLR

[1]The Ohio State University Wexner Medical Center, Columbus, OH. [2]Specialists in Sport and Orthopedic Rehabilitation, Overland Park, KS. [3]Duke University Medical Center Library, Durham, NC. [4]Department of Orthopaedic Surgery, Duke University, Durham, NC. [5]Michael W. Krzyzewski Human Performance Laboratory, Duke Sports Medicine, James R. Urbaniak Sports Sciences Institute, Duke University Health System, Duke University, Durham, NC. This study did not require Institutional Review Board approval or public trials registry. There were no funding sources. 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 Justin M. Losciale, The Ohio State University Wexner Medical Center, 410 West 10th Avenue, Columbus, OH 43210. E-mail: jlosciale8@outlook.com

and may continue for up to 20 years after surgery.[5,31,38,46,52,54,55,61-64,72,73,77] Psychological readiness and perception of knee function have garnered increasing attention as potentially vital components when determining readiness to return to sport.[2,3,13,53,57,79] Visual motor processing compensations and central nervous system connectivity alterations following ACL injury may predispose patients to abnormal biomechanics and increase ACL injury risk.[15,24,25]

Successful RTS criteria should reduce the risk of a second ACL injury. However, despite substantial research,[16,23,33,35,46,69,74,78] there is contradictory evidence associating RTS criteria and safe return to sport.[13] Conflicting evidence for the RTS timeline[16,46,68] and optimal decision metrics[23,35,69,76,80,82] confound this issue. In 2011, Barber-Westin and Noyes[9] reported the prevalence of RTS objective-measure utilization in published ACLR outcome studies. Although objective functional assessments had been reported, there were no studies investigating the association of these assessments with reinjury.[9] Additional studies to assess whether resolving lower-limb functional deficits is effective in reducing ACL reinjury were recommended.[9] Investigators further advocated multifactorial RTS criteria, with study of the validity of these criteria to identify safe return to sport.[13,83]

The current review sought to determine the utility of RTS decisions based on objective criteria and to aggregate the data from studies that resulted in decisions to release patients to unrestricted activity based on their performance during objective RTS testing. The primary purpose of this systematic review was to assess whether objective criteria-based RTS decisions are successful in reducing the risk of a second ACL injury. Additional aims were to (1) report and categorize the criteria used for RTS testing, (2) report passing cutoff scores, (3) determine pass/fail incidence, (4) identify ACL graft and native contralateral ACL injury incidences, and (5) assess the diagnostic accuracy of each RTS criterion. The authors hypothesized that successfully meeting RTS criteria would result in decreased risk of a second ACL injury, with the goal that further understanding RTS criteria and associated second ACL injury risk might assist clinicians in determining factors to use in RTS testing to decrease subsequent ACL injuries.

METHODS

Protocol

The present review and meta-analysis followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.[41,71] The PRISMA statement includes a 27-item checklist that is designed for reporting systematic reviews of randomized trials,[41] but the checklist can also be applied to multiple forms of research methodologies.[71] A MeaSurement Tool to Assess systematic Reviews, version 2 (AMSTAR 2) was used to critically appraise this review.[65]

Literature Search

A medical librarian was consulted to perform a computer-assisted literature search in the PubMed/MEDLINE, CINAHL, Embase, SPORTDiscus, and ProQuest Dissertations and Theses Global databases from inception to March 2018, using database-specific vocabulary and key words related to ACLR and return to sport. The search strategies for all databases are listed in **APPENDIX A** (available at www.jospt.org).

Selection Criteria

To be included, studies were required to (1) involve patients recovering from ACLR with any graft type (may have concomitant meniscus lesion and/or medial collateral ligament lesion), (2) include patients who were between the ages of 10 and 50 years, (3) use clearly defined objective criteria to make the RTS decision, (4) determine and report the number of patients who passed versus failed RTS criteria, (5) track patients for subsequent ACL injury following return to sport, and (6) be written in English. A study was excluded when (1) the patients' average age was 9 years or younger or 51 years or older; (2) patients had posterior cruciate ligament, lateral collateral ligament, or bilateral ACL injury; (3) patients had nonsurgical treatment of ACL injury; (4) data were not reported between 6 months and 10 years post surgery; (5) it was a systematic review, meta-analysis, clinical commentary, or abstract; and (6) it was not written in English.

Titles and abstracts were independently screened by 2 authors. Full-text studies were retrieved if the abstract provided insufficient information to establish eligibility or if the study passed initial eligibility screening. Disagreements were resolved by a third author.

Data Extraction

Two authors independently extracted data using identical customized templates. A third author verified data collection prior to statistical analysis. Discrepancies were resolved by consensus between the 2 initial authors. If further data clarification was required, contact with the corresponding author(s) was attempted. The population size, sex, age, and ACLR characteristics were recorded from each study.

Regardless of surgical procedure used, patients were dichotomized into either hamstring graft or bone-patellar tendon-bone graft, as these were the only 2 graft types used in included studies. All grafts were autografts unless otherwise stated. The researchers further extracted the following information: RTS criteria, ipsilateral ACL injury incidence, contralateral ACL injury incidence, total second ACL injury incidence, pass/fail incidence, and second ACL injury incidence for those who did and did not pass RTS criteria.

Assessment of Study Quality and Overall Strength of the Evidence

The Oxford Centre for Evidence-Based Medicine Levels of Evidence tool was used to assess the level of evidence for each study based on research design.[30]

Study quality assessment was performed by utilizing the modified Downs and Black[17] scale, which has been shown to be a reliable assessment for case-control and cohort studies. The highest total score for the modified version is 16, with a stratified score ranking of 12 or greater as high quality, of 10 to 11 as moderate quality, and of 9 or less as low quality.[43,44]

Strength of the evidence included in this review was determined using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) scale, which shows the overall certainty of the evidence for the outcome being reviewed (**APPENDIX B**, available at www.jospt.org).[7,59] The GRADE scale assesses 5 factors concerning risk of bias, inconsistency (calculated heterogeneity), indirectness (evidence addresses review question), imprecision (width of confidence intervals [CIs]), and publication bias.[59] These factors lead to a reported score of high, moderate, low, or very low quality.[59] The GRADE scale was applied to assess the evidence regarding the association between passing RTS criteria and risk of a second ACL injury. Two authors independently reviewed and scored each study, with disagreements settled through discussion and consensus.

Statistical Analysis

Studies were statistically pooled when 2 or more studies examined the same index test. Data from each study were extracted and dichotomized into categorical variables of "pass" for those who successfully met RTS criteria and returned to sport and "fail" for those who did not successfully meet RTS criteria and return to sport. For the purpose of this review, return to sport was defined as the clearance of a patient for full participation in that patient's defined sport or activity without restrictions (training and competition).[4]

A random-effects proportion meta-analysis (weighted for individual study size) using StatsDirect (StatsDirect Ltd, Cambridge, UK) was conducted to determine the following incidence proportions at the 95% CI: overall second ACL injury, patients who passed RTS testing, patients who failed RTS testing, overall second ACL injury (passed RTS criteria versus failed RTS criteria), ACL graft injury, and contralateral ACL injury. Censoring over time was not performed due to lack of standardization of assessment time points across studies.

Risk difference (RD) of a second ACL injury (combined graft and native contralateral ACL) based on "pass" and "fail" status was determined using a random-effects RD meta-analysis.[22] The RD provides an absolute measure of association between the 2 exposure groups (passed versus failed RTS testing) and determines the difference in total amount of injuries sustained between exposure groups, expressed as a percentage.[48] Failing RTS criteria was labeled the "exposed group" and passing RTS criteria was labeled the "unexposed group"; thus, the calculation was RD = cumulative incidence of second ACL injury$_{failed}$ – cumulative incidence of second ACL injury$_{passed}$.

Anterior cruciate ligament graft injury RD and native contralateral ACL injury RD based on "pass" and "fail" status were similarly calculated. The above calculation determines the association of failing RTS criteria and second ACL injury; the association of passing RTS criteria and second ACL injury is the inverse. Pooled estimates at the 95% CI were summarized in forest plots. Statistical analysis and figures were processed and created using Review Manager Version 5.3 (The Nordic Cochrane Centre, Copenhagen, Denmark).

Statistical heterogeneity was determined for all RD calculations to assess variation across studies and as a component of the GRADE scale.[59] Chi-square (test for heterogeneity), tau-square (between-study variance in random-effects meta-analysis), and the I^2 statistic (percentage of variation across studies due to heterogeneity rather than chance) were calculated at the 95% CI. The categorization to rate the level of heterogeneity was the following: I^2 = 0%, no heterogeneity; I^2 = 1% to 25%, low heterogeneity, not important; I^2 = 26% to 50%, moderate heterogeneity; I^2 = 51% to 75%, high heterogeneity, substantial; I^2 = 76% to 100%, considerable heterogeneity.[29,58]

Diagnostic accuracy for each RTS criterion was determined using a 2-by-2 diagnostic test table with 95% CIs, with report of second ACL injury as the reference standard (StatsDirect Ltd). Positive and negative test results, as well as the definitions of true positive, true negative, false positive, and false negative used in the analysis, are reported in **TABLE 1**. Test sensitivity, specificity, likelihood ratios (positive and negative), and the diagnostic odds ratio (DOR) were calculated. Sensitivity refers to the probability that the result of RTS testing will be positive when the outcome (second ACL injury) occurs. Specificity is the probability that the RTS testing result will be negative when the outcome does not occur. The likelihood ratio statistic reflects changes in posttest probability based on test outcome. The DOR determines the ratio of the odds of a second ACL injury in positive tests relative to the odds of a second ACL injury in negative tests.[20] Values

TABLE 1 Test-Result Definitions

	Sustained Second ACL Injury (Positive)	No Second ACL Injury (Negative)
Positive (failing) test	Fail RTS criteria and sustain a second ACL injury (TP)	Fail RTS criteria and do not sustain a second ACL injury (FP)
Negative (passing) test	Pass RTS criteria and sustain a second ACL injury (FN)	Pass RTS criteria and do not sustain a second ACL injury (TN)

Abbreviations: ACL, anterior cruciate ligament; FN, false negative; FP, false positive; RTS, return to sport; TN, true negative; TP, true positive.

range from 0 to infinity, with higher values indicating enhanced discriminatory ability; a value of 1 indicates no discriminatory value and values less than 1 indicate improper test classification.[20]

RESULTS

Study Selection

The literature search identified 2036 potentially eligible titles. Full-text review of 131 studies was performed, with 4 studies (549 patients) meeting inclusion criteria for this review (**FIGURE 1**).[23,35,47,69] Two corresponding authors were contacted for further clarification regarding ACL injuries, based on pass/fail status, and clarification was received from both authors. Narrative summaries of each included study can be found in **APPENDIX C** (available at www.jospt.org).

Level of Evidence, Study Quality, GRADE, and AMSTAR 2

Three studies were rated 2B (cohort studies)[23,35,47] and 1 was rated 3B (case-control study)[69] (**TABLE 2**). There were no disagreements between the authors on study-level rating. All 4 studies were rated 12/16 or greater (high individual study quality), with moderate agreement between reviewers ($\kappa = 0.54 \pm 0.13$) (**TABLE 2**).[23,35,47,69] Full modified Downs and Black[17] scoring is provided in **APPENDIX D** (available at www.jospt.org). The GRADE scale determined that the quality of evidence for the association of passing RTS criteria with overall second ACL injury risk is "very low quality," due to imprecision of the pooled RD estimate and substantial levels of heterogeneity (**APPENDIX E**, available at www.jospt.org).[59] This review met 11 of 16 criteria (69%), according to AMSTAR 2 (**APPENDIX F**, available at www.jospt.org).[65]

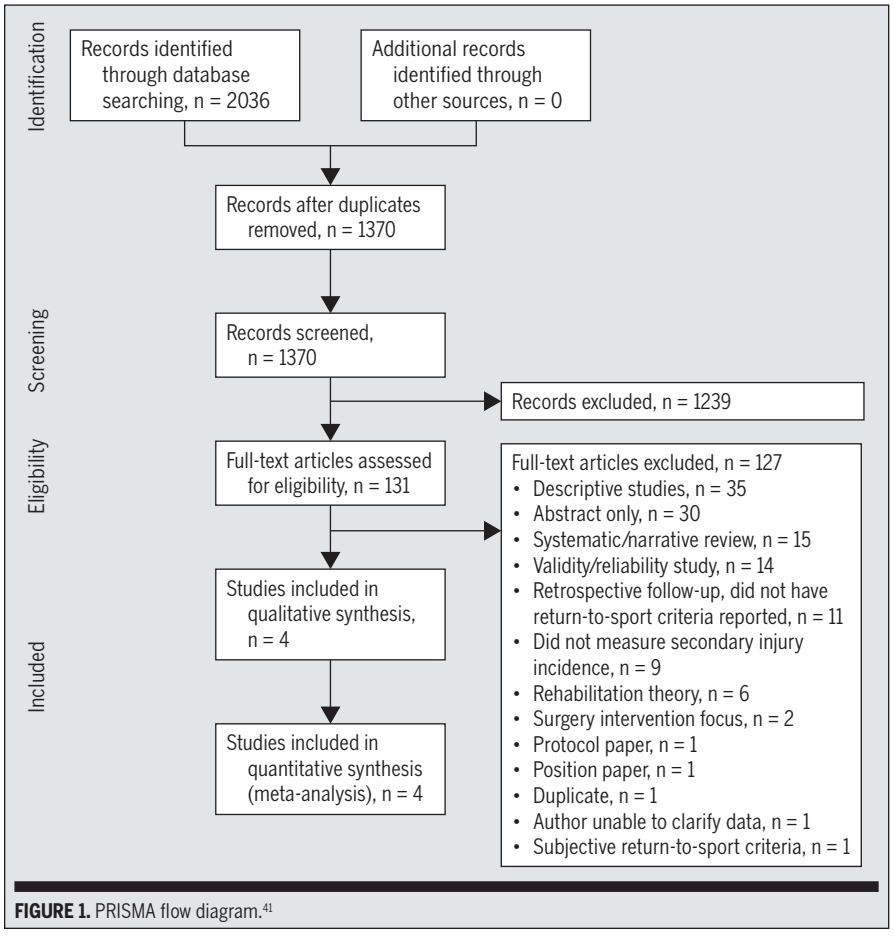


FIGURE 1. PRISMA flow diagram.[41]

The overall confidence in the results is considered moderate.[65]

Incidence Proportion Analysis

The incidence of passing and failing RTS criteria, ACL graft injury (pass versus fail), contralateral ACL injury (pass versus fail), and overall second ACL injury is presented in **TABLE 3**.

Association Between Passing RTS Criteria and Secondary, Graft, and Contralateral ACL Injury Risk

Pooled RD calculations are presented in forest plots (**FIGURES 2** through **4**) at the 95% CI. There was 3% less risk (95% CI: −16%, 10%) of a second ACL injury after passing RTS criteria, with high levels of calculated heterogeneity ($I^2 = 74\%$) (**FIGURE 2**). None of the 3 calculations reached statistical significance.

RTS Criteria

A comprehensive description of each study's RTS criteria and cutoff scores can be found in **TABLE 2**. Between-study comparisons by individual RTS criterion are provided in **TABLE 4**.

Diagnostic Accuracy

The diagnostic accuracy and positive likelihood ratio, negative likelihood ratio, and DOR values of each set of RTS criteria for prediction of a second ACL injury are presented in **TABLE 5**. The RTS criteria tested by Grindem et al[23] demonstrated the best discriminatory ability (DOR = 3.28; 95% CI: 0.40, 154.23), highest sensitivity (0.90; 95% CI: 0.55, 1.00), and the lowest negative likelihood ratio (0.37; 95% CI: 0.06, 1.63), indicating a small decrease in the posttest probability of a second ACL injury after passing RTS criteria.[32]

DISCUSSION

There is an urgent need to develop effective RTS criteria, given the significant risk of a second ACL injury following ACLR.[23,33-36,39,51,66,67,69,83] Previous literature has established the

prevalence of RTS objective-measure utilization[9] and incidence of a second ACL injury, without consideration of RTS testing.[83] The purpose of this review was to examine the association of objective criteria-based RTS decisions with risk of a second ACL injury. The primary finding was a nonsignificant association between passing objective RTS criteria and the risk of a second ACL injury (RD, −3%; $P = .610$), an ACL graft injury (RD, −7%; $P = .140$), and a contralateral ACL injury (RD, 4%; $P = .160$).

The authors of this review elected to calculate the absolute risk of a second ACL injury (RD) compared to a relative measure of association (risk ratio), because absolute risk allows judgment on the clinical relevance of pooled estimates.[48] These results indicate the need for continued research to prospectively examine objective criteria-based RTS decisions. While not statistically significant, there was more risk of a contralateral ACL injury after passing RTS criteria (RD, 4%; $P = .160$). This potentially implies that RTS criteria and comparison metrics may not accurately assess contralateral-limb function and are poor indicators of contralateral ACL injury risk.

This review found higher incidence of ACL graft injury compared to contralateral ACL injury (7.2% versus 5.1%). The ACL graft injury incidence is similar to previously reported values[83]; however, contralateral injury incidence was slightly lower.[83] Most concerning, this review determined that 12% (95% CI: 3%, 26%) of those who failed RTS testing suffered a graft injury, compared to 5.9% (95% CI: 2%, 11%) of patients who passed. Although not statistically significant, there may be a protective association between passing RTS criteria and ACL graft reinjury (RD, −7%; $P = .140$). It is plausible that additional research will demonstrate less risk of an ACL graft injury after passing RTS criteria.

The low number of studies meeting selection criteria and differences in source populations, ages, and competition

TABLE 2 Study Characteristics and RTS Criteria

Study	OCEBM	Downs and Black[17]	Population*	RTS Criteria	Passing Threshold
Kyritsis et al[35]	2B†	12/16‡	n = 158 (all male) Age, 21-22 y Professional Qatari athletes Grafts: HS, n = 108; BPTB, n = 50 Football, n = 105; handball, n = 21; not specified, n = 32	Isokinetic test at 60°/s, 180°/s, and 300°/s Single hop, triple hop, triple crossover hop Running T test Fully complete on-field sport-specific rehabilitation	Quadriceps LSI >90% at 60°/s LSI >90% on all hop tests <11 s on running T test Complete rehabilitation: yes or no
Nawasreh et al[47]	2B†	12/16‡	n = 95 (male, n = 63; female, n = 32) Age, 27.14 ± 10.59 y Level 1 or 2 sport participation Grafts§: BPTB, n = 2; HS, n = 37; allograft, n = 69	Isometric quadriceps strength Single hop, crossover hop, triple hop, 6-m timed hop KOS-ADL Global rating scale of perceived function	QI >90% LSI >90% on all hop tests KOS-ADL >90% Global rating >90%
Grindem et al[23]	2B†	13/16‡	n = 100 (male, n = 46; female, n = 54) Age, 24.3 y; 4.8 mo from injury to surgery Norwegian arm of Delaware-Oslo cohort study Level 1 or 2 sport participation Handball, n = 30; football, n = 53; basketball, n = 6; floorball, n = 11 Level 2 sport, n = 17 Grafts: BPTB, n = 33; HS, n = 67	Isokinetic concentric quadriceps strength at 60°/s Single hop, crossover hop, triple hop, 6-m timed hop KOS-ADL Global rating scale of perceived function	LSI >90% LSI >90% on all hop tests KOS-ADL >90% Global rating >90%
Sousa et al[69]‖	3B¶	14/16‡	n = 223 (male, n = 92; female, n = 131) Isolated ACLR Age: excellent group, 24 ± 12.1 y; delayed group, 27.2 ± 11.7 y Preinjury Tegner: excellent group, 7.2; delayed group, 6.5	Isokinetic quadriceps and hamstring strength at 60°/s and 180°/s Vertical jump, single hop, triple jump	LSI >85% LSI >90% on all hop tests

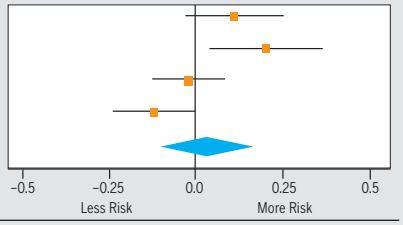
Abbreviations: ACLR, anterior cruciate ligament reconstruction; BPTB, bone-patellar tendon-bone autograft; HS, hamstring autograft; KOS-ADL, Knee Outcome Survey-activities of daily living subscale; LSI, limb symmetry index; OCEBM, Oxford Centre for Evidence-Based Medicine Level of Evidence score; QI, quadriceps index; RTS, return to sport.
**Age values are mean or mean ± SD.*
†Indicates individual cohort study.
‡Score indicates high study methodological quality.
§Only 95 of 108 enrolled patients were analyzed by the study.
‖Only required patients to pass 6 out of 7 RTS criteria.
¶Indicates individual case-control study.

levels could explain the imprecision of pooled estimates and substantial levels of heterogeneity (I^2 = 74%). One study[35] (n = 158) examined competitive athletes (male professional athletes). Higher levels of competition have been shown to increase the baseline risk for ACL injury.[83] Although this review demonstrated "substantial" heterogeneity (I^2 = 51%-75%),[29,58] higher levels were seen in a meta-analysis by Wiggins et al[83] (I^2 = 94%). This finding may be indicative of the inherent heterogeneity of available literature on this population.

Additionally, the time from RTS to reinjury could be an important confounding factor in these estimates. This review did not establish a time point for analysis

Second ACL Injury

Study	Failed Events, n	Failed Total, n	Passed Events, n	Passed Total, n	Weight	Risk Difference IV, Random (95% Confidence Interval)
Grindem et al[23]	9	55	1	18	23.9%	0.11 (−0.04, 0.25)
Kyritsis et al[35]	16	42	21	116	22.1%	0.20 (0.04, 0.36)
Nawasreh et al[47]	3	47	4	48	27.8%	−0.02 (−0.12, 0.09)
Sousa et al[69]	16	171	11	52	26.3%	−0.12 (−0.24, 0.00)
Total*	44	315	37	234	100.0%	0.03 (−0.10, 0.16)

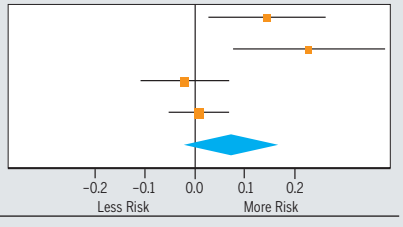


Abbreviations: ACL, anterior cruciate ligament; IV, independent variable.
**Heterogeneity: τ^2 = 0.01, χ^2 = 11.74, df = 3 (P = .008), I^2 = 74%. Test for overall effect: z = 0.51 (P = .61).*

FIGURE 2. Risk difference of a second ACL injury. The pooled effect reflects the association of failing return-to-sport criteria with second ACL injury risk. The pooled estimate describing the association of passing return-to-sport criteria with second ACL injury risk is the inverse of the reported value.

ACL Graft Injury

Study	Failed Events, n	Failed Total, n	Passed Events, n	Passed Total, n	Weight	Risk Difference IV, Random (95% Confidence Interval)
Grindem et al[23]	8	55	0	18	23.2%	0.15 (0.03, 0.26)
Kyritsis et al[35]	14	42	12	116	18.9%	0.23 (0.08, 0.38)
Nawasreh et al[47]	2	47	3	48	27.1%	−0.02 (−0.11, 0.07)
Sousa et al[69]	8	171	2	52	30.9%	0.01 (−0.05, 0.07)
Total*	32	315	17	234	100.0%	0.07 (−0.02, 0.17)

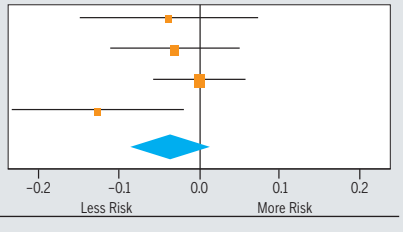


Abbreviations: ACL, anterior cruciate ligament; IV, independent variable.
**Heterogeneity: τ^2 = 0.01, χ^2 = 11.72, df = 3 (P = .008), I^2 = 74%. Test for overall effect: z = 1.49 (P = .14).*

FIGURE 3. Risk difference of ACL graft injury. The pooled effect reflects the association of failing return-to-sport criteria with second ACL injury risk. The pooled estimate describing the association of passing return-to-sport criteria with second ACL injury risk is the inverse of the reported value.

Contralateral ACL Injury

Study	Failed Events, n	Failed Total, n	Passed Events, n	Passed Total, n	Weight	Risk Difference IV, Random (95% Confidence Interval)
Grindem et al[23]	1	55	1	18	16.2%	−0.04 (−0.15, 0.07)
Kyritsis et al[35]	2	42	9	116	26.5%	−0.03 (−0.11, 0.05)
Nawasreh et al[47]	1	47	1	48	40.2%	0.00 (−0.06, 0.06)
Sousa et al[69]	8	171	9	52	17.2%	−0.13 (−0.23, −0.02)
Total*	12	315	20	234	100.0%	−0.04 (−0.09, 0.01)



Abbreviations: ACL, anterior cruciate ligament; IV, independent variable.
**Heterogeneity: τ^2 = 0.00, χ^2 = 4.16, df = 3 (P = .24), I^2 = 28%. Test for overall effect: z = 1.40 (P = .16).*

FIGURE 4. Risk difference of contralateral ACL injury. The pooled effect reflects the association of failing return-to-sport criteria with second ACL injury risk. The pooled estimate describing the association of passing return-to-sport criteria with second ACL injury risk is the inverse of the reported value.

because follow-up times were variable between included studies (range, 24-68 months),[23,35,47,69] potentially influencing RTS estimates. This information may provide insight into how duration of time (or how many exposures to risk) to pass RTS criteria reduces risk following ACLR.[49]

To the researchers' knowledge, there has been no direct comparison between various RTS criteria and their ability to determine the probability of a second ACL injury. Each included study's RTS criteria demonstrated varying discriminatory ability (DOR range, 0.39-3.28) (**TABLE 5**). Grindem et al[23] demonstrated the highest sensitivity (0.90), lowest negative likelihood ratio (0.37), and highest DOR (3.28); however, they had fewer patients for follow-up compared to the other 3 included studies (n = 73).[23,35,47,69] Two studies[47,69] had a DOR of less than 1, indicating no discriminate capabilities.[20] Based on DOR results, the RTS criteria reported by Grindem et al[23] warrant further investigation and potential clinical use. Grindem et al[23] used strict cutoff scores that led to 75.3% of patients failing RTS testing.[23]

TABLE 3 Incidence Proportions by Study and Statistically Pooled Results

Study	Pass/Fail	ACL Graft Injury	Contralateral ACL Injury	Second ACL Injury*	Overall Second ACL Injury†
Kyritsis et al[35]					37/158, 23.4%
Pass	116/158, 73.4%	12/116, 10.3%	9/116, 7.8%	21/116, 18.1%	
Fail	42/158, 26.6%	14/42, 33.3%	2/42, 4.8%	16/42, 38.1%	
Nawasreh et al[47]					7/95, 7.4%
Pass	48/95, 50.5%	3/48, 6.3%	1/48, 2.1%	4/48, 8.3%	
Fail	47/95, 49.5%	2/47, 4.3%	1/47, 2.1%	3/47, 6.4%	
Grindem et al[23]					10/73, 13.7%
Pass	18/73, 24.7%	0/18, 0%	1/18, 5.6%	1/18, 5.6%	
Fail	55/73, 75.3%	8/55, 14.5%	1/55, 1.8%	9/55, 16.4%	
Sousa et al[69]					27/223, 12.1%
Pass	52/223, 23.3%	2/52, 3.9%	9/52, 17.3%	11/52, 21.2%	
Fail	171/223, 76.7%	8/171, 4.7%	8/171, 4.7%	16/171, 9.4%	
Pooled total					
Pass	42.7% (95% CI: 18%, 69%)	5.9% (95% CI: 2%, 11%)	7.5% (95% CI: 1%, 17%)	14.4% (95% CI: 8%, 21%)	
Fail	57.3% (95% CI: 31%, 82%)	12% (95% CI: 3%, 26%)	3.5% (95% CI: 2%, 6%)	15.6% (95% CI: 6%, 29%)	
Pooled incidence‡		7.2% (95% CI: 4%, 11%)	5.1% (95% CI: 3%, 8%)	13.9% (95% CI: 8%, 21%)	

Abbreviations: ACL, anterior cruciate ligament; CI, confidence interval.
**Second ACL injury = ACL injury$_{graft}$ + ACL injury$_{contralateral}$.*
†Overall second ACL injury = second ACL injury$_{pass}$ + second ACL injury$_{fail}$.
‡Pooled incidence = pooled total$_{pass}$ + pooled total$_{fail}$.

TABLE 4 Components of Return-to-Sport Criteria by Study

Study	Isokinetic Testing*	Isometric Testing†	Single Hop	Triple Hop	Triple Crossover Hop	6-m Timed Hop	Vertical Jump	Triple Jump	T Test	Sport-Specific Rehabilitation	KOS-ADL	GRSPF
Kyritsis et al[35]	✓‡	–	✓	✓	✓	–	–	–	✓	✓	–	–
Grindem et al[23]	✓§	–	✓	✓	✓	✓	–	–	–	–	✓	✓
Nawasreh et al[47]	–	✓	✓	✓	✓	✓	–	–	–	–	✓	✓
Sousa et al[69]	✓‖	–	✓	–	–	–	✓	✓	–	–	–	–

Abbreviations: ✓, used as return-to-sport criterion; –, not used as return-to-sport criterion; GRSPF, global rating scale of perceived function; KOS-ADL, Knee Outcome Survey-activities of daily living subscale.
**Isokinetic quadriceps and/or hamstring strength testing at 60°/s, 180°/s, and/or 300°/s.*
†Isometric strength testing via maximal voluntary isometric contraction of the quadriceps.
‡Isokinetic quadriceps and hamstring strength testing at 60°/s, 180°/s, and 300°/s.
§Isokinetic quadriceps strength testing at 60°/s.
‖Isokinetic quadriceps and hamstring strength testing at 60°/s and 180°/s.

Researchers have suggested that optimal cutoff scores (isokinetic strength and hop tests) should be 90% or greater to 100% on a limb symmetry index (LSI) for competitive athletes; however, no included study used a passing LSI of greater than 90%.[16,76] This could have diminished test sensitivity and the DOR.

This review examined the difference in risk and the probability of a second ACL injury after passing all RTS criteria, but not the impact of partially meeting RTS criteria on second ACL injury risk. It may be valuable to further investigate how partially meeting RTS criteria (ie, passing 5 of 7 criteria or failing a specific test) may alter risk. Two included studies[23,35] identified independent risk factors for an ACL graft injury (more symmetrical quadriceps strength and improved hamstring-to-quadriceps strength ratio).

Time from surgery to return to sport may be a key moderator of second ACL injury risk, based on evidence from 2 included studies.[23,69] There is no consensus on the optimal timing of return to sport. Grindem et al[23] determined that for every month return to sport was delayed (up to 9 months), the incidence of any knee reinjury was reduced by 51%. Returning to sport at 6 months was shown to be an independent predictor of contralateral ACL injury.[69] Delayed return to sport was shown to be protective of second ACL injury in a pediatric population.[14] Patients in competitive sports who were allowed to self-select when to return to sport following ACLR chose to return to preinjury activity levels at approximately 8 months.[33]

It has been proposed that biological healing, neuromuscular control, and proprioceptive and strength recovery require up to 2 years to normalize following ACLR.[46] The timeline of cortical dysfunction recovery following ACLR is unknown.[15] A window of highest incidence of second ACL injury, seen between 6 months and 2 years post ACLR, has been described, coinciding with the described healing phase.[37,46] Time from surgery to return to sport is likely a surrogate measure of multiple variables, including increased time addressing strength and kinematic deficits, recovering proprioceptive loss, additional tissue healing, and overcoming any psychological or cortical impairments.

Unresolved ipsilateral deficits can place additional demands, potentially above the physiologic capacity, on the contralateral limb. This is a possible cause of the greater risk for a contralateral ACL injury found in this review. Time from surgery should be an important consideration in RTS decision making, based on available evidence. Returning to sport prior to 9 months following ACLR could be detrimental to the patient.

Twelve different RTS tests were reported in this review. The most common criterion (100%) was the single-leg hop test.[23,35,47,69] Although single-leg hop LSI score was a key component in RTS testing, its ability to alter second ACL injury risk and predict future knee injury has not been established.[8,9,16,21,26,76] Two included studies found that no hop test was predictive of a second ACL injury, even though passing LSI scores were in accordance with published recommendations.[23,35,69,76] Recent evidence demonstrates that using percentage of body height as a normalized hop distance has some predictive validity.[53] These comparison metrics (LSI versus percent of body height) have not been directly compared to determine enhanced predictive ability. Additionally, quantitative measurement (distance and LSI) may not provide enough information to optimize test sensitivity.

Valgus loading[27,28,56] and altered postural stability[54,56] during landing tasks have been shown to predict future injury. Adolescent athletes following ACLR met hop test symmetry by hopping a shorter distance on the contralateral limb and demonstrated lower knee energy absorption compared to controls.[84] Assessing kinematic variables during hop testing may be warranted for RTS testing based on limitations with current comparison metrics (LSI).[45] Validated clinical evaluations of jump-landing mechanics have been previously reported and may provide valuable information when making RTS decisions.[10,16,19,21,50]

Concentric isokinetic quadriceps strength testing at 60°/s was the second most common RTS criterion (75%).[23,35,69] Cutoff scores differed across included studies. Passing scores have been variably reported in the literature (85%-100% on the LSI), likely contributing to observed differences.[16,23,35,69,76] Thomeé et al[77] determined that individuals 12 months post ACLR have significant difficulty achieving 90% or greater on the LSI for strength tests. Evidence suggests that patients may

TABLE 5 Diagnostic Accuracy of Return-to-Sport Criteria*

Study	Sensitivity	Specificity	+LR	–LR	DOR
Kyritsis et al[35]	0.43 (0.27, 0.61)	0.79 (0.70, 0.85)	2.01 (1.20, 3.26)	0.73 (0.52, 0.93)	2.76 (1.17, 6.50)
Grindem et al[23]	0.90 (0.55, 1.00)	0.27 (0.17, 0.40)	1.23 (0.80, 1.52)	0.37 (0.06, 1.63)	3.28 (0.40, 154.23)
Nawasreh et al[47]	0.43 (0.10, 0.82)	0.50 (0.39, 0.61)	0.86 (0.31, 1.60)	1.14 (0.49, 1.84)	0.75 (0.10, 4.73)
Sousa et al[69]	0.59 (0.39, 0.78)	0.21 (0.15, 0.27)	0.75 (0.51, 0.97)	1.95 (1.10, 3.15)	0.39 (0.15, 1.00)

Abbreviations: DOR, diagnostic odds ratio; –LR, negative likelihood ratio; +LR, positive likelihood ratio.
**Values are mean (95% confidence interval).*

have difficulty meeting higher LSI scores, even 2 years following ACLR.[16,23,46,47,76,77]

Nawasreh et al[47] measured isometric quadriceps strength and demonstrated that patients who did not meet RTS criteria (including a quadriceps index of 90% or greater) at 6 months following ACLR were not able to achieve the quadriceps index score at 12 or 24 months. Isokinetic LSI comparisons were shown to overestimate muscle function.[82] Wellsandt et al[82] described the estimated preinjury capacity statistic and found that only 28.6% of their cohort met a score of 90% of the estimated preinjury capacity at 6 months, versus 57.1% who scored 90% on the LSI. Further research is warranted to understand the utility of estimated preinjury capacity measurement in this population.

It has been reported that persistent psychological impairments reduce the capability of an individual to successfully return to preinjury competition level.[5,13] Two included studies (50%) used patient-reported outcome measures as part of their RTS criteria.[23,47] Qualitative studies suggest that individuals following ACLR are most concerned with fear of reinjury.[5,57,79] Psychological measurements are advocated as essential aspects of a biopsychosocial approach to RTS decisions.[3,5,16,38,75,78] The gold standard assessment tool and necessary "amount of confidence" are currently unknown. This is best highlighted by a recent study,[53] which determined that high knee confidence was included in a "high ACL reinjury risk" profile. Further research to identify the appropriate assessment tool and cutoff scores is warranted.

Limitations

This systematic review identified a limited number of studies (n = 4), reducing the total sample size and number of ACL injury events, which potentially impacted the ability to reach statistical significance in RD analyses. High levels of heterogeneity (I^2 = 74%), likely attributable to clinical diversity[22] between studies (demographics, competition levels, RTS criteria, RTS time frames), decreased the value of pooled estimates and might have affected the ability to reach statistical significance. Heterogeneity was considered during GRADE assessment, and the results of the meta-analysis were discussed in the context of the confidence of the pooled estimate. Meta-analysis was reported despite high levels of heterogeneity, due to lack of prospective review protocol registration and to eliminate concerns of reporting bias.

The variability of return to sport was reported in included studies (potential for reporting bias), none of which matched the 2016 consensus statement on RTS definition,[4] due to similar publication dates for 3 included studies. Further, the competition level of patients who returned to sport was not controlled for and could have significantly affected risk of reinjury. Follow-up times were also different among studies. Longer follow-up times after return to sport could result in increased exposures to higher-risk activities. Due to variation in reporting, this review did not establish a follow-up duration cut point to perform the analysis, possibly confounding the results by increasing the incidence of a second ACL injury. The risk of a second ACL injury was not determined based on graft type or sex due to inconsistent categorization of patients in the included studies.

CONCLUSION

This review demonstrated that current objective criteria-based RTS decisions did not show an association with the risk of a second ACL injury. This conclusion was based on a very low quality of evidence due to observed heterogeneity and imprecision between the included studies. This review cannot confidently conclude that there is no association between passing objective RTS criteria and risk of a second ACL injury. Studies included in this review demonstrated clinically important findings regarding RTS decisions that warrant attention. Additional high-quality studies are encouraged and may alter these conclusions. ◉

KEY POINTS

FINDINGS: This review demonstrates that there are few studies examining criteria-based return-to-sport (RTS) decisions following anterior cruciate ligament reconstruction and that RTS criteria may be suboptimal at reducing the risk of a second anterior cruciate ligament injury in a heterogeneous population.

IMPLICATIONS: The current evidence indicates that there may be a need for continued research to determine optimal RTS criteria.

CAUTION: This review was based on a limited number of studies and a very low quality of evidence when examining different subgroups after anterior cruciate ligament reconstruction, thus preventing definitive conclusions on this topic.

REFERENCES

1. Ajuied A, Wong F, Smith C, et al. Anterior cruciate ligament injury and radiologic progression of knee osteoarthritis: a systematic review and meta-analysis. *Am J Sports Med*. 2014;42:2242-2252. https://doi.org/10.1177/0363546513508376
2. Ardern C, Taylor N, Feller J, Webster K. Does fear of re-injury impact on sports participation in the medium-term after anterior cruciate ligament reconstruction surgery? [abstract]. *J Sci Med Sport*. 2011;14 suppl 1:e98-e99. https://doi.org/10.1016/j.jsams.2011.11.206
3. Ardern C, Taylor N, Feller J, Whitehead T, Webster K. The psychological responses of athletes before, and early after recovery from ACL reconstruction predict returning to the pre-injury level sport [abstract]. *J Sci Med Sport*. 2013;16 suppl 1:e90. https://doi.org/10.1016/j.jsams.2013.10.216
4. Ardern CL, Glasgow P, Schneiders A, et al. 2016 consensus statement on return to sport from the First World Congress in Sports Physical Therapy, Bern. *Br J Sports Med*. 2016;50:853-864. https://doi.org/10.1136/bjsports-2016-096278
5. Ardern CL, Österberg A, Tagesson S, Gauffin H, Webster KE, Kvist J. The impact of psychological readiness to return to sport and recreational activities after anterior cruciate ligament reconstruction. *Br J Sports Med*. 2014;48:1613-1619. https://doi.org/10.1136/bjsports-2014-093842
6. Ardern CL, Taylor NF, Feller JA, Whitehead TS, Webster KE. Sports participation 2 years after

anterior cruciate ligament reconstruction in athletes who had not returned to sport at 1 year: a prospective follow-up of physical function and psychological factors in 122 athletes. *Am J Sports Med*. 2015;43:848-856. https://doi.org/10.1177/0363546514563282

7. Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. *BMJ*. 2004;328:1490. https://doi.org/10.1136/bmj.328.7454.1490
8. Barber-Westin SD, Noyes FR. Factors used to determine return to unrestricted sports activities after anterior cruciate ligament reconstruction. *Arthroscopy*. 2011;27:1697-1705. https://doi.org/10.1016/j.arthro.2011.09.009
9. Barber-Westin SD, Noyes FR. Objective criteria for return to athletics after anterior cruciate ligament reconstruction and subsequent reinjury rates: a systematic review. *Phys Sportsmed*. 2011;39:100-110. https://doi.org/10.3810/psm.2011.09.1926
10. Bell DR, Smith MD, Pennuto AP, Stiffler MR, Olson ME. Jump-landing mechanics after anterior cruciate ligament reconstruction: a Landing Error Scoring System study. *J Athl Train*. 2014;49:435-441. https://doi.org/10.4085/1062-6050-49.3.21
11. Culvenor AG, Collins NJ, Guermazi A, et al. Early knee osteoarthritis is evident one year following anterior cruciate ligament reconstruction: a magnetic resonance imaging evaluation. *Arthritis Rheumatol*. 2015;67:946-955. https://doi.org/10.1002/art.39005
12. Culvenor AG, Collins NJ, Vicenzino B, et al. Predictors and effects of patellofemoral pain following hamstring-tendon ACL reconstruction. *J Sci Med Sport*. 2016;19:518-523. https://doi.org/10.1016/j.jsams.2015.07.008
13. Czuppon S, Racette BA, Klein SE, Harris-Hayes M. Variables associated with return to sport following anterior cruciate ligament reconstruction: a systematic review. *Br J Sports Med*. 2014;48:356-364. https://doi.org/10.1136/bjsports-2012-091786
14. Dekker TJ, Godin JA, Dale KM, Garrett WE, Taylor DC, Riboh JC. Return to sport after pediatric anterior cruciate ligament reconstruction and its effect on subsequent anterior cruciate ligament injury. *J Bone Joint Surg Am*. 2017;99:897-904. https://doi.org/10.2106/JBJS.16.00758
15. Diekfuss JA, Grooms DR, Yuan W, et al. Does brain functional connectivity contribute to musculoskeletal injury? A preliminary prospective analysis of a neural biomarker of ACL injury risk. *J Sci Med Sport*. 2019;22:169-174. https://doi.org/10.1016/j.jsams.2018.07.004
16. Dingenen B, Gokeler A. Optimization of the return-to-sport paradigm after anterior cruciate ligament reconstruction: a critical step back to move forward. *Sports Med*. 2017;47:1487-1500. https://doi.org/10.1007/s40279-017-0674-6
17. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *J Epidemiol Community Health*. 1998;52:377-384.
18. Frobell RB, Roos HP, Roos EM, Roemer FW, Ranstam J, Lohmander LS. Treatment for acute anterior cruciate ligament tear: five year outcome of randomised trial. *BMJ*. 2013;346:f232. https://doi.org/10.1136/bmj.f232
19. Garrison JC, Shanley E, Thigpen C, Geary R, Osler M, DelGiorno J. The reliability of the Vail Sport Test as a measure of physical performance following anterior cruciate ligament reconstruction. *Int J Sports Phys Ther*. 2012;7:20-30.
20. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PM. The diagnostic odds ratio: a single indicator of test performance. *J Clin Epidemiol*. 2003;56:1129-1135. https://doi.org/10.1016/S0895-4356(03)00177-X
21. Gokeler A, Welling W, Zaffagnini S, Seil R, Padua D. Development of a test battery to enhance safe return to sports after anterior cruciate ligament reconstruction. *Knee Surg Sports Traumatol Arthrosc*. 2017;25:192-199. https://doi.org/10.1007/s00167-016-4246-3
22. Grindem H, Mansournia MA, Øiestad BE, Ardern CL. Was it a good idea to combine the studies? Why clinicians should care about heterogeneity when making decisions based on systematic reviews. *Br J Sports Med*. In press. https://doi.org/10.1136/bjsports-2018-099516
23. Grindem H, Snyder-Mackler L, Moksnes H, Engebretsen L, Risberg MA. Simple decision rules can reduce reinjury risk by 84% after ACL reconstruction: the Delaware-Oslo ACL cohort study. *Br J Sports Med*. 2016;50:804-808. https://doi.org/10.1136/bjsports-2016-096031
24. Grooms D, Appelbaum G, Onate J. Neuroplasticity following anterior cruciate ligament injury: a framework for visual-motor training approaches in rehabilitation. *J Orthop Sports Phys Ther*. 2015;45:381-393. https://doi.org/10.2519/jospt.2015.5549
25. Grooms DR, Chaudhari A, Page SJ, Nichols-Larsen DS, Onate JA. Visual-motor control of drop landing after anterior cruciate ligament reconstruction. *J Athl Train*. 2018;53:486-496. https://doi.org/10.4085/1062-6050-178-16
26. Hegedus EJ, McDonough S, Bleakley C, Cook CE, Baxter GD. Clinician-friendly lower extremity physical performance measures in athletes: a systematic review of measurement properties and correlation with injury, part 1. The tests for knee function including the hop tests. *Br J Sports Med*. 2015;49:642-648. https://doi.org/10.1136/bjsports-2014-094094
27. Hewett TE, Di Stasi SL, Myer GD. Current concepts for injury prevention in athletes after anterior cruciate ligament reconstruction. *Am J Sports Med*. 2013;41:216-224. https://doi.org/10.1177/0363546512459638
28. Hewett TE, Myer GD, Ford KR, et al. Biomechanical measures of neuromuscular control and valgus loading of the knee predict anterior cruciate ligament injury risk in female athletes: a prospective study. *Am J Sports Med*. 2005;33:492-501. https://doi.org/10.1177/0363546504269591
29. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327:557-560. https://doi.org/10.1136/bmj.327.7414.557
30. Howick J, Chalmers I, Glasziou P, et al. Explanation of the 2011 Oxford Centre for Evidence-Based Medicine (OCEBM) Levels of Evidence (background document). Available at: https://www.cebm.net/2011/06/explanation-2011-ocebm-levels-evidence/. Accessed December 17, 2018.
31. Ithurburn MP, Paterno MV, Ford KR, Hewett TE, Schmitt LC. Young athletes with quadriceps femoris strength asymmetry at return to sport after anterior cruciate ligament reconstruction demonstrate asymmetric single-leg drop-landing mechanics. *Am J Sports Med*. 2015;43:2727-2737. https://doi.org/10.1177/0363546515602016
32. Jaeschke R, Guyatt G, Lijmer J. Diagnostic tests. In: Guyatt G, Rennie D, eds. *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. Chicago, IL: American Medical Association Press; 2002:121-140.
33. Jenny JY, Clement X. Patient-based decision for resuming activity after ACL reconstruction: a single-centre experience. *Eur J Orthop Surg Traumatol*. 2016;26:929-935. https://doi.org/10.1007/s00590-016-1861-0
34. Kamath GV, Murphy T, Creighton RA, Viradia N, Taft TN, Spang JT. Anterior cruciate ligament injury, return to play, and reinjury in the elite collegiate athlete: analysis of an NCAA Division I cohort. *Am J Sports Med*. 2014;42:1638-1643. https://doi.org/10.1177/0363546514524164
35. Kyritsis P, Bahr R, Landreau P, Miladi R, Witvrouw E. Likelihood of ACL graft rupture: not meeting six clinical discharge criteria before return to sport is associated with a four times greater risk of rupture. *Br J Sports Med*. 2016;50:946-951. https://doi.org/10.1136/bjsports-2015-095908
36. Lai CCH, Ardern CL, Feller JA, Webster KE. Eighty-three per cent of elite athletes return to preinjury sport after anterior cruciate ligament reconstruction: a systematic review with meta-analysis of return to sport rates, graft rupture rates and performance outcomes. *Br J Sports Med*. 2018;52:128-138. https://doi.org/10.1136/bjsports-2016-096836
37. Lind M, Menhert F, Pedersen AB. Incidence and outcome after revision anterior cruciate ligament reconstruction: results from the Danish registry for knee ligament reconstructions. *Am J Sports Med*. 2012;40:1551-1557. https://doi.org/10.1177/0363546512446000
38. Logerstedt D, Di Stasi S, Grindem H, et al. Self-reported knee function can identify athletes who fail return-to-activity criteria up to 1 year after anterior cruciate ligament reconstruction: a Delaware-Oslo ACL cohort study. *J Orthop Sports Phys Ther*. 2014;44:914-923. https://doi.org/10.2519/jospt.2014.4852
39. Maletis GB, Inacio MC, Funahashi TT. Risk factors associated with revision and contralateral

anterior cruciate ligament reconstructions in the Kaiser Permanente ACLR registry. *Am J Sports Med*. 2015;43:641-647. https://doi.org/10.1177/0363546514561745
40. Mather RC, 3rd, Koenig L, Kocher MS, et al. Societal and economic impact of anterior cruciate ligament tears. *J Bone Joint Surg Am*. 2013;95:1751-1759. https://doi.org/10.2106/JBJS.L.01705
41. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Int J Surg*. 2010;8:336-341. https://doi.org/10.1016/j.ijsu.2010.02.007
42. Morgan MD, Salmon LJ, Waller A, Roe JP, Pinczewski LA. Fifteen-year survival of endoscopic anterior cruciate ligament reconstruction in patients aged 18 years and younger. *Am J Sports Med*. 2016;44:384-392. https://doi.org/10.1177/0363546515623032
43. Mosler AB, Agricola R, Weir A, Hölmich P, Crossley KM. Which factors differentiate athletes with hip/groin pain from those without? A systematic review with meta-analysis. *Br J Sports Med*. 2015;49:810. https://doi.org/10.1136/bjsports-2015-094602
44. Munn J, Sullivan SJ, Schneiders AG. Evidence of sensorimotor deficits in functional ankle instability: a systematic review with meta-analysis. *J Sci Med Sport*. 2010;13:2-12. https://doi.org/10.1016/j.jsams.2009.03.004
45. Myer GD, Paterno MV, Ford KR, Quatman CE, Hewett TE. Rehabilitation after anterior cruciate ligament reconstruction: criteria-based progression through the return-to-sport phase. *J Orthop Sports Phys Ther*. 2006;36:385-402. https://doi.org/10.2519/jospt.2006.2222
46. Nagelli CV, Hewett TE. Should return to sport be delayed until 2 years after anterior cruciate ligament reconstruction? Biological and functional considerations. *Sports Med*. 2017;47:221-232. https://doi.org/10.1007/s40279-016-0584-z
47. Nawasreh Z, Logerstedt D, Cummer K, Axe MJ, Risberg MA, Snyder-Mackler L. Do patients failing return-to-activity criteria at 6 months after anterior cruciate ligament reconstruction continue demonstrating deficits at 2 years? *Am J Sports Med*. 2017;45:1037-1048. https://doi.org/10.1177/0363546516680619
48. Nielsen RO, Bertelsen ML, Verhagen E, et al. When is a study result important for athletes, clinicians and team coaches/staff? *Br J Sports Med*. 2017;51:1454-1455. https://doi.org/10.1136/bjsports-2017-097759
49. Nielsen RØ, Malisoux L, Møller M, Theisen D, Parner ET. Shedding light on the etiology of sports injuries: a look behind the scenes of time-to-event analyses. *J Orthop Sports Phys Ther*. 2016;46:300-311. https://doi.org/10.2519/jospt.2016.6510
50. Padua DA, DiStefano LJ, Beutler AI, de la Motte SJ, DiStefano MJ, Marshall SW. The Landing Error Scoring System as a screening tool for an anterior cruciate ligament injury-prevention program in elite-youth soccer athletes. *J Athl Train*. 2015;50:589-595. https://doi.org/10.4085/1062-6050-50.1.10
51. Paterno MV. Incidence and predictors of second anterior cruciate ligament injury after primary reconstruction and return to sport. *J Athl Train*. 2015;50:1097-1099. https://doi.org/10.4085/1062-6050-50.10.07
52. Paterno MV, Ford KR, Myer GD, Heyl R, Hewett TE. Limb asymmetries in landing and jumping 2 years following anterior cruciate ligament reconstruction. *Clin J Sport Med*. 2007;17:258-262. https://doi.org/10.1097/JSM.0b013e31804c77ea
53. Paterno MV, Huang B, Thomas S, Hewett TE, Schmitt LC. Clinical factors that predict a second ACL injury after ACL reconstruction and return to sport: preliminary development of a clinical decision algorithm. *Orthop J Sports Med*. 2017;5:2325967117745279. https://doi.org/10.1177/2325967117745279
54. Paterno MV, Kiefer AW, Bonnette S, et al. Prospectively identified deficits in sagittal plane hip–ankle coordination in female athletes who sustain a second anterior cruciate ligament injury after anterior cruciate ligament reconstruction and return to sport. *Clin Biomech (Bristol, Avon)*. 2015;30:1094-1101. https://doi.org/10.1016/j.clinbiomech.2015.08.019
55. Paterno MV, Kiefer AW, Bonnette SH, et al. Deficits in hip-ankle coordination in female athletes who suffer a second anterior cruciate ligament (ACL) injury after ACL reconstruction and return to sport. *Orthop J Sports Med*. 2014;2 suppl 2:2325967114S2325900039. https://doi.org/10.1177/2325967114S00039
56. Paterno MV, Schmitt LC, Ford KR, et al. Biomechanical measures during landing and postural stability predict second anterior cruciate ligament injury after anterior cruciate ligament reconstruction and return to sport. *Am J Sports Med*. 2010;38:1968-1978. https://doi.org/10.1177/0363546510376053
57. Ross CA, Clifford A, Louw QA. Factors informing fear of reinjury after anterior cruciate ligament reconstruction. *Physiother Theory Pract*. 2017;33:103-114. https://doi.org/10.1080/09593985.2016.1271847
58. Ryan R, Cochrane Consumers and Communication Review Group. Heterogeneity and subgroup analyses in Cochrane Consumers and Communication Group reviews: planning the analysis at protocol stage. Available at: http://cccrg.cochrane.org/sites/cccrg.cochrane.org/files/public/uploads/heterogeneity_subgroup_analyses_revising_december_1st_2016.pdf. Accessed March 10, 2018.
59. Ryan R, Hill S, Cochrane Consumers and Communication Group. How to GRADE the quality of the evidence. Available at: https://cc.cochrane.org/sites/cc.cochrane.org/files/public/uploads/how_to_grade.pdf. Accessed December 17, 2018.
60. Sanders TL, Maradit Kremers H, Bryan AJ, et al. Incidence of anterior cruciate ligament tears and reconstruction: a 21-year population-based study. *Am J Sports Med*. 2016;44:1502-1507. https://doi.org/10.1177/0363546516629944
61. Schmitt LC, Paterno MV, Ford KR, Myer GD, Hewett TE. Altered knee mechanics during landing at the time of return to sport following ACL reconstruction: effects of quadriceps strength asymmetry [abstract]. *J Orthop Sports Phys Ther*. 2010;40:A42-A43. https://doi.org/10.2519/jospt.2010.40.1.A40
62. Schmitt LC, Paterno MV, Ford KR, Myer GD, Hewett TE. Strength asymmetry and landing mechanics at return to sport after anterior cruciate ligament reconstruction. *Med Sci Sports Exerc*. 2015;47:1426-1434. https://doi.org/10.1249/MSS.0000000000000560
63. Schmitt LC, Paterno MV, Hewett T. Functional performance at the time of return to sport following ACL reconstruction: the impact of quadriceps strength asymmetry [abstract]. *J Orthop Sports Phys Ther*. 2009;39:A103. https://doi.org/10.2519/jospt.2009.39.1.A101
64. Schmitt LC, Paterno MV, Hewett TE. The impact of quadriceps femoris strength asymmetry on functional performance at return to sport following anterior cruciate ligament reconstruction. *J Orthop Sports Phys Ther*. 2012;42:750-759. https://doi.org/10.2519/jospt.2012.4194
65. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*. 2017;358:j4008. https://doi.org/10.1136/bmj.j4008
66. Shelbourne KD, Benner RW, Gray T. Return to sports and subsequent injury rates after revision anterior cruciate ligament reconstruction with patellar tendon autograft. *Am J Sports Med*. 2014;42:1395-1400. https://doi.org/10.1177/0363546514524921
67. Shelbourne KD, Gray T, Haro M. Incidence of subsequent injury to either knee within 5 years after anterior cruciate ligament reconstruction with patellar tendon autograft. *Am J Sports Med*. 2009;37:246-251. https://doi.org/10.1177/0363546508325665
68. Shelbourne KD, Klootwyk TE, DeCarlo MS. Update on accelerated rehabilitation after anterior cruciate ligament reconstruction. *J Orthop Sports Phys Ther*. 1992;15:303-308. https://doi.org/10.2519/jospt.1992.15.6.303
69. Sousa PL, Krych AJ, Cates RA, Levy BA, Stuart MJ, Dahm DL. Return to sport: does excellent 6-month strength and function following ACL reconstruction predict midterm outcomes? *Knee Surg Sports Traumatol Arthrosc*. 2017;25:1356-1363. https://doi.org/10.1007/s00167-015-3697-2
70. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*. 2016;355:i4919. https://doi.org/10.1136/bmj.i4919
71. Swartz MK. The PRISMA statement: a guideline

for systematic reviews and meta-analyses. *J Pediatr Health Care*. 2011;25:1-2. https://doi.org/10.1016/j.pedhc.2010.09.006

72. Tengman E, Brax Olofsson L, Nilsson KG, Tegner Y, Lundgren L, Häger CK. Anterior cruciate ligament injury after more than 20 years: I. Physical activity level and knee function. *Scand J Med Sci Sports*. 2014;24:e491-e500. https://doi.org/10.1111/sms.12212

73. Tengman E, Brax Olofsson L, Stensdotter AK, Nilsson KG, Häger CK. Anterior cruciate ligament injury after more than 20 years. II. Concentric and eccentric knee muscle strength. *Scand J Med Sci Sports*. 2014;24:e501-e509. https://doi.org/10.1111/sms.12215

74. Thomeé P, Senorski E, Thomeé R. Return to pre-injury knee strenuous sports after an ACL reconstruction, a report from a rehabilitation outcome registry on patient characteristics [abstract]. *Physiotherapy*. 2015;101 suppl 1:e1515. https://doi.org/10.1016/j.physio.2015.03.1499

75. Thomeé P, Währborg P, Börjesson M, Thomeé R, Eriksson BI, Karlsson J. Self-efficacy of knee function as a pre-operative predictor of outcome 1 year after anterior cruciate ligament reconstruction. *Knee Surg Sports Traumatol Arthrosc*. 2008;16:118-127. https://doi.org/10.1007/s00167-007-0433-6

76. Thomeé R, Kaplan Y, Kvist J, et al. Muscle strength and hop performance criteria prior to return to sports after ACL reconstruction. *Knee Surg Sports Traumatol Arthrosc*. 2011;19:1798-1805. https://doi.org/10.1007/s00167-011-1669-8

77. Thomeé R, Neeter C, Gustavsson A, et al. Variability in leg muscle power and hop performance after anterior cruciate ligament reconstruction. *Knee Surg Sports Traumatol Arthrosc*. 2012;20:1143-1151. https://doi.org/10.1007/s00167-012-1912-y

78. Thomeé R, Waldén M, Hägglund M. Return to sports after anterior cruciate ligament injury: neither surgery nor rehabilitation alone guarantees success—it is much more complicated. *Br J Sports Med*. 2015;49:1422. https://doi.org/10.1136/bjsports-2015-094793

79. Tjong VK, Murnaghan ML, Nyhof-Young JM, Ogilvie-Harris DJ. A qualitative investigation of the decision to return to sport after anterior cruciate ligament reconstruction: to play or not to play. *Am J Sports Med*. 2014;42:336-342. https://doi.org/10.1177/0363546513508762

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

81. Webster KE, Feller JA. Exploring the high reinjury rate in younger patients undergoing anterior cruciate ligament reconstruction. *Am J Sports Med*. 2016;44:2827-2832. https://doi.org/10.1177/0363546516651845

82. Wellsandt E, Failla MJ, Snyder-Mackler L. Limb symmetry indexes can overestimate knee function after anterior cruciate ligament injury. *J Orthop Sports Phys Ther*. 2017;47:334-338. https://doi.org/10.2519/jospt.2017.7285

83. Wiggins AJ, Grandhi RK, Schneider DK, Stanfield D, Webster KE, Myer GD. Risk of secondary injury in younger athletes after anterior cruciate ligament reconstruction: a systematic review and meta-analysis. *Am J Sports Med*. 2016;44:1861-1876. https://doi.org/10.1177/0363546515621554

84. Wren TAL, Mueske NM, Brophy CH, et al. Hop distance symmetry does not indicate normal landing biomechanics in adolescent athletes with recent anterior cruciate ligament reconstruction. *J Orthop Sports Phys Ther*. 2018;48:622-629. https://doi.org/10.2519/jospt.2018.7817

APPENDIX A

COMPUTER-ASSISTED SEARCHES

PubMed/MEDLINE

(surgery[tiab] OR surgical[tiab] OR repair[tiab] OR reconstructive[tiab] OR reconstruction[tiab] OR "surgery" [Subheading] OR "Reconstructive Surgical Procedures"[Mesh] OR graft[tiab] OR autograft[tiab]) AND ("Anterior Cruciate Ligament"[Mesh] OR "Anterior Cruciate Ligament"[tiab] OR ACL[tiab] OR ACLR[tiab]) AND ("athletes"[MeSH Terms] OR "athletes"[tiab] OR "Athletic Injuries"[Mesh] OR athletic[tiab] OR "sport"[tiab] OR "sports"[MeSH Terms] OR "sports"[tiab] OR "Military Personnel"[Mesh] OR military[tiab] OR soldier[tiab] OR soldiers[tiab]) AND ("return to sport"[MeSH Terms] OR "Return to sport"[tiab] OR "Return to sports"[tiab] OR "return to play"[tiab] OR "return to competition"[tiab] OR "return to"[tiab] OR "return to duty"[tiab] OR "return to activity"[tiab]) NOT (Editorial[ptyp] OR Letter[ptyp] OR Case Reports[ptyp] OR Comment[ptyp]) NOT (animals[mh] NOT humans[mh])

Embase

(surgery:ab,ti OR surgical:ab,ti OR repair:ab,ti OR reconstructive:ab,ti OR reconstruction:ab,ti OR graft:ab,ti OR autograft:ab,ti) AND ('anterior cruciate ligament'/de OR acl:ab,ti OR aclr:ab,ti) AND ('athlete'/de OR athletes:ab,ti OR 'sport injury'/de OR athletic:ab,ti OR sport:ab,ti OR sports:ab,ti OR 'sport'/de OR 'military'/de OR 'army'/de OR 'soldier'/de OR 'soldiers') AND ('return to sport'/de OR 'return to sport':ab,ti OR 'return to sports':ab,ti OR 'return to play':ab,ti OR 'return to competition':ab,ti OR 'return to':ab,ti OR 'return to duty':ab,ti OR 'return to activity':ab,ti) AND [humans]/lim AND [english]/lim AND [embase]/lim

SPORTDiscus

(DE "ANTERIOR cruciate ligament" OR "anterior cruciate ligament") AND (surgery OR surgical OR repair OR reconstructive OR reconstruction OR OR graft OR autograft) AND ("return to sport" OR "Return to sports" OR "return to play" OR "return to competition" OR "return to" OR "return to duty" OR "return to activity")

CINAHL

((MH "Surgery, Reconstructive+") OR surgery OR surgical OR repair OR reconstructive OR reconstruction OR graft OR autograft) AND ((MH "Anterior Cruciate Ligament") OR (MH "Anterior Cruciate Ligament Injuries") OR (MH "Anterior Cruciate Ligament Reconstruction") OR "Anterior Cruciate Ligament" OR ACL OR ACLR) AND ((MH "Sports+") OR (MH "Athletes+") OR (MH "Athletic Injuries") OR (MH "Military Personnel") OR "athletes" OR athletic OR "sport" OR "sports" OR military OR soldier OR soldiers) AND ((MH "Sports Re-Entry") OR "Return to sport" OR "Return to sports" OR "return to play" OR "return to competition" OR "return to" OR "return to duty" OR "return to activity")

ProQuest Dissertations and Theses Global

all(surgery OR surgical OR repair OR reconstructive OR reconstruction OR graft OR autograft) AND all("Anterior Cruciate Ligament" OR ACL OR ACLR) AND all(athletes OR athletic OR sport OR sports OR military OR soldier OR soldiers) AND all("Return to sport" OR "Return to sports" OR "return to play" OR "return to competition" OR "return to" OR "return to duty" OR "return to activity")

APPENDIX B

GRADING OF RECOMMENDATIONS ASSESSMENT, DEVELOPMENT AND EVALUATION SCALE[59]

Quality	Interpretation
High	Very confident that the true effect lies close to that of the estimate of the effect
Moderate	Moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low	Confidence in the effect estimate is limited; the true effect may be substantially different from the estimate of the effect
Very low	Very little confidence in the effect estimate; the true effect is likely to be substantially different from the estimate of effect

APPENDIX C

SUMMARY OF INCLUDED STUDIES

For all summaries, patient demographics, return-to-sport (RTS) criteria, and cutoff scores can be found in **TABLE 2**. Reinjury statistics and pass and fail incidences can be found in **TABLE 3**.

Kyritsis et al[35]

Kyritsis et al[35] published data from 158 male professional athletes treated at Aspetar Orthopaedic and Sports Medicine Hospital following primary anterior cruciate ligament reconstruction (ACLR). They tested 6 RTS criteria and tracked their patients to determine who sustained an anterior cruciate ligament (ACL) graft injury (**TABLE 2**). Those who successfully met RTS criteria were considered "fully discharged." They reported a secondary ACL injury incidence of 10.3% (12/116) for the fully discharged patients and 33.3% (14/42) for patients who were not discharged (**TABLE 3**). They also reported 11 contralateral ACL injuries, which were not included in their analysis. The time from RTS until secondary injury for 65% (17/26) of patients was within 6 months of discharge. Regression modeling determined that patients who had a lower hamstring-to-quadriceps ratio were at risk for injury (every 10% decrease in that ratio equaled a 10.6-fold higher risk). The authors concluded that athletes who did not meet RTS criteria had a 4-fold greater risk of sustaining an ACL graft rupture.

Grindem et al[23]

Grindem et al[23] published data from 100 patients in the Norwegian arm of the Delaware-Oslo ACL cohort study following ACLR. They utilized 7 RTS criteria (**TABLE 2**). Sixty-nine patients completed functional testing. They sought to determine the risk of all knee injuries following RTS. The secondary knee injury incidence was 22 of 74 for those who returned to level 1 sports. Failure to pass RTS criteria occurred in 55 of 73 (75.3%) patients (**TABLE 3**). Twenty-one patients who failed RTS criteria suffered a knee injury (38.2%). Time from RTS to reinjury was between 3 and 22 months post ACLR (median, 13 months). Individuals who returned to a level 1 sport were 4 times more likely to experience a knee injury in the first 2 years. The authors found that a 1% increase in quadriceps limb symmetry index was correlated with a 3% reduced reinjury risk. They also determined that within the first 9 months, a later RTS was associated with a lower reinjury risk (for every 1-month delay in RTS, knee injury incidence was reduced by 51%).

Sousa et al[69]

Sousa et al[69] published data from a case-control retrospective review of 223 patients who underwent primary ACLR by a single surgeon from 1998 to 2005. The patients followed identical rehabilitation protocols and were eligible for undergoing a 7-item RTS test battery at 6 months post surgery (**TABLE 2**). A satisfactory result in 6 of 7 tests would categorize them as the "excellent 6-month group"; if they did not reach satisfactory performance, then they were categorized as the "delayed 6-month group" (**TABLE 2**). Fifty-two of 223 (23%) patients were in the excellent group and 171 of 223 (77%) were in the delayed 6-month group (**TABLE 3**). Ipsilateral graft reinjury incidence for the entire population was 10 of 223, and contralateral ACL injury occurred in 17 of 223 (8%) patients at a mean of 44 months (range, 22-123 months) post ACLR (**TABLE 3**). The authors concluded that patients returning to sport 6 months after ACLR were at increased risk for a contralateral ACL injury.

Nawasreh et al[47]

Nawasreh et al[47] published data from a cohort study of 95 patients who underwent ACLR and completed RTS testing after injury during level 1 or 2 sport competition. Of the 108 initially enrolled, full data sets (6-, 12-, and 24-month follow-ups) were available for 60 patients. They utilized 7 RTS criteria (**TABLE 2**). Forty-eight of 95 (51%) passed RTS criteria at 6 months post ACLR. Seven of 95 (7.4%) patients suffered a second ACL injury, 3 of 7 (43%) patients failed RTS criteria, and 4 of 7 (57%) passed RTS criteria. The authors concluded that a battery of RTS tests that include performance-based and patient-reported outcomes could be utilized to identify those with persistent dysfunction and possible higher risk of a second ACL injury.

APPENDIX D

MODIFIED DOWNS AND BLACK[17] SCORES FOR ALL INCLUDED STUDIES*

	Modified Downs and Black Score Distribution†															
Study	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12‡**	**13**	**14**	**15**	**Total**
Kyritsis et al[35]	1	1	N	1	1	1	1	N	1	1	1	2	N	N	1	12
Grindem et al[23]	1	1	1	N	1	1	1	N	1	1	1	2	1	N	1	13
Nawasreh et al[47]	1	1	1	N	1	N	1	N	1	1	1	2	1	N	1	12
Sousa et al[69]	1	1	1	1	1	1	1	N	1	1	1	2	1	N	1	14

Abbreviation: N, criterion not met.

**κ = 0.54 ± 0.13 (moderate agreement).*

†Assesses study quality based on a set of 15 questions, with a total score of 16. A score of 1 indicates that the criterion was met. Items: 1, Hypothesis/aim/objective of the study clearly described; 2, Characteristics of the patients clearly described; 3, Patient sample representative of patients treated in routine clinical practice; 4, Is there information on the possibility of selection bias? 5, Was a comparison group identified and clearly defined? 6, Are the main outcomes clearly described in the Introduction or Methods? 7, Were the main outcome measures used accurate (valid and reliable)? 8, Was there any attempt to blind those measuring the main outcomes? 9, Are the main findings of the study clearly described? 10, Does the study provide estimates of random variability? 11, Were the statistical tests used to assess the main outcomes appropriate? 12, Are the distribution of principal confounders in each group of subjects to be compared clearly described? 13, Was there adequate adjustment for confounding in the analyses from which the main findings were drawn? 14, Was a sample-size calculation reported? 15, Was there sufficient power to detect a clinically important effect where the probability value for a difference being due to chance is less than 5%?

‡Item 12 had score options of 2, fully described and 1, partially described.

APPENDIX E

GRADE TABLE FOR RISK OF SECOND ACL INJURY OUTCOME[59]

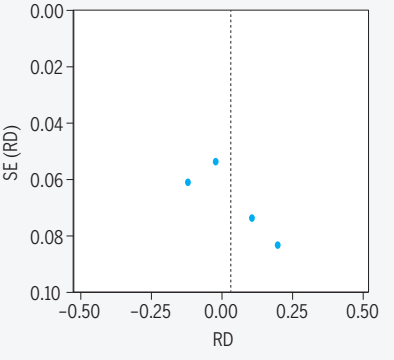
GRADE Criteria	Possible Rating	Rating Score Given	Justification (Explanation of Downgrading or Upgrading)
Study design	RCT (starts at +4, high quality), NRCT (starts at +2, low quality)	+2	Only NRCT studies available
Risk of bias	No, serious (–1), very serious (–2)	0	Any plausible bias unlikely to seriously alter this outcome
Inconsistency	No, serious (–1), very serious (–2)	–1	Although inconsistency above acceptable standards was noted (I^2 = 74%, P = .008), there are factors that explain this inconsistency (source population, return-to-sport criteria tested, cutoff scores)
Indirectness	No, serious (–1), very serious (–2)	0	No serious concerns for indirectness of the evidence
Imprecision	No, serious (–1), very serious (–2)	–1	There is concern for imprecision of the estimate due to wide confidence intervals, with each end indicating an opposite effect
Publication bias	Undetected, strongly suspected (–1)	0	Undetected, and variation in funnel-plot symmetry likely due to heterogeneity
Upgrading factors	Large effect (+1 or +2), dose response (+1 or +2), no plausible confounding (+1 or +2)	0	No upgrading factors
Overall quality-of-evidence score*	...	0†	...

Abbreviations: ACL, anterior cruciate ligament; GRADE, Grading of Recommendations Assessment, Development and Evaluation scale; NRCT, nonrandomized controlled trial; RCT, randomized controlled trial.
**Sum of rating score given for all GRADE criteria. High quality, 4 or greater; moderate quality, 3; low quality, 2; very low quality, 1 or less.*
†Very low quality.

RISK-OF-BIAS TABLE, ADAPTED FROM ROBINS-I[70] AND USED IN GRADE DETERMINATION

Study	Confounding	Selection of Participants	Missing Data	Measurement of Outcome	Selection of Reported Result	Overall Bias
Kyritsis et al[35]	Low	Low	Low	Low	Low	Low
Grindem et al[23]	Low	Low	Low	Low	Low	Low
Nawasreh et al[47]	Low	Low	Moderate	Low	Low	Low
Sousa et al[69]	Low	Moderate	Low	Low	Low	Low

Abbreviations: GRADE, Grading of Recommendations Assessment, Development and Evaluation scale; ROBINS-I, Risk Of Bias In Non-randomised Studies-of Interventions.



Funnel plot for second anterior cruciate ligament injury RD analysis.
Abbreviations: RD, risk difference; SE, standard error.

APPENDIX F

AMSTAR, VERSION 2[65]

Question*	Criteria Met	Justification
1	Yes	PICO established in research question and selection criteria
2	No	Methods were not established prior to conducting this review
3	No	RCTs and NRSI were allowed due to low number of studies on the topic
4	Yes	Searched 5 databases, considered gray literature, provided all searches for reviewers, and searched within 24 months
5	Yes	Two reviewers independently performed study selection
6	Yes	Two reviewers independently performed data extraction
7	Partial yes	A list of reasons for excluded studies was provided, but not a complete list of references
8	Yes	Adequate and thorough description of included studies was provided
9	Yes	Risk of bias was reported and used to factor into GRADE scoring
10	Yes	Funding sources of included studies were reported
11	Yes	Appropriate statistical analysis methodology was used
12	Yes	Risk of bias was considered during GRADE scoring, which impacts the quality-of-evidence rating and the strength of this review's conclusion
13	Partial yes	Risk of bias was considered during GRADE scoring, but not individually discussed
14	Yes	Satisfactory explanation for heterogeneity was provided
15	Partial yes	Publication bias was considered during GRADE scoring, but not further discussed
16	Yes	We reported no competing interests

*Abbreviations: AMSTAR, A MeaSurement Tool to Assess systematic Reviews; GRADE, Grading of Recommendations Assessment, Development and Evaluation scale; NRSI, nonrandomized studies of health care interventions; PICO, patient/problem, intervention, comparison, outcome; RCT, randomized controlled trial. *Framework for development of research questions: 1, Did the research question and inclusion criteria include PICO? 2, Were review methods/protocol established prior to conduct of review? 3, Did the authors explain selection criteria based on study design? 4, Was there a comprehensive literature search? 5, Was the study selection in duplicate? 6, Was data extraction in duplicate? 7, Was a list of excluded studies provided? 8, Did the authors describe included studies in adequate detail? 9, Was there a satisfactory technique to assess risk of bias? 10, Did the authors report sources of funding for included studies? 11, Were appropriate statistical methods used for meta-analysis? 12, Did the authors assess the impact of risk of bias on results of meta-analysis? 13, Was risk of bias accounted for when discussing results? 14, Was there discussion of heterogeneity? 15, Was there investigation into publication bias? 16, Was any conflict of interest reported?*

POTENTIAL SOURCES OF CONFLICT AND SOURCES OF FUNDING FOR INCLUDED STUDIES

Included Studies	Potential Sources of Support (if reported in study)
Kyritsis et al[35]	Funding: none reported. The authors did not declare any conflicts of interest
Grindem et al[23]	Funding: National Institutes of Health (R37 HD037985). The authors did not declare any conflicts of interest
Nawasreh et al[47]	"One or more of the authors has declared the following potential conflict of interest or source of funding: this ongoing prospective cohort study was funded by the National Institutes of Health (NIH R37HD37985, P30 GM103333) and a Promotion of Doctoral Studies (PODS) I Scholarship"
Sousa et al[69]	Funding: none reported. The authors did not declare any conflicts of interest

MICHAEL P. REIMAN, PT, PhD, OCS, SCS, ATC, FAAOMPT, CSCS[1,2] · **SIMON DÉCARY,** PT, PhD[3]
BENOY MATHEW, MSc, MAACP, MCSP[4] • **CARLY K. REIMAN**[5]

Accuracy of Clinical and Imaging Tests for the Diagnosis of Hip Dysplasia and Instability: A Systematic Review

BACKGROUND: Evidence concerning the accurate clinical or imaging methods to diagnose hip instability or hip dysplasia is currently scarce.

OBJECTIVE: To summarize the diagnostic accuracy of clinical and imaging tests for the diagnosis of hip dysplasia and instability.

METHODS: A computer-assisted literature search of the MEDLINE, CINAHL, and Embase databases, using key words related to diagnostic accuracy of the hip joint, was conducted on March 6, 2018. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were used for the searching and reporting phases of the study. Quality assessment of bias and applicability was conducted using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool. Diagnostic accuracy, including sensitivity, specificity, likelihood ratio, and 95% confidence interval, was summarized.

RESULTS: Out of 3109 citations, 7 articles were included. Two studies reported on 5 clinical tests for hip instability and 5 studies reported on 5 radiographic measures for hip dysplasia. Only 1 study was not of low methodological quality. The prone instability test moderately improved positive posttest probability by 38% to diagnose hip instability. The Shenton line moderately to highly improved posttest probability by 41% to 60% to diagnose hip dysplasia.

CONCLUSION: This systematic review summarizes the diagnostic accuracy of various clinical tests and radiographic measures for hip instability and hip dysplasia. Further high-quality studies are necessary to examine the diagnostic accuracy of the clinical examination and radiography to assist in ruling in or ruling out the diagnoses of hip dysplasia and instability. Consensus is required to standardize the definitions of these diagnoses and their reference standards. The study was registered with the International Prospective Register of Systematic Reviews (CRD42018089019).

LEVEL OF EVIDENCE: Diagnosis, level 3b.
J Orthop Sports Phys Ther 2019;49(2):87-97. Epub 30 Nov 2018. doi:10.2519/jospt.2019.8476

KEY WORDS: *center-edge angle, diagnostic accuracy, prone instability test, radiography, Shenton line*

Hip pain is a commonly encountered reason for consultation in physical therapy clinics. The current focus of nonarthritic hip joint–related pathology is on femoroacetabular impingement (FAI) syndrome, with publications reporting surgical outcomes having increased 2600% over a 13-year period.[29] This has driven recent efforts to reach consensus on the definition of nonarthritic hip joint–related pathologies, such as FAI syndrome[29] or labral tears. Systematic reviews and meta-analyses have been published to summarize the diagnostic accuracy of clinical and imaging tests for these pathologies; however, because individuals with other hip pathologies, such as hip dysplasia and instability, may have similar patient characteristics to those with FAI syndrome,[31] these conditions need to be included in a complete differential diagnosis of hip pain. These diagnoses have been left largely ignored in the literature.

The term *hip dysplasia* refers to misalignment between the femoral head and the acetabulum secondary to changes in their shape, size, and orientation,[40] resulting in structural instability that causes mechanical overloading of the acetabular rim during normal activities.[9,12] Hip dysplasia, at its core, is a condition of instability[40] and has been linked with development of early hip osteoarthritis.[4,14] Dysplasia, especially borderline dysplasia, has variable definitions,[42] and the reliability of analyzing various radiographic measures is questionable.[24,34]

Traditionally, the definition has focused on radiography, with a lateral center-edge angle (CEA) measurement of less than 20° to 25° and/or an anterior CEA of less than 20°.[31] The variability in a consistent definition of dysplasia has led to a proposed diagnostic framework, grouping symptomatic dysplastic hips into 3 categories based on the primary direction of instability.[40] This incorporation of instability is more likely to encapsulate the condition these individuals present

[1]Department of Orthopaedic Surgery, Duke University Medical Center, Durham, NC. [2]Duke University Orthopaedic Fellowship, Duke University Medical Center, Durham, NC. [3]Faculty of Medicine, Université Laval, Quebec, Canada. [4]Department of Trauma and Orthopaedics, Croydon University Hospital, London, UK. [5]Department of Biological Sciences, North Carolina State University, Raleigh, NC. The study was registered with the International Prospective Register of Systematic Reviews (CRD42018089019). 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 Michael P. Reiman, Duke University School of Medicine, Department of Orthopaedic Surgery, Doctor of Physical Therapy Division, DUMC 104002, Durham, NC 27710. E-mail: reiman.michael@gmail.com

with. Although instability of the hip joint has been classically associated with hip dysplasia, multiple intra-articular and extra-articular factors may contribute to hip instability.[16,32]

The diagnostic process is an ongoing assessment of whether sufficient information has been collected. The goal of information gathering in the diagnostic process is to reduce the diagnostic uncertainty enough to make optimal decisions for subsequent care.[17] The diagnoses of hip dysplasia and instability lack precision. To our knowledge, a systematic synthesis of the diagnostic accuracy regarding the clinical or imaging tests for these pathologies does not exist. Therefore, the purpose of this systematic review was to summarize the diagnostic accuracy of clinical and imaging tests for the diagnosis of hip dysplasia and instability.

METHODS

Registration

The study was registered on February 19, 2018 with the International Prospective Register of Systematic Reviews (PROSPERO), a database of prospectively registered systematic reviews for health and social topics. The study was registered after the pilot search and prior to the updated data search (CRD42018089019).

Data Sources

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses of diagnostic test accuracy studies (PRISMA-DTA) guidelines[25] were utilized during the searching and reporting phases of this review. The PRISMA-DTA guidelines comprise a stand-alone extension of the original 27-item checklist that was designed to be used as a basis for reporting systematic reviews of randomized trials,[26] but can also be applied to multiple research methodologies.[33] The PRISMA-DTA guideline checklist is designed to facilitate transparent reporting of reviews of diagnostic accuracy studies.[25]

Identification and Selection of the Literature

A computer-assisted literature search of the MEDLINE, CINAHL, and Embase databases was performed from inception of each respective database to March 6, 2018. As the goal was to optimize the sensitivity of our search strategy,[38,39] and to increase the likelihood that all appropriate studies were identified, we also searched Google Scholar and hand searched included studies for additional references. The search strategy was developed in collaboration with a biomedical librarian and used controlled vocabulary and key words related to diagnostic accuracy of the clinical examination measures relative to hip dysplasia and/or instability. Screening filters were initially used during assessment of title, abstract, and full-text documents. The search was further limited to humans and English- or French-language publications. Findings from systematic reviews suggest that there is no evidence of bias for conventional medicine studies when studies written in languages other than English are excluded.[27]

The full search strategy for MEDLINE is listed in **APPENDIX A** (available at www.jospt.org). To be included in the systematic review, the studies had to satisfy the following criteria.

Participants Participants in the included studies were between 13 and 65 years of age and had to have hip pain suspected to be related to hip dysplasia or instability, as diagnosed by at least 1 imaging or clinical test utilizing an appropriate reference standard. Studies that used examination measures with specialized instrumentation, included participants with congenitally related conditions (eg, Ehlers-Danlos or Marfan syndrome) or who were infants/toddlers or cadavers, or were not written in English or French were excluded.

Reference Standard for the Diagnosis of Dysplasia There is a lack of consensus on the reference standard for diagnosis of hip dysplasia, as multiple radiographic measurements are used[3,34] and the adult acetabular anatomy varies according to sex and ethnicity.[20] The lateral CEA, first described by Wiberg,[37] is the most commonly utilized radiographic measure. A CEA less than 20°, as measured on an anteroposterior radiograph of the pelvis, can be utilized to diagnose dysplasia.[1,8,22] Values between 20° and 25° have been classified as borderline dysplasia.[22]

Diagnosis of Hip Instability Defined as extraphysiological hip motion that causes pain and impairs function,[32] hip instability is a multifactorial condition that encompasses a broad range of causes, from trauma, generalized ligamentous laxity, collagen disorders, bone abnormalities, to soft tissue laxity.[6] Clinical diagnosis of hip instability can be challenging due to lack of specific signs and symptoms, and the presentation can be quite subtle.[16] At present, there is no established objective or radiological signs specific to hip instability.[19]

Intervention Examination studies (clinical or radiologic examination) reporting diagnostic estimates for the diagnosis of hip dysplasia and/or instability were included.

Comparator Analysis of comparisons or of subgroups is included where appropriate. Clinical or radiological examination comparing the diagnostic accuracy of hip dysplasia versus instability and comparisons between clinical and radiological examinations for the diagnosis of either hip dysplasia or instability were included.

Outcomes Studies that reported diagnostic accuracy (eg, sensitivity, specificity, likelihood ratios), pretest and posttest probability, as well as the degree of posttest probability shifts were included. Secondary outcomes included study level of evidence, study purpose, definitions of dysplasia/instability by each study, the type of clinician interpreting diagnosis, and the reliability of examination measures.

Time All time frames reporting diagnostic accuracy for clinical and/or radiological examination of hip dysplasia and/or instability were included.

Study Design

A literature search was conducted for diagnostic accuracy studies (primary

experimental evidence) investigating either clinical or imaging examination measures for the diagnosis of hip dysplasia or instability, published in either English or French (the authors' native language). Two authors (M.P.R., C.K.R.) independently performed the search. As computerized search results for diagnostic accuracy data frequently omit relevant articles,[7] the reference lists of all selected publications were checked to retrieve relevant publications not identified in the computerized search. Gray literature was also hand searched and included publications, posters, abstracts, and conference proceedings. Duplicate studies were removed. To identify relevant articles, titles and abstracts of all identified citations were independently screened. Full-text articles were retrieved if the abstract provided insufficient information to establish eligibility or if the article passed the first eligibility screening. Reasons for article exclusion are provided in **APPENDIX B** (available at www.jospt.org). Disagreements among the reviewers were discussed and resolved by consensus. Interobserver agreement of study inclusion between 2 authors was assessed using kappa statistics, with values less than 0 indicating no agreement, and 0.00 to 0.20 slight, 0.21 to 0.40 fair, 0.41 to 0.60 moderate, 0.61 to 0.80 substantial, and 0.81 to 1.00 almost perfect agreement.[21]

Quality Assessment

Each of the full-text articles was independently reviewed by 2 reviewers (M.P.R., S.D.) and scored with the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool.[35] Disagreements among the reviewers were discussed and resolved during a consensus meeting. The QUADAS-2 is a quality assessment tool composed of 4 domains: patient selection, index test, reference standard, and flow and timing. The risk of bias is assessed in each of the domains, while the first 3 domains are also assessed for applicability by indicating a "low," "high," or "unclear" rating. Applicability in the QUADAS-2 refers to whether certain aspects of an individual study are matching or not matching the review question. The QUADAS-2 utilizes an overall judgment of "low," "high," or "unclear" risk. An overall risk rating of "low risk of bias" or "low concern regarding applicability" requires the study to be ranked as low on all relevant domains. A high or unclear rating in 1 or more domains may require that the study be rated as "at risk of bias" or having "concerns regarding applicability."[35]

Data Extraction and Analysis

All authors independently extracted information and data regarding the pathology, study population, settings, diagnostic reference standard, and clinical and imaging test accuracy, including number of true positives, false positives, false negatives, and true negatives, for calculation of sensitivity, specificity, positive likelihood ratio (+LR), and negative likelihood ratio (–LR) when not provided. Extracted data were reviewed and confirmed by a second independent author.

To determine posttest probability with 95% confidence intervals (CIs) using LRs as described by Jaeschke et al,[15] we first identified pretest probability. It represents the probability that a specific patient with a specific past history, presenting in a specific clinical setting with a specific symptom complex, has a specific pathology.[15] Posttest probability can be altered to a minimal degree with +LRs of 1.0 to 2.0 and –LRs of 0.5 to 1.0, to a small degree with +LRs of 2.0 to 5.0 and –LRs of 0.2 to 0.5, to a moderate degree with +LRs of 5.0 to 10.0 and –LRs of 0.1 to 0.2, and to a large and almost conclusive degree with +LRs greater than 10.0 and –LRs less than 0.1.[15] The pretest probability (prevalence) was calculated by adding the number of true-positive and false-negative findings for each test or measure and study. The posttest probability of instability or dysplasia was calculated utilizing pretest probability (prevalence) and the +LR. The posttest probability of not having symptomatic hip dysplasia or instability was calculated utilizing pretest probability and the –LR. The probability shifts were reported as the difference in pretest-to-posttest probability for both hip dysplasia or instability presence and absence as a result of utilizing the particular modality in each study.

Heterogeneity of data (eg, different reference standards, different clinical and imaging examinations) precluded meta-analysis. Only 2 studies assessed clinical examinations, each assessing different clinical examinations. The 5 studies assessing imaging examinations either used different imaging measures or lacked consistency in the description of acetabular dysplasia.[30]

RESULTS

Selection of Studies and Methodological Assessment

The systematic search identified 7[2,11,13,18,28,30,41] studies that met the inclusion criteria after review of the title, abstract, and full text (**FIGURE 1**). Conflicts of interest for included studies are provided in **APPENDIX C** (available at www.jospt.org). Two studies[13,28] reported on the diagnostic accuracy of clinical tests to diagnose hip instability. Five[2,11,18,30,41] studies reported on the diagnostic accuracy of radiological measures to diagnose hip dysplasia. Interrater reliability for study inclusion/exclusion between the 2 reviewers was $\kappa = 0.72$ (95% CI: 0.59, 0.85), indicating substantial interrater agreement.

FIGURE 2 presents the methodological assessment of all included studies. **APPENDIX D** (available at www.jospt.org) presents the methodological assessment of the individual included studies based on the QUADAS-2 tool. Interrater reliability between the 2 reviewers for agreement of QUADAS-2 scoring was $\kappa = 0.60$ (95% CI: 0.40, 0.80), indicating moderate agreement.

TABLE 1 presents the characteristics of the 7 included studies.[2,11,13,18,28,30,41] There were 965 subjects across 7 studies investigating hip instability or hip dysplasia

as the source of hip pain. For the 2 studies reporting on clinical tests for hip instability, the sample sizes were 109 and 199 patients, with 62 and 54 cases of hip instability, respectively. For the 5 studies reporting on radiographic measures for hip dysplasia, the sample size ranged from 21 to 241. There were 3[13,18,30] level II studies, 3 level III[11,28,41] studies, and 1 level IV[2] study. The prevalence of pathology ranged from 17%[18] to 80%[41] in individual studies.

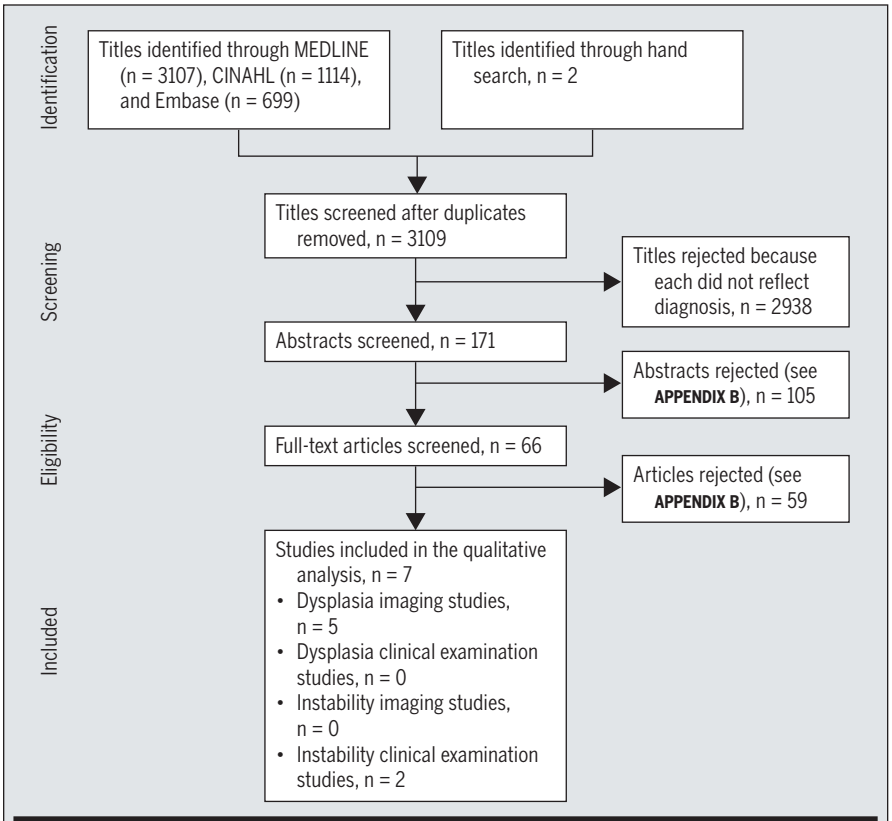


FIGURE 1. PRISMA flow diagram for study inclusion.

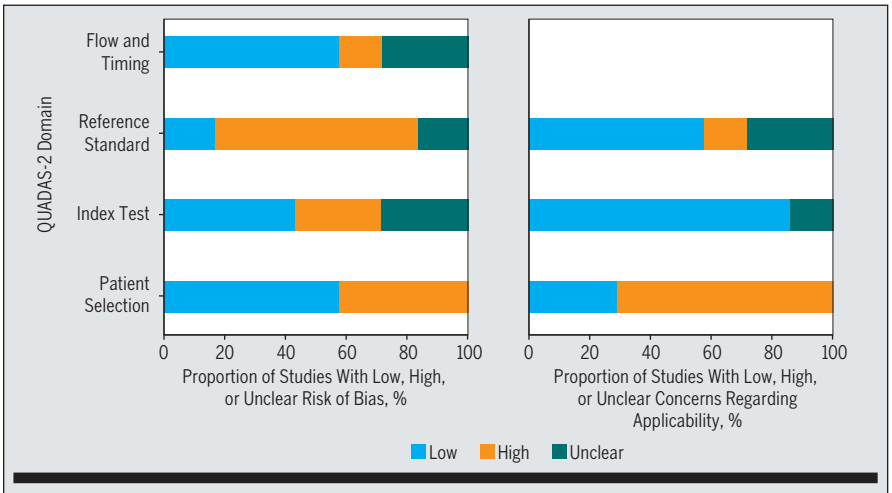


FIGURE 2. Proportion of included studies with low, high, or unclear risk of bias and concerns regarding applicability. Abbreviation: QUADAS-2, Quality Assessment of Diagnostic Accuracy Studies-2 tool.

TABLE 2 presents the clinical tests and radiographic measures assessed in the individual studies. Studies could not be pooled for meta-analysis because of variability between studies. There were 5 different clinical tests for hip instability and 5 different radiological measures for hip dysplasia. Definition use for the diagnoses of hip instability and hip dysplasia, as well as the selected reference standard, was not consistent across all studies. Orthopaedic surgeons interpreted the reference standard in most studies (n = 6), and interrater reliability was only assessed in 2 studies.

Diagnostic Accuracy of Clinical Tests for Hip Instability

TABLE 3 provides the diagnostic accuracy of 5 clinical tests for hip instability assessed in 2 studies.[13,28] Positive LRs to diagnose or rule in hip instability ranged from 2.2 (foot progression angle walking test) to 15.9 (prone instability test) (**FIGURE 3**). Negative LRs to exclude or rule out hip instability ranged from 0.68 (prone instability test) to 0.22 (abduction-hyperextension-external rotation [AB-HEER] test). Only 1 study[13] combined tests, which resulted in an increase in specificity and a decrease in sensitivity.

Diagnostic Accuracy of Radiographic Measures for Hip Dysplasia

TABLE 4 provides the diagnostic accuracy of 5 radiographic measures for hip dysplasia in 5 studies, including the crossover sign,[2] iliofemoral line,[18] Shenton line,[18,30] iliocapsularis-to-rectus femoris ratio,[11] and the Femoro-Epiphyseal Acetabular Roof (FEAR) index.[41] Positive LRs to diagnose or rule in hip dysplasia ranged from 1.2 (Shenton line for borderline dysplasia) to 53.0 (Shenton line for acetabular dysplasia) (**FIGURE 4**). Negative LRs to exclude or rule out hip dysplasia ranged from 0.99 (Shenton line for borderline dysplasia) to 0.17 (Shenton line for acetabular dysplasia). No study combined radiographic measures to improve diagnostic accuracy.

DISCUSSION

The purpose of our systematic review was to summarize the diagnostic accuracy of clinical and imaging tests for the diagnosis of hip instability and hip dysplasia. We found 2 clinical studies reporting on 5 different clinical tests to diagnose hip instability and 5 imaging studies reporting on 5 radiographic measures to diagnose hip dysplasia.

Clinical Tests to Diagnose Hip Instability

Five different clinical tests were described to diagnose hip instability. When positive, the prone instability test resulted in a large shift in probability for the presence of hip instability, from 57% to 95%. When negative, the AB-HEER test resulted in a small shift in probability for the absence of hip instability as the cause of symptoms, from 57% to 22%. This evidence must be interpreted with caution, as the diagnostic statistics from these 2 tests come from only 1 low-quality study.[13]

Radiographic Measures to Diagnose Hip Dysplasia

Five different radiographic measures were described to diagnose hip dysplasia. When positive, the Shenton line resulted in a small to large shift in probability for the presence of hip dysplasia, from 17% to 50% (depending on the definition of dysplasia) to 77% to 91%. When negative, the Shenton line resulted in a minimal to moderate shift in probability for the absence of hip dysplasia as the cause of hip pain, with a post-test probability ranging from 22% to 15%. The evidence for these estimates comes from low-quality studies.[18,30]

Methodological Quality Assessment and Impact on Accuracy

All studies in this review showed some risk of bias and applicability concerns. No study achieved an overall low risk of bias.[35]

The included studies had greater risk of bias compared to concerns regarding applicability (**FIGURE 2**). The imaging studies generally had greater risk of bias than the clinical studies. The greatest risk of bias was relative to reference standard and index test. The imaging study with the largest +LR and lowest −LR also had the greatest risk of bias and concerns regarding applicability.[30]

Risk of bias has been shown to overestimate diagnostic accuracy in previous studies, especially sensitivity.[36] The risk of bias for index tests (imaging studies) and reference standard (both clinical and imaging studies) could affect diagnostic accuracy estimates. Lack of adequate

TABLE 1 Characteristics of Included Studies

Study	Level of Evidence	Study Design/Purpose	Risk of Bias[35]	Applicability Concerns[35]	Study Population
Clinical studies					
Hoppe et al[13]	II	Diagnostic cohort study Diagnostic accuracy of 3 physical examination tests for hip microinstability	High	Unclear	109 patients (65 female); mean age, 27.8 y (range, 13-58 y)
Ranawat et al[28]	III	Prospective cohort study Assess efficacy of FPAW test to identify hip pathology related to FAI or hip instability	Unclear	Unclear	199 patients (114 female); mean ± SD age, 35.4 ± 11.8 y
Imaging studies					
Bellaïche et al[2]	IV	Prospective cohort study Describe radiographic criteria of acetabular dysplasia on standard radiography	High	High	241 patients younger than 50 y; 57 patients for crossover sign
Haefeli et al[11]	III	Prognostic study Assess whether the iliocapsularis-to-rectus femoris ratio for cross-sectional area, thickness, width, and circumference is increased in hip dysplasia	High	High	45 patients (65% female); mean ± SD age, 34 ± 10 y
Kraeutler et al[18]	II	Cohort study Define and validate a novel radiographic parameter (iliofemoral line) in the detection of hip dysplasia	Unclear	High	222 patients (162 female); mean ± SD age, 33.8 ± 11.4 y
Rhee et al[30]	II	Cohort study Determine the reliability of the Shenton line for the diagnosis of hip dysplasia	High	High	128 patients; mean age, 41 y (range, 13-49 y)
Wyatt et al[41]	III	Diagnostic study To compare a new radiographic measurement (FEAR index) with the LCEA and AI in hip dysplasia	High	High	21 patients (61% women); mean age, 31 y

Abbreviations: AI, acetabular index; FAI, femoroacetabular impingement; FEAR, Femoro-Epiphyseal Acetabular Roof; FPAW, foot progression angle walking; LCEA, lateral center-edge angle.

blinding likely overestimates accuracy.[36] Test interpretation with knowledge of index-test findings is a significant bias in diagnostic accuracy studies. A meta-review[23] and 1 review[5] have reported that overall accuracy was higher in the presence of diagnostic review bias (the person interpreting the reference standard was aware of the index-test results). Last, the case-control design found in several imaging studies also overestimates accuracy, because healthy controls have low probability of causing false-positive or false-negative findings.[36]

Overall, improvement in the methodological quality, using appropriate design to assess diagnostic accuracy with adequate blinding and standardization of definition and reference standard, and avoiding a case-control design are required to improve the evidence on the diagnosis of hip instability and hip dysplasia. At this point, we cannot be confident in the precision of diagnostic accuracy estimates for both clinical tests for hip instability and radiographic measures for hip dysplasia. Generally, the CIs were narrow for +LRs, −LRs, and posttest probability, although a few cases were noteworthy, especially with lower-bound estimates for the CIs on +LRs and −LRs (AB-HEER test and prone instability test for clinical examination[13]). Additionally, both the clinical and radiological examinations were assessed in single studies, requiring further studies to validate findings.

TABLE 2 Clinical Tests and Radiographic Measures Utilized

Study	Examination	Dysplasia/Instability Definition	Reference Standard	Interpreting Clinician	Interrater Reliability
Clinical studies					
Hoppe et al[13]	AB-HEER test, prone instability test, HEER test	Criteria for intraoperative diagnosis of hip instability (≥1) • Distraction of the hip under general anesthesia, with body weight traction alone • Adequate distraction, with <11 turns of fine traction • Inability to fully reduce hip after hip is vented • Arthroscopic findings: - Tearing of ligamentum teres - Straight anterior labral tears (4 to 2 o'clock) - Anterior inside-out chondral wear pattern	Diagnosis of hip microinstability made at the time of surgery	Orthopaedic surgeon	NR
Ranawat et al[28]	FPAW test, FABER test	Hip instability: discomfort associated with terminal range of motion; as a result of capsular laxity, structural bony abnormality related to dysplasia; or posttraumatic sequelae leading to subluxation or dislocation Hip dysplasia: LCEA of <25°	History and physical examination in combination with plain-film radiographs	Orthopaedic surgeon	NR
Imaging studies					
Bellaïche et al[2]	Crossover sign (sign of acetabular roof retroversion)	VCEA >20°, indicating insufficient external/anterior coverage of the femoral head	Acetabular retroversion on cross-sectional imaging (arthroscan and/or MRI)	Specialized radiologist	NR
Haefeli et al[11]	Iliocapsularis-to-rectus femoris ratio	LCEA of <25°, with a minimal acetabular index of 14°	MRI	Orthopaedic surgeon	NR
Kraeutler et al[18]	Iliofemoral line, Shenton line	Frank dysplasia: LCEA of <20° Borderline dysplasia: LCEA of 20°-24.9°	Radiograph (AP pelvis)	Orthopaedic surgeon, radiologist	ICC = 0.96-0.99
Rhee et al[30]	Shenton line, "broken" Shenton line (inferior femoral neck projection is cephalad to the superior arch of the obturator foramen)	Center-edge angle of <25° and hip center distance >10 mm, "broken" Shenton line (inferior femoral neck projection is cephalad to the superior arch of the obturator foramen)	Radiograph (AP pelvis)	Orthopaedic surgeons (n = 6)	κ = 0.80 (0.75, 0.84)*
Wyatt et al[41]	FEAR index	LCEA of <25° (hip dysplasia), "broken" Shenton line, migration of femoral head on radiograph, positive crescent sign on MRA	Radiograph (AP pelvis) and MRA	Orthopaedic surgeon	NR

Abbreviations: AB-HEER, abduction-hyperextension-external rotation; AP, anteroposterior; FABER, flexion, abduction, external rotation; FEAR, Femoro-Epiphyseal Acetabular Roof; FPAW, foot progression angle walking; HEER, hyperextension-external rotation; ICC, intraclass correlation coefficient; LCEA, lateral center-edge angle; MRA, magnetic resonance arthrography; MRI, magnetic resonance imaging; NR, not reported; VCEA, vertical-center-external angle.
**Values in parentheses are 95% confidence interval.*

Definition of Pathology

Similar to FAI syndrome,[10] the definitions of symptomatic hip instability and hip dysplasia were variable and complex. The CEA is traditionally regarded as the imaging diagnosis for dysplasia.[31] It was the most consistent measurement as part of the dysplasia pathology definition in the included studies of this review. Three studies utilized a lateral CEA less than 25°,[11,28,41] 1 discriminated between borderline dysplasia and frank dysplasia with the lateral CEA,[18] 1 utilized the Shenton line,[30] and 1 utilized the vertical-center-external angle.[2] All studies, except 1,[28] utilized multiple imaging measurements for the definition of dysplasia, and none were consistent.

As noted previously, hip dysplasia is also a condition of instability. Traditionally, hip instability has been poorly defined.[6,19] The Ottawa classification for acetabular dysplasia[40] incorporates both clinical and radiographic findings indicative of either anterior or posterior instability. Only 2 studies[13,28] in this review examined hip instability with various clinical examination tests. Both had either unclear or high risk of bias and concerns regarding applicability. One of

TABLE 3 Diagnostic Accuracy of Clinical Tests for Hip Instability*

Measure/Study	Sensitivity, %	Specificity, %	Positive LR	Negative LR	Positive Posttest Probability†/ Probability Shift	Negative Posttest Probability†/ Probability Shift
AB-HEER test						
Hoppe et al[13]	80.6 (70.8, 90.5)	89.4 (80.5, 98.2)	7.6 (3.3, 17.5)†	0.22 (0.13, 0.36)†	Pretest, 57%; posttest, 91% (81%, 96%) Moderate	Pretest, 57%; posttest, 22% (15%, 32%) Small
Prone instability test						
Hoppe et al[13]	33.9 (22.1, 45.7)	97.9 (93.7, 100.0)	15.9 (2.2, 114.2)†	0.68 (0.56, 0.81)†	Pretest, 57%; posttest, 95% (85%, 99%) Large	Pretest, 57%; posttest, 47% (42%, 52%) Minimal
HEER test						
Hoppe et al[13]	71.0 (59.7, 82.3)	85.1 (74.9, 95.3)	4.8 (2.4, 9.6)†	0.34 (0.23, 0.51)†	Pretest, 57%; posttest, 86% (76%, 93%) Small	Pretest, 57%; posttest, 31% (23%, 40%) Small
Combination of hip instability tests (AB-HEER test, prone instability test, HEER test)						
Hoppe et al[13]						
≥1 positive test	87.1 (78.8, 95.4)	78.7 (67.0, 90.4)	4.1 (2.2, 7.7)†	0.16 (0.08, 0.32)†	Pretest, 57%; posttest, 84% (75%, 90%) Small	Pretest, 57%; posttest, 17% (10%, 30%) Moderate
≥2 positive tests	67.7 (56.1, 79.4)	95.7 (91.7, 99.8)	15.9 (4.1, 62.5)†	0.34 (0.23, 0.39)†	Pretest, 57%; posttest, 95% (84%, 99%) Large	Pretest, 57%; posttest, 31% (23%, 39%) Small
All 3 positive tests	30.6 (19.2, 42.1)	97.9 (94.7, 100.0)	14.4 (2.0, 104.8)†	0.71 (0.60, 0.84)†	Pretest, 57%; posttest, 95% (72%, 99%) Large	Pretest, 57%; posttest, 48% (44%, 52%) Minimal
Foot progression angle walking test						
Ranawat et al[28]	67 (53, 79)†	70 (62, 77)†	2.2 (1.6, 3.0)†	0.5 (0.32, 0.71)†	Pretest, 27%; posttest, 45% (38%, 53%) Small	Pretest, 27%; posttest, 15% (11%, 21%) Minimal
FABER test						
Ranawat et al[28]	54 (40, 67)†	90 (84, 94)†	5.4 (3.1, 9.0)†	0.5 (0.39, 0.69)†	Pretest, 27%; posttest, 67% (54%, 78%) Moderate	Pretest, 27%; posttest, 16% (12%, 20%) Small

Abbreviations: AB-HEER, abduction-hyperextension-external rotation; FABER, flexion, abduction, external rotation; HEER, hyperextension-external rotation; LR, likelihood ratio.
**Values in parentheses are 95% confidence interval.*
†Values (in part) calculated by authors of this study.

the studies[13] utilized surgery as a reference standard, although it was unclear whether the reference standard was independent of the clinical examination tests, a risk of incorporation bias potentially inflating sensitivity.[36]

The highest-quality study in this review[28] might be argued to be at risk of bias relative to the reference standard. The combination of subject history, physical examination, and radiographs was utilized as a reference standard. While this could describe differential verification bias, where more than 1 reference standard is used, a diagnosis (similar to FAI syndrome)[10] is likely to require a combination of different domains to be appropriate. Additionally, this study utilized a combined reference standard rather than separate reference standards. However, the concern is that the description of "discomfort associated with terminal range of motion, as a result of capsular laxity, structural bony abnormality related to dysplasia, or posttraumatic sequelae leading to subluxation or dislocation," while inclusive, is compared to a reference standard of a lateral CEA of less than 25°. Though it is a commonly accepted reference standard for dysplasia,[31] it is unclear whether it is able to define instability without clinical signs and symptoms.

Overall, our systematic review identified various pathology definitions for both hip instability and hip dysplasia. Standardization will be required to improve subsequent studies in this field.

Limitations

Due to heterogeneity of index-test utilization, we were unable to perform meta-analysis. There was also a lack of comparison of subject inclusion and exclusion across the studies, contributing to variability in the definition of included pathology.

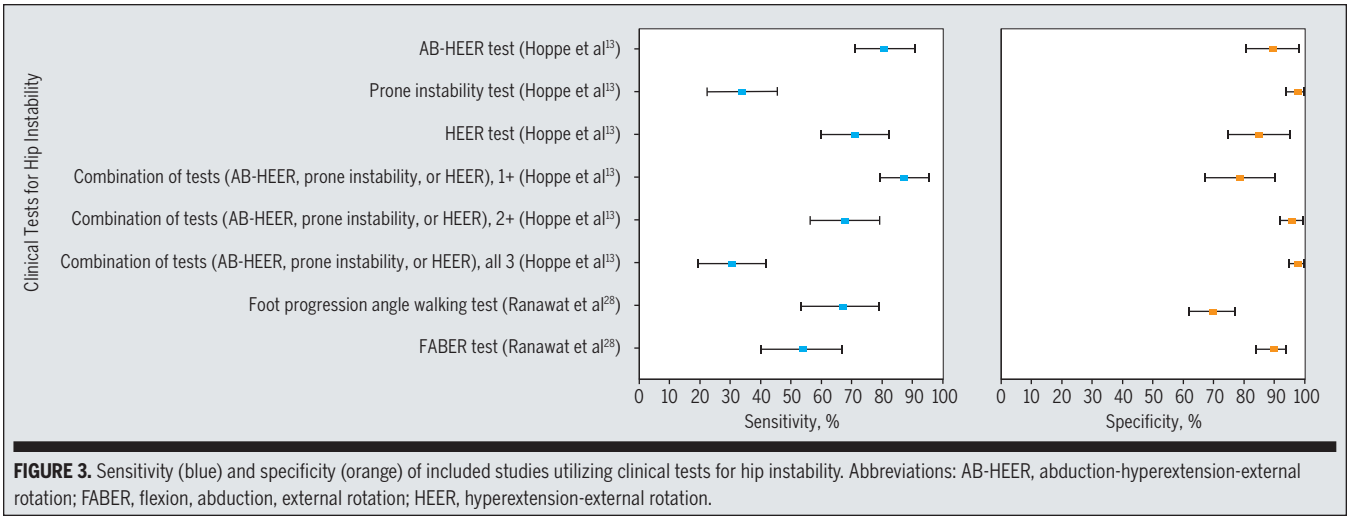


FIGURE 3. Sensitivity (blue) and specificity (orange) of included studies utilizing clinical tests for hip instability. Abbreviations: AB-HEER, abduction-hyperextension-external rotation; FABER, flexion, abduction, external rotation; HEER, hyperextension-external rotation.

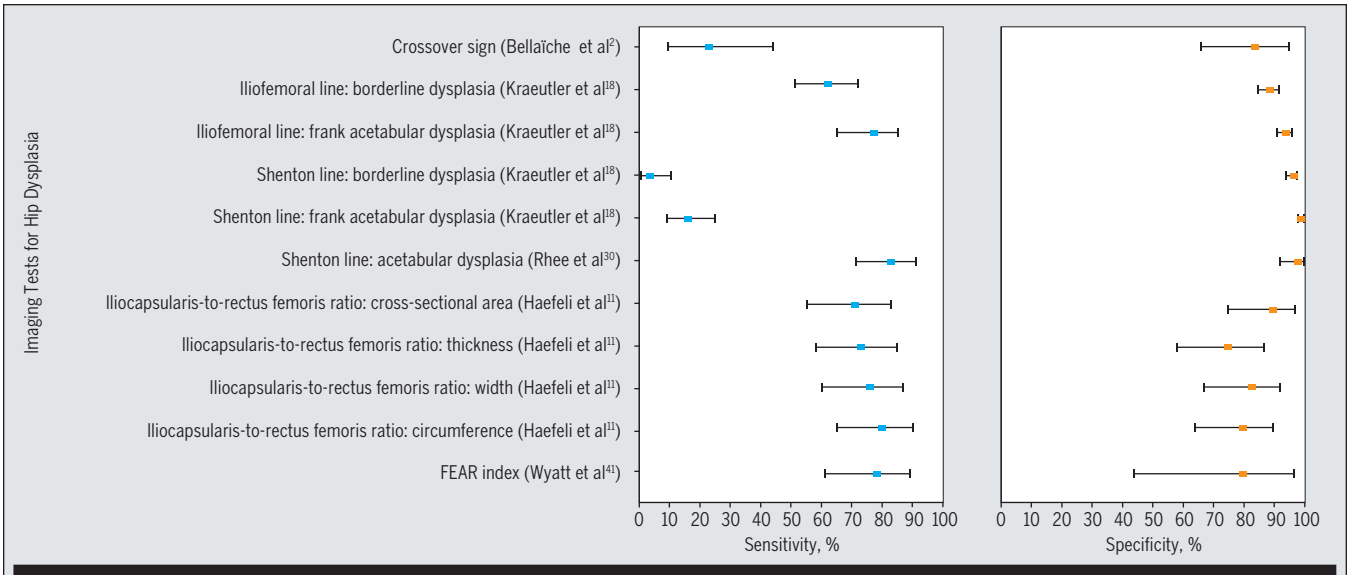


FIGURE 4. Sensitivity (blue) and specificity (orange) of included studies utilizing imaging tests for hip dysplasia. Abbreviation: FEAR, Femoro-Epiphyseal Acetabular Roof.

CONCLUSION

This systematic review summarizes the diagnostic accuracy of clinical tests for hip instability and radiographic measures for hip dysplasia. These tests and measures may support the differential diagnosis to identify the cause of hip pain. However, current evidence is scarce and based primarily on low-quality studies.

KEY POINTS

FINDINGS: The clinical examination can potentially result in small to substantial shifts in probability of having hip instability and minimal to moderate shifts in

TABLE 4 Diagnostic Accuracy of Single Studies Investigating Diagnostic Imaging*

Measure/Study	Sensitivity, %	Specificity, %	Positive LR	Negative LR	Positive Posttest Probability†/ Probability Shift	Negative Posttest Probability†/ Probability Shift
Crossover sign						
Bellaïche et al[2]	23 (9, 44)†	84 (66, 95)†	1.4 (0.49, 4.2)†	0.92 (0.71, 1.19)†	Pretest, 45%; posttest, 55% (29%, 78%) Minimal	Pretest, 45%; posttest, 44% (37%, 50%) Minimal
Iliofemoral line						
Kraeutler et al[18]						
Borderline dysplasia (15%-22% medialization)	62 (51, 72)	89 (85, 92)	5.6 (4.0, 7.9)†	0.43 (0.32, 0.57)†	Pretest, 19%; posttest, 56% (48%, 64%) Moderate	Pretest, 19%; posttest, 9% (7%, 12%) Small
Frank acetabular dysplasia (>22% medialization)	77 (65, 85)	94 (91, 96)	13 (8.4, 20)†	0.24 (0.16, 0.37)†	Pretest, 17%; posttest, 73% (64%, 81%) Large	Pretest, 17%; posttest, 5% (3%, 7%) Small
Shenton line						
Kraeutler et al[18]						
Borderline dysplasia (15%-22% medialization)	3.7 (0.1, 10.3)	97 (94, 98)	1.2 (0.35, 4.3)†	0.99 (0.95, 1.0)†	Pretest, 19%; posttest, 22% (7%, 50%) Small	Pretest, 19%; posttest, 18% (18%, 19%) Minimal
Frank acetabular dysplasia (>22% medialization)	16 (8.9, 25)	99 (98, 100)	16 (5.1, 51)†	0.85 (0.77, 0.94)†	Pretest, 17%; posttest, 77% (51%, 91%) Large	Pretest, 17%; posttest, 15% (14%, 16%) Minimal
Rhee et al[30]						
Acetabular dysplasia	83 (71, 91)†	98 (92, 100)†	53.0 (7.6, 371.7)†	0.17 (0.10, 0.30)†	Pretest, 50%; posttest, 91% (81%, 96%) Large	Pretest, 50%; posttest, 22% (15%, 32%) Moderate
Iliocapsularis-to-rectus femoris ratio						
Haefeli et al[11]						
Cross-sectional area	71 (55, 83)	90 (75, 97)	7.1 (2.8, 18)†	0.32 (0.20, 0.51)†	Pretest, 53%; posttest, 89% (76%, 95%) Moderate	Pretest, 53%; posttest, 26% (18%, 36%) Small
Thickness	73 (58, 85)	75 (58, 87)	2.9 (1.7, 5.2)†	0.36 (0.21, 0.60)†	Pretest, 53%; posttest, 77% (65%, 85%) Small	Pretest, 53%; posttest, 29% (19%, 40%) Small
Width	76 (60, 87)	83 (67, 92)	4.3 (2.2, 8.6)†	0.30 (0.14, 0.50)†	Pretest, 53%; posttest, 83% (71%, 91%) Small	Pretest, 53%; posttest, 25% (16%, 36%) Small
Circumference	80 (65, 90)	80 (64, 90)	4.0 (2.1, 7.6)†	0.25 (0.14, 0.46)†	Pretest, 53%; posttest, 82% (70%, 89%) Small	Pretest, 53%; posttest, 22% (14%, 34%) Small
FEAR index (5°)						
Wyatt et al[41]	78 (61, 89)†	80 (44, 97)†	3.9 (1.1, 14)†	0.27 (0.14, 0.54)†	Pretest, 80%; posttest, 94% (81%, 98%) Small	Pretest, 80%; posttest, 51% (35%, 68%) Small

Abbreviations: FEAR, Femoro-Epiphyseal Acetabular Roof; LR, likelihood ratio.
**Values in parentheses are 95% confidence interval.*
†Values (in part) calculated by authors of this study.

probability of not having hip instability. The radiological examination can potentially result in minimal to substantial shifts in probability of having hip dysplasia and minimal to moderate shifts in probability of not having hip dysplasia. **IMPLICATIONS:** Both clinical and radiological examination tests were better able to improve diagnosis existence than nonexistence. The clinical tests included in this review are easy to perform and could provide the clinician with the ability to increase probability of the existence of hip instability between a minimal and a moderate degree. The radiographic measures included in this review are more complex and detailed, requiring clinician expertise in this area. **CAUTION:** Conclusions from this study are based primarily on studies of either unclear or high risk of bias and having concerns regarding applicability.

ACKNOWLEDGMENTS: *The authors thank Leila Ledbetter, MLIS for assisting with the literature search for this study.*

REFERENCES

1. Armbuster TG, Guerra J, Resnick D, et al. The adult hip: an anatomic study. Part I: the bony landmarks. *Radiology*. 1978;128:1-10. https://doi.org/10.1148/128.1.1
2. Bellaïche L, Lequesne M, Gedouin JE, Laude F, Boyer T, French Arthroscopy Society. Imaging data in a prospective series of adult hip pain in under-50 year-olds. *Orthop Traumatol Surg Res*. 2010;96:S44-S52. https://doi.org/10.1016/j.otsr.2010.09.008
3. Beltran LS, Rosenberg ZS, Mayo JD, et al. Imaging evaluation of developmental hip dysplasia in the young adult. *AJR Am J Roentgenol*. 2013;200:1077-1088. https://doi.org/10.2214/AJR.12.9360
4. Birrell F, Silman A, Croft P, et al. Syndrome of symptomatic adult acetabular dysplasia (SAAD syndrome). *Ann Rheum Dis*. 2003;62:356-358. https://doi.org/10.1136/ard.62.4.356
5. Brealey SD, Scally AJ, Hahn S, Godfrey C. Evidence of reference standard related bias in studies of plain radiograph reading performance: a meta-regression. *Br J Radiol*. 2007;80:406-413. https://doi.org/10.1259/bjr/41006673
6. Dangin A, Tardy N, Wettstein M, May O, Bonin N. Microinstability of the hip: a review. *Orthop Traumatol Surg Res*. 2016;102:S301-S309. https://doi.org/10.1016/j.otsr.2016.09.002
7. Doust JA, Pietrzak E, Sanders S, Glasziou PP. Identifying studies for systematic reviews of diagnostic tests was difficult due to the poor sensitivity and precision of methodologic filters and the lack of information in the abstract. *J Clin Epidemiol*. 2005;58:444-449. https://doi.org/10.1016/j.jclinepi.2004.09.011
8. Fredensborg N. The CE angle of normal hips. *Acta Orthop Scand*. 1976;47:403-405. https://doi.org/10.3109/17453677608988709
9. Genda E, Iwasaki N, Li G, MacWilliams BA, Barrance PJ, Chao EY. Normal hip joint contact pressure distribution in single-leg standing—effect of gender and anatomic parameters. *J Biomech*. 2001;34:895-905. https://doi.org/10.1016/S0021-9290(01)00041-0
10. Griffin DR, Dickenson EJ, O'Donnell J, et al. The Warwick Agreement on femoroacetabular impingement syndrome (FAI syndrome): an international consensus statement. *Br J Sports Med*. 2016;50:1169-1176. https://doi.org/10.1136/bjsports-2016-096743
11. Haefeli PC, Steppacher SD, Babst D, Siebenrock KA, Tannast M. An increased iliocapsularis-to-rectus-femoris ratio is suggestive for instability in borderline hips. *Clin Orthop Relat Res*. 2015;473:3725-3734. https://doi.org/10.1007/s11999-015-4382-y
12. Henak CR, Abraham CL, Anderson AE, et al. Patient-specific analysis of cartilage and labrum mechanics in human hips with acetabular dysplasia. *Osteoarthritis Cartilage*. 2014;22:210-217. https://doi.org/10.1016/j.joca.2013.11.003
13. Hoppe DJ, Truntzer JN, Shapiro LM, Abrams GD, Safran MR. Diagnostic accuracy of 3 physical examination tests in the assessment of hip microinstability. *Orthop J Sports Med*. 2017;5:2325967117740121. https://doi.org/10.1177/2325967117740121
14. Jacobsen S, Sonne-Holm S. Hip dysplasia: a significant risk factor for the development of hip osteoarthritis. A cross-sectional survey. *Rheumatology (Oxford)*. 2005;44:211-218. https://doi.org/10.1093/rheumtology/keh436
15. Jaeschke R, Guyatt GH, Sackett DL. 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
16. Kalisvaart MM, Safran MR. Microinstability of the hip—it does exist: etiology, diagnosis and treatment. *J Hip Preserv Surg*. 2015;2:123-135. https://doi.org/10.1093/jhps/hnv017
17. Kassirer JP. Our stubborn quest for diagnostic certainty. A cause of excessive testing. *N Engl J Med*. 1989;320:1489-1491. https://doi.org/10.1056/NEJM198906013202211
18. Kraeutler MJ, Ashwell ZR, Garabekyan T, et al. The iliofemoral line: a radiographic sign of acetabular dysplasia in the adult hip. *Am J Sports Med*. 2017;45:2493-2500. https://doi.org/10.1177/0363546517708983
19. Kraeutler MJ, Garabekyan T, Pascual-Garrido C, Mei-Dan O. Hip instability: a review of hip dysplasia and other contributing factors. *Muscles Ligaments Tendons J*. 2016;6:343-353. https://doi.org/10.11138/mltj/2016.6.3.343
20. Laborie LB, Engesaeter IØ, Lehmann TG, et al. Radiographic measurements of hip dysplasia at skeletal maturity—new reference intervals based on 2,038 19-year-old Norwegians. *Skeletal Radiol*. 2013;42:925-935. https://doi.org/10.1007/s00256-013-1574-y
21. 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
22. Lee CB, Kim YJ. Imaging hip dysplasia in the skeletally mature. *Orthop Clin North Am*. 2012;43:329-342. https://doi.org/10.1016/j.ocl.2012.05.007
23. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA*. 1999;282:1061-1066. https://doi.org/10.1001/jama.282.11.1061
24. Malviya A, Raza A, Witt JD. Reliability in the diagnosis of femoroacetabular impingement and dysplasia among hip surgeons: role of surgeon volume and experience. *Hip Int*. 2016;26:284-289. https://doi.org/10.5301/hipint.5000349
25. McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, PRISMA-DTA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses of diagnostic test accuracy studies: the PRISMA-DTA statement. *JAMA*. 2018;319:388-396. https://doi.org/10.1001/jama.2017.19163
26. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Int J Surg*. 2010;8:336-341. https://doi.org/10.1016/j.ijsu.2010.02.007
27. Morrison A, Polisena J, Husereau D, et al. The effect of English-language restriction on systematic review-based meta-analyses: a systematic review of empirical studies. *Int J Technol Assess Health Care*. 2012;28:138-144. https://doi.org/10.1017/S0266462312000086
28. Ranawat AS, Gaudiani MA, Slullitel PA, Satalich J, Rebolledo BJ. Foot progression angle walking test: a dynamic diagnostic assessment for femoroacetabular impingement and hip instability. *Orthop J Sports Med*. 2017;5:2325967116679641. https://doi.org/10.1177/2325967116679641
29. Reiman MP, Peters S, Sylvain J, Hagymasi S, Ayeni OR. Prevalence and consistency in surgical outcome reporting for femoroacetabular impingement syndrome: a scoping review. *Arthroscopy*. 2018;34:1319-1328.e9. https://doi.org/10.1016/j.arthro.2017.11.037
30. Rhee PC, Woodcock JA, Clohisy JC, et al. The Shenton line in the diagnosis of acetabular dysplasia in the skeletally mature patient. *J Bone Joint Surg Am*. 2011;93 suppl 2:35-39. https://doi.org/10.2106/JBJS.J.01717
31. Sankar WN, Duncan ST, Baca GR, et al.

Descriptive epidemiology of acetabular dysplasia: the Academic Network of Conservational Hip Outcomes Research (ANCHOR) periacetabular osteotomy. *J Am Acad Orthop Surg*. 2017;25:150-159. https://doi.org/10.5435/JAAOS-D-16-00075

32. Shu B, Safran MR. Hip instability: anatomic and clinical considerations of traumatic and atraumatic instability. *Clin Sports Med*. 2011;30:349-367. https://doi.org/10.1016/j.csm.2010.12.008

33. Swartz MK. The PRISMA statement: a guideline for systematic reviews and meta-analyses. *J Pediatr Health Care*. 2011;25:1-2. https://doi.org/10.1016/j.pedhc.2010.09.006

34. Tannast M, Hanke MS, Zheng G, Steppacher SD, Siebenrock KA. What are the radiographic reference values for acetabular under- and overcoverage? *Clin Orthop Relat Res*. 2015;473:1234-1246. https://doi.org/10.1007/s11999-014-4038-3

35. 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

36. Whiting PF, Rutjes AW, Westwood ME, Mallett S, QUADAS-2 Steering Group. A systematic review classifies sources of bias and variation in diagnostic test accuracy studies. *J Clin Epidemiol*. 2013;66:1093-1104. https://doi.org/10.1016/j.jclinepi.2013.05.014

37. Wiberg G. Studies on dysplastic acetabula and congenital subluxation of the hip joint with special reference to the complication of osteoarthritis. *Acta Chir Scand*. 1939;83:53-68.

38. Wilczynski NL, Haynes RB. Optimal search strategies for detecting clinically sound prognostic studies in EMBASE: an analytic survey. *J Am Med Inform Assoc*. 2005;12:481-485. https://doi.org/10.1197/jamia.M1752

39. Wilczynski NL, Haynes RB, Hedges Team. Developing optimal search strategies for detecting clinically sound prognostic studies in MEDLINE: an analytic survey. *BMC Med*. 2004;2:23. https://doi.org/10.1186/1741-7015-2-23

40. Wilkin GP, Ibrahim MM, Smit KM, Beaulé PE. A contemporary definition of hip dysplasia and structural instability: toward a comprehensive classification for acetabular dysplasia. *J Arthroplasty*. 2017;32:S20-S27. https://doi.org/10.1016/j.arth.2017.02.067

41. Wyatt M, Weidner J, Pfluger D, Beck M. The Femoro-Epiphyseal Acetabular Roof (FEAR) index: a new measurement associated with instability in borderline hip dysplasia? *Clin Orthop Relat Res*. 2017;475:861-869. https://doi.org/10.1007/s11999-016-5137-0

42. Yeung M, Kowalczuk M, Simunovic N, Ayeni OR. Hip arthroscopy in the setting of hip dysplasia: a systematic review. *Bone Joint Res*. 2016;5:225-231. https://doi.org/10.1302/2046-3758.56.2000533

SEND Letters to the Editor-in-Chief

JOSPT welcomes **letters related to professional issues or articles published in the *Journal***. The Editor-in-Chief reviews and selects letters for publication based on the topic's relevance, importance, appropriateness, and timeliness. Letters should include a summary statement of any conflict of interest, including financial support related to the issue addressed. In addition, letters are copy edited, and the correspondent is not typically sent a version to approve. Letters to the Editor-in-Chief should be sent electronically to **jospt@jospt.org**. Authors of the relevant manuscript are given the opportunity to respond to the content of the letter.

APPENDIX A

MEDLINE SEARCH STRATEGY

("Joint Instability"[Mesh] OR instability[tiab] OR unstable[tiab] OR lax[tiab] OR laxity[tiab] OR subluxation[tiab] OR "Hip Dysplasia, Beukes Type" [Supplementary Concept] OR dysplasia[tiab] OR dysplastic[tiab]) AND ("Diagnostic Imaging"[Mesh] OR "diagnostic imaging" [Subheading] OR MRI[tiab] OR "magnetic resonance imaging"[tiab] OR "magnetic resonance arthrography"[tiab] OR "magnetic resonance arthrogram"[tiab] OR MRA[tiab] OR "computed tomography"[tiab] OR ct[tiab] OR bone scan[tiab] OR "radiography"[tiab] OR "radiography"[MeSH Terms] OR radiograph[tiab] OR "plain film"[tiab] OR "x ray"[tiab] OR "x-rays"[MeSH Terms] OR "x-rays"[tiab] OR arthrogram[tiab] OR "bone scan"[tiab] OR "bone scintigraphy"[tiab] OR sonography[tiab] OR "Ultrasonography"[Mesh] OR ultrasound[tiab] OR "Clinical Examination" OR "Clinical Exam"[tiab] OR Physical Examination OR "Physical Exam"[tiab] OR "Orthopedic Examination"[tiab] OR "Orthopedic Exam"[tiab] OR musculo-skeletal examination OR musculoskeletal examination OR musculoskeletal exam OR musculo-skeletal exam OR "Clinical evaluation"[tiab] OR "Physical evaluation"[tiab] OR musculoskeletal evaluation OR musculo-skeletal evaluation OR "Clinical inspection"[tiab] OR "Physical inspection"[tiab] OR musculoskeletal inspection OR musculo-skeletal inspection) AND (Hip[Mesh] OR Hip[tiab] OR hips[tiab] OR "Hip Joint"[Mesh] OR "coxofemoral joint"[tiab]) AND (reliability[tiab] OR accuracy[tiab] OR accurate[tiab] OR Sensitivity[tiab] OR specificity[tiab] OR "Sensitivity and Specificity"[Mesh] OR valid[tiab] OR validity[tiab] OR validation[tiab] OR predict[tiab] OR predictive[tiab] OR predicts[tiab] OR predicted[tiab] OR diagnosis[tiab] OR diagnostic[tiab] OR diagnosed[tiab] OR diagnosis[MeSH] OR diagnosis[sh] OR "diagnostic accuracy"[tiab]) AND (randomized controlled trial[pt] OR controlled clinical trial[pt] OR randomized[tiab] OR randomised[tiab] OR randomization[tiab] OR randomisation[tiab] OR randomly[tiab] OR trial[tiab] OR groups[tiab] OR Clinical trial[pt] OR "clinical trial"[tiab] OR "clinical trials"[tiab] OR "evaluation studies"[Publication Type] OR "evaluation studies as topic"[MeSH Terms] OR "evaluation study"[tiab] OR evaluation studies[tiab] OR "intervention study"[tiab] OR "intervention studies"[tiab] OR "case-control studies"[MeSH Terms] OR "case-control"[tiab] OR "cohort studies"[MeSH Terms] OR cohort[tiab] OR "longitudinal studies"[MeSH Terms] OR "longitudinal"[tiab] OR longitudinally[tiab] OR "prospective"[tiab] OR prospectively[tiab] OR "retrospective studies"[MeSH Terms] OR "retrospective"[tiab] OR "follow up"[tiab] OR "comparative study"[Publication Type] OR "comparative study"[tiab] OR "Cross sectional studies"[mesh] OR "cross sectional"[tiab]) NOT (Editorial[ptyp] OR Letter[ptyp] OR Comment[ptyp]) AND (English[lang] OR Fre[LA])

APPENDIX B

REASONS FOR EXCLUSION

Reasons for exclusion after abstract screening (n = 105)

1. Not a diagnostic study design (n = 68)
2. Not reporting on diagnostic estimates (n = 12)
3. Not the appropriate age group (n = 15)
4. Not a diagnosis of hip dysplasia or instability (n = 4)
5. Study on cadaver (n = 2)
6. Asymptomatic patients (n = 3)
7. Not in English (n = 1)

Reasons for exclusion after full text (n = 59)

1. Not a diagnostic study design (n = 17)
2. Not reporting on diagnostic estimates (n = 31)
3. Not a diagnosis of hip dysplasia or instability (n = 9)
4. Asymptomatic patients (n = 1)
5. Conference abstract (n = 1)

APPENDIX C

CONFLICTS OF INTEREST FOR INCLUDED STUDIES

Study	Reported Conflict of Interest by Study
Hoppe et al[13]	The authors declared that they have no conflicts of interest in the authorship and publication of this contribution
Ranawat et al[28]	One or more of the authors has declared the following potential conflict of interest or source of funding: "A.S.R. is a paid consultant for Arthrex Inc, CONMED Linvatec, DePuy Mitek, and Stryer MAKO and receives IP royalties from ConforMIS"
Bellaïche et al[2]	No conflict-of-interest statement
Haefeli et al[11]	One or more of the authors has received funding from the Deutsche Arthrose-Hilfe e.V. (S.D.S.) and the Swiss National Science Foundation (M.T.)
Kraeutler et al[18]	One or more of the authors has declared the following potential conflict of interest or source of funding: "J.N.O. receives royalties from the Extra Fixation Cup, part of the ADEPT Hip Resurfacing System. O.M.-D. holds stock or stock options in MITA and is a paid consultant for and receives research support from Smith & Nephew"
Rhee et al[30]	The authors did not receive any outside funding or grants in support of their research for or preparation of this work. One or more of the authors, or a member of his or her immediate family, received, in any 1 year, payments or other benefits in excess of $10 000 or a commitment or agreement to provide such benefits from commercial entities (DePuy, Wright)
Wyatt et al[41]	Each author certifies that he or she, or a member of his or her immediate family, has no funding or commercial associations (eg, consultancies, stock ownership, equity interest, patent/licensing arrangements, etc) that might pose a conflict of interest in connection with the submitted article

APPENDIX D

RISK OF BIAS OF THE STUDIES INCLUDED IN THE REVIEW[35]

	Risk of Bias				Applicability Concerns		
Study	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Clinical studies							
Hoppe et al[13]	L	L	H	U	L	L	U
Ranawat et al[28]	L	L	U	L	L	L	U
Imaging studies							
Bellaïche et al[2]	L	H	H	H	H	L	L
Haefeli et al[11]	H	L	L	L	H	L	L
Kraeutler et al[18]	L	U	U	L	H	L	L
Rhee et al[30]	H	U	H	L	H	U	H
Wyatt et al[41]	H	H	H	U	H	L	L

Abbreviations: H, high risk; L, low risk; U, unclear risk.