





Original Research

Longitudinal Invariance Testing Of The Knee Injury Osteoarthritis Outcome Score For Joint Replacement Scale (KOOS-JR)

Alexandra Dluzniewski¹, Caleb Allred¹ , Madeline P Casanova¹ , Jonathan D Moore¹, Adam C Cady² , Russell T Baker¹ ^a

¹ WWAMI Medical Education, University of Idaho, ² Kaiser Permanente

Keywords: structural validity, psychometrics, arthroplasty, patient reported outcome measures

<https://doi.org/10.26603/001c.86129>

International Journal of Sports Physical Therapy

Vol. 18, Issue 5, 2023

Background

The Knee Osteoarthritis Outcome Score for Joint Replacement (KOOS-JR) is a seven-item patient reported outcome measure used to assess perceived knee health. Though commonly used, the longitudinal psychometric properties of the KOOS-JR have not been established and further characterization of its structural validity and multi-group invariance properties is warranted.

Purpose

The purpose of this study was to evaluate psychometric properties of the KOOS-JR in a large sample of patients who received care for knee pathology.

Study Design

Original research.

Methods

Longitudinal data extracted from the Surgical Outcome System (SOS) database of 13,470 knee pathology patients who completed the KOOS-JR at baseline, three-months, six-months, and one-year. Scale structure was assessed with confirmatory factor analysis (CFA), while multi-group and longitudinal invariance properties were assessed with CFA-based procedures. Latent group means were compared with statistical significance set at $\alpha \leq .05$ and Cohen's d effect size as $d = 0.2$ (small), $d = 0.5$ (medium), and $d = 0.8$ (large).

Results

CFA results exceeded goodness-of-fit indices at all timepoints. Multi-group invariance properties passed test requirements. Longitudinal analysis identified a biased item resulting in removal of item #1; the retained six-item model (KOOS-JR-6) passed longitudinal invariance requirements. KOOS-JR-6 scores significantly changed over time ($p \leq .001$, $M_{diff} = 1.08$, Cohen's $d = 0.57$): the highest scores were at baseline examination and the lowest at 12-month assessment.

Conclusions

The KOOS-JR can be used to assess baseline differences between males and females, middle and older aged adults, and patients receiving total knee arthroplasty or

^a Corresponding Author:
Russell T. Baker, PhD, DAT, AT
University of Idaho
875 Perimeter Drive
Moscow, ID 83844
Email: russellb@uidaho.edu
Phone: 208-885-2065

non-operative care. Caution is warranted if the KOOS-JR is used longitudinally due to potential measurement error associated with item #1. The KOOS-JR-6 may be a more viable option to assess change over time; however, more research is warranted.

Level of Evidence

3

© The Author(s)

INTRODUCTION

Osteoarthritis (OA) is a debilitating disease that causes activity limitations for an estimated 14 million Americans.^{1,2} Most individuals suffering from OA are over the age of 65, and a substantial portion (~three million) are racial or ethnic minorities.^{1,2} Those who suffer from OA experience diminished quality of life (QOL)² and often need to undergo total knee replacement (TKR)¹; many may also suffer from the development of depressive symptoms³ and cardiovascular disease.⁴ Therefore, understanding the patient's perception of their OA treatment is essential to providing quality care for these individuals, while also allowing providers to better support a patient's health status and QOL.

One method for gaining insight into the OA patient experience is the use of patient-reported outcome measures (PROMs). PROMs are often used to assess the patients' perspective of their QOL, pain, symptoms, or functional status,^{5,6} which can then be used by clinicians to inform patient care.⁶⁻⁸ Though the implementation of PROMs to guide care for OA has been recommended, providers must carefully select PROMs with sound psychometric properties to effectively measure the patient experience or assess treatment effectiveness. A specific PROM developed for patients with OA or those who have had a TKR is the Knee Injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS-JR). The KOOS-JR is a seven-item short form developed from the 42-item Knee Injury and Osteoarthritis Outcome Score (KOOS) to reduce patient response burden and improve implementation with OA patients in clinical practice.⁹⁻¹¹

Numerous studies have been performed to validate the KOOS-JR and there has been growing support to extend the use of the KOOS-JR in clinical and academic settings.¹² Initial work has identified moderate (0.46)¹³ to preferred (0.84)⁹ internal consistency, while construct validity was established by correlating the KOOS-JR to the KOOS Pain (Spearman's correlation coefficient 0.89) and KOOS Activities of Daily Living (Spearman's correlation coefficient 0.90) subscales.¹² A recent study utilizing confirmatory factor analysis (CFA) procedures found acceptable structural validity (CFI = 0.976, TLI = 0.964, IFI = 0.976, RMSEA = 0.067¹⁴; and invariant multi-group solutions, indicating the KOOS-JR can be used to measure differences across certain sub-groups (i.e., sex, older adults, intervention groups).¹⁴ However, concerns with ceiling effects,¹⁵ an inability to detect differences between groups,¹⁵ validity concerns when used in later stages of recovery,¹³ and an inability to assess outcomes in younger active patients¹² suggests further assessment is needed. Another substantial concern is the lack of longitudinal assessment (e.g., CFA at multiple

time points, longitudinal invariance testing) necessary for establishing scale measurement properties to guide scale use to assess group differences and patient recovery over time.¹⁶⁻¹⁸

Further CFA assessment using responses from repeated patient assessment would help to establish the latent structure and structural validity of the KOOS-JR across time, while invariance (i.e., multi-group and longitudinal) testing would ensure the instrument is valid across groups and time for assessing group differences or change over time.¹⁶⁻¹⁸ Specifically, performing CFAs across multiple administrations would benefit clinicians and researchers by confirming scale structural validity across repeated use to address concerns that scale structure was biased by the timing of scale administration.¹⁶⁻¹⁸ Multi-group invariance testing would provide additional evidence that the items and dimensions were being operationalized in a similar fashion across sub-groups of the population (e.g., do males and females interpret the items in a similar fashion), which allows for substantive research questions regarding group differences to be answered when using the scale.¹⁶⁻¹⁸ Finally, longitudinal invariance testing establishes if the items and latent constructs are adequately measured (i.e., the items and constructs being operationalized similarly) across repeated testing to ensure participant response change is not a byproduct of item bias or measurement error, which allows the KOOS-JR to be used to assess perceived knee health at various stages throughout the injury recovery process.¹⁶⁻¹⁸

While the KOOS-JR is widely used in research and clinical practice, further assessment of the multigroup and longitudinal psychometric properties of the KOOS-JR is warranted to conduct these additional analyses to further establish KOOS-JR measurement properties.¹⁵⁻¹⁹ Therefore, the purpose of this study was to evaluate psychometric properties of the KOOS-JR in a large sample of patients who received care for knee pathology. This occurred in three steps: (1) perform CFAs in a large and diverse patient population to further evaluate the structural validity of the KOOS-JR across multiple assessments; (2) conduct multi-group invariance testing to confirm the validity of the KOOS-JR in specific sub-groups; and (3) perform longitudinal invariance testing to establish the longitudinal properties of the KOOS-JR for use across time to assess if the scale can be used to measure improvement across repeated measures.

METHODS

DATA SOURCE

The Surgical Outcome System²⁰ is an international deidentified patient-reported outcome database that adheres to the Health Insurance Portability and Accountability Act (HIPPA) and has already received IRB approval. The SOS allows for retrospective analysis of the collected data from patients who provide informed consent for data use. The University Institutional Review Board (IRB) indicated IRB approval was not required as the deidentified dataset was not considered human-subject research; IRB approval was granted from the Cedar-Sinai Office of Research Compliance and Quality Improvement as part of a larger research project utilizing SOS data. The dataset used included KOOS-JR responses at four time points: 1) baseline, prior to receiving care (i.e., knee arthroplasty, non-operative care), 2) three-months post-intervention, 3) six-months post-intervention, and 4) one-year post-intervention.

INSTRUMENTATION

The KOOS-JR⁹ is comprised of seven items to assess stiffness, pain, and function of the knee [9]. Patients respond to items using a 5-point Likert scale (none = 0, mild = 1, moderate = 2, severe = 3, extreme = 4). The KOOS-JR is scored by summing the raw scores (0-28); higher scores correspond to worse knee health (i.e., 0 = "perfect knee health"; 28 = "total knee disability"). KOOS-JR scores may also be converted to an interval score (0-100), where a converted interval score of 100 represents "perfect knee health" while a score of 0 represents "total knee disability".⁹ Raw scores (i.e., Likert scale responses) and the 0-28 scale were used for the purposes of this study.

STATISTICAL ANALYSIS

All data and demographic information were extracted from the SOS database in Excel and uploaded to the Statistical Package for Social Sciences (IBM SPSS Statistics for Windows, Version 27.0. Armonk, NY: IBM Corp) and Analysis of Moment Structure (AMOS, SPSS, Inc.) Version 27 for data analysis. Individuals with incomplete KOOS-JR responses at baseline evaluation were removed from the dataset; however, responses from individuals missing only demographic information were retained for analysis. Individuals who did not respond to the KOOS-JR at all time points were used for initial analysis but were excluded from longitudinal analyses. The dataset was then assessed for outliers across all time points: univariate and multivariate outliers were assessed using z-scores (± 3.3) and Mahalanobis distance (cases with a $p < 0.001$ according to the Chi-square test) were removed from the dataset. Data normality was also assessed using histograms and descriptive statistics (i.e., skewness and kurtosis values).

SCALE STRUCTURE

Scale structure of the KOOS-JR was assessed using AMOS to conduct a CFA at each time point. The CFA was specified

as a unidimensional seven-item factor.¹⁴ Model fit was assessed with the following *a priori* criteria^{18,21,22}: Comparative Fit Index (CFI; ≥ 0.95), Tucker-Lewis Index (TLI; ≥ 0.95), Standardized Root Mean Square Residual (SRMR ≤ 0.08), Root Mean Square Error of Approximation (RMSEA ≤ 0.06), and Bollen's Incremental Fit Index (IFI; ≥ 0.95). Greater weight in assessment of model fit was given to CFI and SRMR because those criteria are less susceptible to effects from the small degrees of freedom present when performing CFA on the KOOS-JR.²³ Model fit was also assessed by considering localized areas of strain, as well as the interpretability, size, and statistical significance of the model's parameter estimates (i.e., factor variances, covariances, and indicator errors).¹⁷

MULTI-GROUP INVARIANCE TESTING

Multi-group invariance testing was performed to assess whether items were being interpreted equally across subgroups (i.e., age, sex, knee group) at the initial examination (i.e., baseline exam; time point 1). Multi-group invariance testing was completed across a multi-step process where each step was progressively more restricted¹⁸: configural model (i.e., to assess equal factor structure), metric model (i.e., to assess equal factor loadings), and scalar model (i.e., to assess equal loadings and intercepts). The CFI difference test (CFI_{diff}) and Chi-square difference test (χ^2_{diff}) were used to assess invariance. Model fit was considered adequate at each step if CFI_{diff} was ≤ 0.01 when compared back to the configural model. While the χ^2_{diff} was assessed with each model, CFI_{diff} was given greater weight in assessing model fit because of the sensitivity of χ^2_{diff} with large sample sizes.^{18,21} Thus, if a model exceeded the χ^2_{diff} test recommendation but passed the CFI_{diff} test, invariance testing procedures continued. If the measurement properties met the criteria, then substantive analyses (e.g., comparing group latent means) were performed. Latent group means were compared with statistical significance set at $p \leq .05$; Cohen's d effect sizes were calculated and evaluated using the guidelines of $d = 0.2$ as a small effect size, $d = 0.5$ as a medium effect size, and $d = 0.8$ as a large effect size.²⁴

LONGITUDINAL INVARIANCE TESTING

Longitudinal invariance testing was also performed using the analysis procedures outlined in the multi-group invariance section; however, the analysis was now performed to confirm similar interpretation of items and common factors across time points for all participants. If all tested measurement parameters (e.g., metric, intercepts) met the criteria, the model was further tested to assess if substantive properties (e.g., change over time) could be evaluated, allowing for assessment of KOOS-JR scores over time (e.g., did scores change from baseline to 12-months post-arthroplasty). Latent group means were compared with statistical significance set at $p \leq .05$; Cohen's d effect size was calculated and evaluated using the guidelines of $d = 0.2$ as a small effect size, $d = 0.5$ as a medium effect size, and $d = 0.8$ as a large effect size.²⁴

Table 1. Demographics

Characteristics	N	%
Sex		
Male	5,354	40.9
Female	7,189	54.9
Unknown	548	4.2
Age		
<18 years	7	0.1
18-25 years	16	0.1
26-40 years	114	0.9
41-65 years	6933	53.0
66+ years	5728	43.8
Unknown	293	2.2
Knee Intervention Classification		
Arthroplasty	11,268	86.1
Non-operative	1,823	13.9
Unknown	0	0.0

RESULTS

Of the 13,470 cases, five had missing data and 374 were flagged as univariate and multivariate outliers across all time points. A total of 379 cases, which consisted of 295 (78.9%) arthroplasty knee group participants, 208 (58.6%) females (mean age of 63.30 ± 9.57 years), were removed from the dataset, leaving 13,091 cases for analysis. The mean age of the sample was 64.82 ± 9.10 years (range = 12-89 years) with males accounting for 40.9% ($n = 5,354$) and females accounting for 54.9% ($n = 7,189$) of the sample. Additionally, most respondents ($n = 11,268$, 86.1%) were classified in the knee arthroplasty intervention group. In the knee arthroplasty group, 42.7% ($n = 4,601$) were males and 57.3% ($n = 6,171$) were females; and 52.7% ($n = 5,802$) were in the middle-aged adult group (i.e., 41-65 years) and 46.7% ($n = 5,146$) were in the older aged adult group (i.e., 66+ years). A full participant demographic breakdown is presented in [Table 1](#).

SCALE STRUCTURE

A total of 13,091 participants completed the KOOS-JR at baseline and were used for analysis. The baseline model met goodness-of-fit indices ($\chi^2 (14) = 801.332$; CFI = 0.970; TLI = 0.954; IFI = 0.970; SRMR = 0.029; RMSEA = 0.066; [Table 2](#)) as did the three-month model (CFI = 0.981; $\chi^2 (14) = 587.407$; TLI = 0.972; IFI = 0.981; SRMR = 0.026; RMSEA = 0.056; [Table 2](#)), the six-month model (CFI = 0.986; $\chi^2 (14) = 537.443$; TLI = 0.979; IFI = 0.986; SRMR = 0.019; RMSEA = 0.053; [Table 2](#)), and the one-year model (CFI = 0.984; $\chi^2 (14) = 757.884$; TLI = 0.976; IFI = 0.984; SRMR = 0.020; RMSEA = 0.064; [Table 2](#)).

MULTI-GROUP INVARIANCE

AGE GROUP ANALYSIS

A total of 12,661 individuals reported their age (middle-aged adults [41-65 years] $n = 6,933$; older adults [66+ years] $n = 5,728$) and were used for analysis. The configural model

(i.e., equal form) goodness-of-fit indices met recommended values (CFI = 0.969; $\chi^2 (28) = 791.22$; RMSEA = 0.046; [Table 3](#)). The metric model (i.e., equal loadings) passed both the χ^2_{diff} test and the CFI_{diff} test, allowing for testing of the equal latent variance model. The equal latent variance model passed both the χ^2_{diff} test and the CFI_{diff} test, indicating variances were equal across groups. The scalar model (i.e., equal indicator intercepts) did not pass the χ^2_{diff} test but passed the CFI_{diff} test. As the CFI_{diff} test was weighted more heavily, the equal means model was tested for substantive group differences. When means were not constrained to be equal, the CFI_{diff} was greater than 0.01, indicating the differences in mean scores between groups was statistically significant. Follow-up analyses indicated the middle-aged adult group had significantly ($p \leq .001$) higher scores (i.e., high "total knee disability"/low knee health) on the KOOS-JR than the older adult group at baseline examination ($M_{\text{diff}} = 0.13$, Cohen's $d = 0.23$).

SEX GROUP ANALYSIS

A total of 12,543 individuals reported their sex (males $n = 5,354$; females $n = 7,189$) and were used for analysis. The mean age of males was 64.82 ± 9.07 with 54.1% in the middle-aged adult group and 44.7% in the older adult age group. The mean age of females was 64.84 ± 9.12 with 54.1% in the middle-aged adult group and 44.9% in the older adult age group. The configural model (i.e., equal form) goodness-of-fit indices met recommended values (CFI = 0.970; $\chi^2 (28) = 752.187$; RMSEA = 0.045; [Table 4](#)). The metric model (i.e., equal loadings) did not pass the χ^2_{diff} test, but passed the CFI_{diff} test, allowing for testing of an equal latent variance model. The equal latent variance model did not pass the χ^2_{diff} test, but passed the CFI_{diff} test, indicating variances were equal across groups ([Table 4](#)). The scalar model (i.e., equal indicator intercepts), did not pass the χ^2_{diff} test, but passed the CFI_{diff} test allowing for testing of the equal latent means model. When means were not constrained to be equal, the CFI_{diff} was greater than 0.01, indicating the differences in mean scores between groups was statistically significant. Follow-up analy-

Table 2. Goodness-of-Fit Indices for KOOS-JR at Each Time Point

	χ^2	df ^a	CFI	TLI	IFI	SRMR	RMSEA
Baseline (n = 13091)	801.332	14	0.970	0.954	0.970	0.029	0.066
3 months (n = 13091)	587.407	14	0.981	0.972	0.981	0.022	0.056
6 months (n = 13091)	537.443	14	0.986	0.979	0.986	0.019	0.053
12 months (n = 13091)	757.884	14	0.984	0.976	0.984	0.020	0.064

CFI= Confirmatory Factor Analysis, TLI= Tucker Lewis index, IFI= Bollen's Incremental Fit Index, SRMR= Standardized Root Mean Square Residual, RMSEA= Root Mean Square Error of Approximation

Table 3. Goodness-of-Fit Indices for Measurement Invariance Analyses Across Age Groups

	χ^2	df ^a	χ^2 difference (df)	CFI	CFI difference	TLI	IFI	SRMR	RMSEA
Middle aged adults (n = 6,933)	444.623	14	b	0.969	b	0.954	0.969	0.290	0.067
Older adults (n = 5,728)	346.499	14	b	0.968	b	0.952	0.968	0.030	0.064
Configural (equal form)	791.22	28	b	0.969	b	0.953	0.969	0.029	0.046
Metric (equal loadings)	801.626	34	10.406 (6)	0.969	0.000	0.961	0.969	0.029	0.042
Equal factor variances	806.437	35	15.21(7)	0.968	0.001	0.962	0.968	0.030	0.042
Scalar (equal indicator intercepts)	942.582	40	151.362(12)	0.963	0.006	0.961	0.963	0.029	0.042
Equal Means	1084.847	41	293.627(13)	0.957	0.012	0.956	0.957	0.029	0.045

^a df = degrees of freedom

^b Indicates the value is not calculated at this step.

^c Indicates the model did not pass invariance criteria.

CFI= Confirmatory Factor Analysis, TLI= Tucker Lewis index, IFI= Bollen's Incremental Fit Index, SRMR= Standardized Root Mean Square Residual, RMSEA= Root Mean Square Error of Approximation

Bold indicates that χ^2 difference criterion was exceeded.

Table 4. Goodness-of-Fit Indices for Measurement Invariance Analyses Across Sex

	χ^2	Df ^a	χ^2 difference (df)	CFI	CFI difference	TLI	IFI	SRMR	RMSEA
Males (n = 5,354)	299.847	14	b	0.974	b	0.962	0.974	0.027	0.062
Females (n = 7,189)	452.341	14	b	0.967	b	0.950	0.967	0.030	0.066
Configural (equal form)	752.187	28	b	0.970	b	0.955	0.970	0.027	0.045
Metric (equal loadings)	772.936	34	20.749 (6)	0.970	0.000	0.963	0.970	0.028	0.042
Equal factor variances	778.196	35	26.009 (7)	0.969	0.001	0.963	0.969	0.029	0.041
Scalar (equal indicator intercepts)	936.819	40	184.632 (12)	0.963	0.007	0.961	0.963	0.028	0.042
Equal Means	1131.552	41	379.365(13)	0.955	0.015	0.954	0.955	0.029	0.046

^a df = degrees of freedom

^b Indicates the value is not calculated at this step.

^c Indicates the model did not pass invariance criteria.

CFI= Confirmatory Factor Analysis, TLI= Tucker Lewis index, IFI= Bollen's Incremental Fit Index, SRMR= Standardized Root Mean Square Residual, RMSEA= Root Mean Square Error of Approximation

Bold indicates that χ^2 difference criterion was exceeded.

ses indicated the female group had significantly ($p \leq .001$) higher scores (i.e., higher "total knee disability"/lower knee health) on the KOOS-JR than the male group at baseline examination ($M_{\text{diff}} = 0.15$, Cohen's $d = 0.27$).

INTERVENTION GROUP ANALYSIS

Because the knee arthroplasty group included 86.1% of the total sample and invariance testing recommendations include having subgroups with a similar number of participants in each group [16,18], a random subsample of the knee arthroplasty group was selected. A total of 2,636 par-

Table 5. Goodness-of-Fit Indices for Measurement Invariance Analyses Across Intervention Group

	χ^2	df ^a	χ^2 difference (df)	CFI	CFI difference	TLI	IFI	SRMR	RMSEA
Knee arthroplasty (n = 1,363)	105.924	14	b	0.973	b	0.959	0.973	0.029	0.069
Knee non-operative (n = 1,273)	84.779	14	b	0.976	b	0.965	0.977	0.027	0.063
Configural (equal form)	190.703	28	b	0.974	b	0.962	0.975	0.029	0.047
Metric (equal loadings)	218.248	34	27.545 (6)	0.971	0.003	0.964	0.971	0.032	0.045
Equal factor variances	335.489	35	144.786 (7)	0.970	0.004	0.964	0.970	0.038	0.045
Scalar (equal indicator intercepts)	269.554	40	78.851 (12)	0.964	0.010	0.962	0.964	0.033	0.047
Equal Means	484.382	41	293.679(13)	0.930	0.044	0.929	0.930	0.037	0.064

^a df = degrees of freedom^b Indicates the value is not calculated at this step.^c Indicates the model did not pass invariance criteria.

CFI= Confirmatory Factor Analysis, TLI= Tucker Lewis index, IFI= Bollen's Incremental Fit Index, SRMR= Standardized Root Mean Square Residual, RMSEA= Root Mean Square Error of Approximation

Bold indicates that χ^2 difference criterion was exceeded.

ticipants (i.e., knee arthroplasty n = 1,363; knee non-operative n = 1,273) were used for analysis. The knee arthroplasty group was composed of 643 males (51.4%) and 608 females (48.6%) with 50.0% (n = 682) in the middle-aged adult group (i.e., 44-65 years) and 50.0% (n = 681) in the older adult group (i.e., 66+). The knee non-operative group was composed of 627 males (51.4%) and 594 females (48.6%), with 48.0% (n = 602) in the middle-aged adult group and 46.4% (n = 582) in the older adults group.

The configural model (i.e., equal form) goodness-of-fit indices met recommended values (CFI = 0.974; χ^2 (28) = 190.703; SRMR = 0.029; RMSEA = 0.047; Table 5). The metric model (i.e., equal loadings) did not pass the χ^2_{diff} test, but passed the CFI_{diff} test, which supports testing the equal latent variance model. The equal latent variance model did not pass the χ^2_{diff} test, but passed the CFI_{diff} test, indicating variances were equal between groups. The scalar model (i.e., equal indicator intercepts) did not pass the χ^2_{diff} test but passed the CFI_{diff} test which supports assessing the equal latent means model. When means were not constrained to be equal, the CFI_{diff} criterion was exceeded, indicating the difference in means between groups was statistically significant. Follow-up analyses found that the knee arthroplasty group had significantly higher scores (i.e., higher “total knee disability”/lower knee health) on the KOOS-JR than the knee non-operative group at baseline examination (M_{diff} = 0.38, Cohen's d = 0.60).

LONGITUDINAL INVARIANCE

A total of 13,091 participants completed the KOOS-JR at all four time points and were used for analysis. The configural model (i.e., equal form) goodness-of-fit indices met recommended values (CFI = 0.984; χ^2 (302) = 3208.074; RMSEA = 0.027; Table 6) indicating equal form across repeated assessment. The metric model (i.e., equal loadings) did not pass the χ^2_{diff} test, but passed the CFI_{diff} test, warranting analysis of an equal latent variance model. The equal latent variance model did not pass the χ^2_{diff} test, but passed

the CFI_{diff} test, indicating variances were equal across time. The scalar model (i.e., equal indicator intercepts), exceeded both the χ^2_{diff} test and CFI_{diff} test, which prevents comparison of reported levels of the latent variable (i.e., “total knee disability”) across repeated assessment and suggested item-level bias across repeated use of the scale. Upon inspection of the model, item #1 was found to be the source of non-invariance; when item #1 was not constrained to be equal, the model passed the CFI_{diff} test. Therefore, item #1 (i.e., “How severe is your knee stiffness after first wakening in the morning?”) was identified as a problematic item and was removed; longitudinal invariance was retested with the new structure (i.e., KOOS-JR-6).

The new configural model (KOOS-JR-6) met recommended values (CFI = 0.986; χ^2 (210) = 2388.504; RMSEA = 0.028; Table 7). Both the metric and equal factor variances models exceeded the χ^2_{diff} test but passed the CFI_{diff} test, indicating variances were similar across time. The scalar model exceeded the χ^2_{diff} test, but passed the CFI_{diff} test, allowing assessment of equal latent means. When means were not constrained to be equal, the CFI_{diff} test was exceeded, indicating means were significantly different across time. Follow-up analyses found that at baseline, scores were the highest (i.e., higher “total knee disability”/lower knee health) and group means incrementally decreased (i.e., improved) across time with the lowest group mean scores (i.e., lowest “total knee disability”) being reported at the 12-month assessment (M_{diff} = 1.08, Cohen's d = 0.57).

DISCUSSION

The purpose of this study was to assess the psychometric properties of the KOOS-JR using a large and diverse longitudinal sample of patient responses. Using maximum likelihood CFA, we assessed the structural validity of the KOOS-JR and conducted invariance analysis across groups and time in a large sample of patients who sought care for various knee pathologies to ensure the KOOS-JR can be used

Table 6. Goodness-of-Fit Indices for Measurement Invariance Analyses Across Time Points

	χ^2	df ^a	χ^2 difference (df)	CFI	CFI difference	TLI	IFI	SRMR	RMSEA
Baseline (n = 13091)	801.332	14	b	0.970	b	0.954	0.970	0.029	0.066
3 months (n = 13091)	587.407	14	b	0.981	b	0.972	0.981	0.022	0.056
6 months (n = 13091)	537.443	14	b	0.986	b	0.979	0.986	0.019	0.053
12 months (n = 13091)	757.884	14	b	0.984	b	0.976	0.984	0.020	0.064
Configural (equal form)	3208.074	302	b	0.984	b	0.980	0.984	0.019	0.027
Metric (equal loadings)	3888.173	320	680.099 (18)	0.981	0.003	0.977	0.981	0.023	0.029
Equal factor variances	4196.661	323	988.587 (21)	0.979	0.005	0.975	0.979	0.028	0.030
Scalar (equal indicator intercepts)	6254.532	338	3046.458 (36)^c	0.968	0.016^c	0.964	0.968	0.027	0.037

^a df = degrees of freedom^b Indicates the value is not calculated at this step.^c Indicates the model did not pass invariance criteria.

CFI= Confirmatory Factor Analysis, TLI= Tucker Lewis index, IFI= Bollen's Incremental Fit Index, SRMR= Standardized Root Mean Square Residual, RMSEA= Root Mean Square Error of Approximation

Bold indicates that CFI_{diff} or χ^2 difference criterion was exceeded.**Table 7. Goodness-of-Fit Indices for Measurement Invariance Analyses Across Time Points with Item #1 Removed**

	χ^2	df ^a	χ^2 difference (df)	CFI	CFI difference	TLI	IFI	SRMR	RMSEA
Baseline (n = 13091)	493.932	9	b	0.977	b	0.962	0.977	0.025	0.064
3 months (n = 13091)	435.897	9	b	0.983	b	0.972	0.983	0.022	0.060
6 months (n = 13091)	410.508	9	b	0.987	b	0.979	0.987	0.019	0.058
12 months (n = 13091)	651.245	9	b	0.984	b	0.973	.984	0.021	0.074
Configural (equal form)	2388.504	210	b	0.986	b	0.981	0.986	0.019	0.028
Metric (equal loadings)	2908.794	225	520.29 (15)	0.983	0.003	0.979	0.983	0.023	0.030
Equal factor variances	3168.579	228	780.075 (18)	0.981	0.005	0.977	0.981	0.027	0.031
Scalar (equal indicator intercepts)	4194.661	240	1806.157 (30)	0.974	0.010	0.971	0.974	0.025	0.035
Equal Means	21025.091	243	18,636.587(33)	0.866	0.12	0.847	0.866	0.228	0.081

^a df = degrees of freedom^b Indicates the value is not calculated at this step.^c Indicates the model did not pass invariance criteria.

CFI= Confirmatory Factor Analysis, TLI= Tucker Lewis index, IFI= Bollen's Incremental Fit Index, SRMR= Standardized Root Mean Square Residual, RMSEA= Root Mean Square Error of Approximation

Bold indicates that χ^2 difference criterion was exceeded.

between groups and across time. Contemporary analytic methods were used to assess multi-group and longitudinal model fit and structural validity of the KOOS-JR,^{17,18,21} with the multi-group analysis being conducted in a larger and more heterogeneous population than previously used in the literature.¹⁴ Multi-group and longitudinal invariance results suggest the KOOS-JR demonstrates structural validity and can be used with specific sub-groups of the population (e.g., different sexes, age groups). Longitudinal analysis, however, identified a biased item resulting in a modified

version of the KOOS-JR (i.e., KOOS-JR-6); the modified version met longitudinal analysis recommendations.

STRUCTURAL VALIDITY - CONFIRMATORY FACTOR ANALYSIS

The CFA results indicated sound structural properties of the KOOS-JR in a large, heterogeneous sample of patients who completed the scale during a baseline examination when seeking care. The model fit exceeded recommended fit indices^{18,21}; thus, the findings supported prior Rasch

analysis⁹ and CFA¹⁴ assessment, with a structurally sound unidimensional model at initial (i.e., baseline) patient completion being found. Identification of a sound structural model justified further multi-group and longitudinal invariance testing to further determine scale measurement properties and guide use of the scale in clinical practice and research related to hypothesis testing, assessing group differences, and examining change across time.

MULTI-GROUP INVARIANCE ANALYSIS ACROSS AGE, SEX, AND INTERVENTION GROUPS

The presence of multi-group invariance supports scale use for hypothesis testing (e.g., are levels of “total knee disability” different across sexes), while providing valuable insight on if the scale items or underlying construct (i.e., total knee disability) are being operationalized similarly across the groups.^{17,18,21} The multi-group invariance testing results confirmed the structural validity of the scale across the tested groups (e.g., sex, age groups), which then allowed between group differences to be assessed at baseline examination.^{14,18} The findings provide clinicians and researchers with evidence that identified group differences are true group differences as opposed to differences that may result from measurement error (e.g., how an item is interpreted, how a latent construct is operationalized, etc.).

The results confirmed prior multi-group invariance testing¹⁴ that found the KOOS-JR to be invariant in an older adult population (i.e., 41 year of age or older). Because the KOOS-JR was invariant in middle-aged and older (i.e., 41+ years of age) populations, assessment of latent mean differences was warranted in these groups.^{17,18} We found significant latent mean differences across groups at the baseline examination: the middle-aged adult group (41-66 years) reported substantially higher scores (i.e., worse knee health) than the older adult group (66+ years) at the baseline examination. The results confirmed prior research findings¹⁴ of the middle-aged group reporting higher levels of self-reported knee disability than the older age group on the KOOS-JR. The findings, however, conflict with prior research which indicated older age groups (i.e., 65 years or age or older) perceived greater impairments of knee health on the KOOS^{25,26} and that patient reported functional impairment increases across the life span.²⁵⁻²⁷ The differences could be the result of sample differences or the KOOS-JR having fewer items designed for those who will undergo knee arthroplasty. It may also be important to note that while the difference was statistically significant, the effect size was small, and the group differences may not be that meaningful in clinical practice or research without further research that also considers physical activity levels and how this might influence KOOS or KOOS-JR scores.

Lower levels of physical activity level before the onset of OA or before a total hip arthroplasty (THA) intervention in older (i.e., 66+ years) age groups have been reported.²⁸ Self-reported physical activity decreases related to knee health impairment could be explained by numerous factors (e.g., greater levels of joint degeneration, number of comorbidities, body composition, etc.) in patient population^{28,29} and it is conceivable that increased prevalence of these vari-

ables (e.g., greater knee degeneration, etc.) and decreased physical activity would result in greater perceived impairments in knee health as measured by the KOOS or KOOS-JR. A limitation of the SOS data available for this study was a lack of demographic patient information; thus, further analysis to explore how these variables affected KOOS-JR scores at baseline or across time could not be performed. Further research is warranted to better understand the influence of these variables on KOOS-JR scores prior to or after intervention or the onset of OA. Additional multi-group invariance testing is also warranted with younger patient populations. Assessment of group differences in KOOS-JR scores in patient populations under the age of 41 should be performed with caution because multi-group invariance testing could not be conducted in this population due to insufficient sample sizes in the data.

The multi-group invariance testing between sexes also confirmed prior findings of the KOOS-JR being invariant between males and females.¹⁴ Thus, differences in latent mean scores can be viewed as true group differences as opposed to measurement error, and comparison of group mean scores differences across the sexes is supported.^{14,18} The analysis identified statistically significant group mean differences between males and females: male participants reported lower scores (i.e., less perceived “total knee disability”) than female participants at the baseline examination. The findings support prior KOOS-JR findings of females reporting greater levels of knee disability on the KOOS-JR compared to males.¹⁴ Sex differences at the baseline examination could be related to differences in psychological variables, such as coping strategies. For example, females have been reported to have reduced capacity to cope with musculoskeletal pain and this may explain higher baseline levels of perceived knee health impairment on the KOOS-JR.³⁰ Other research, however, found females to have higher pain acceptance and more social support than males, while males were reported to have higher levels of kinesophobia, more mood disturbances, and lower activity levels than females.³¹

It should be noted that the effect size of the latent mean difference between sexes in the study was small and differences in condition, response to pain, prior treatment adherence, age, and physical activity level could explain the latent mean sex difference.³¹⁻³⁴ The role of physical activity and patient awareness of physical limitations for participating in physical activity (e.g., sports) with an injury or degenerative joint condition might be relevant in understanding this phenomenon. For example, researchers have reported that females have higher levels of physical activity compared to males,³¹ while other researchers indicated males reported higher levels of physical activity before the onset of OA and prior to THA.²⁸ Limitations in the SOS dataset prevent further analysis of the role of these variables (e.g., physical activity levels, coping strategies, etc.) in affecting KOOS-JR scores and further research is warranted to identify when these sex differences occur and better understand how other variables influence or predict KOOS-JR scores.

The findings are also congruent with prior research¹⁴ which indicated the KOOS-JR was invariant when tested with a sample of patients who received knee arthroplasty or non-operative care. The assessment of latent mean scores indicated the arthroplasty group reported higher perceived knee disability (i.e., lower levels of perceived knee health) at baseline than the non-operative group and the difference was statistically significant with a medium effect size. The two groups had similar demographic profiles for age and sex, indicating the identified group difference was unlikely to be explained by sex or age differences. This finding confirms prior research¹⁴ and fits the expectation that those who have knee degeneration and warrant surgical intervention would demonstrate higher scores on the KOOS-JR. The findings provide preliminary evidence that KOOS-JR scores may be elevated in those with more substantial pathology requiring more substantial intervention; however, the demographic information available in the SOS database does not allow for further group comparison (e.g., pathology, pathology severity, length of symptomology, psychosocial variable assessment, etc.) to better understand the variables or antecedents that influence patient responses on the KOOS-JR at baseline examination or for predicting who will respond favorably to specific interventions. Further research into patient perceptions of knee health and relevant variables and antecedents may be useful to determine when these differences arise and what might be the mechanism for these differences. Additional research is also needed to determine if diagnostic-cut-off criteria or other clinical guidelines could be created to aid clinicians in using patient reported scores to inform the intervention decision-making process.

LONGITUDINAL INVARIANCE

The study also provides novel insight into the longitudinal properties of the KOOS-JR and its validity for assessing post-intervention effects across time. Longitudinal invariance was established for the equal forms, equal loadings, and equal variances models, but did not pass the equal intercepts model. Failure to meet this standard indicates the respondents did not interpret the construct (i.e., “total knee disability”) similarly across time, and assessment of changes in mean scores was not warranted without further inspection of the model and individual items. The presence of this finding identifies measurement bias which creates a challenge in assessing levels (“i.e., “amounts”) of knee health/disability over time with the KOOS-JR. Subsequent analysis identified item #1 (i.e., “How severe is your knee stiffness after first wakening in the morning?”) as the problematic item and indicates respondents are not interpreting this item similarly across repeated assessment. Thus, caution is warranted when examining changes across time or patient recovery with the KOOS-JR because score changes may not be the result of change over time or improvement (i.e., healing) from an intervention alone.¹⁸ The removal of the problematic item from the model, however, resulted in a more psychometrically sound scale that met the contemporary recommendations for longitudinal measurement invariance.

The new model (i.e., KOOS-JR-6) was invariant across each step of the longitudinal measurement invariance process, indicating this model can be used to assess changes in patient recovery across time or examine group mean changes across time. Thus, the results supported examining the mean scores across repeated assessment on the KOOS-JR-6 to determine if scores changed after receiving treatment. The findings indicated the participants reported statistically significant and meaningful improvements in knee health across repeated measures: the lowest scores (i.e., highest “total knee disability”) were reported at the baseline examination and the highest scores (i.e., lowest “total knee disability”) occurred at the 12-month follow-up. The KOOS-JR-6 findings provide some support for scale validity as patients who receive surgery or who participate in the rehabilitation process would be expected to identify significant improvement over time, whether from the effects of intervention, placebo, or natural healing. The findings are congruent with researchers^{35,36} who have previously reported patient improvement on the KOOS after patients received care (e.g., arthroplasty, arthroscopy, and exercise therapy, etc.) from six months to two years post-intervention.

LIMITATIONS AND FUTURE RESEARCH

While the current study had many strengths, limitations also existed. For example, lack of complete demographic information from the dataset prohibited analysis of all possible subgroups (e.g., surgical approach; younger populations; ethnicity; socioeconomic status; psychosocial variables, etc.), thus limiting invariance testing across all relevant subgroups and the understanding of potential mechanisms for the identified group differences. Therefore, clinicians and researchers should exercise caution when examining KOOS-JR score group differences for populations where multi-group invariance is not yet established. Further, the lack of other relevant demographic information (e.g., pathology, pathology severity, surgical intervention, or approach, etc.) prevents the completion of other analyses to answer other measurement (e.g., multi-group invariance across surgical approaches) or substantive (e.g., assess differences in intervention effectiveness) questions valuable to research and clinical practice. Finally, despite the strengths of using a large, heterogenous sample of patient responses, it should be noted that instrument validation is a multi-step process. The study provides strong evidence for the tested measurement properties of the KOOS-JR-6; however, further research is necessary to establish other needed scale measurement properties (e.g., responsiveness, reliability, minimal clinically important differences [MCIDs], etc.).

Future research should test multi-group invariance across additional subgroups to further inform use of the KOOS-JR in those populations. Establishing multi-group invariance in other relevant subgroups (e.g., different socioeconomic groups, age groups, activity levels, health literacy levels, pathologies, etc.) could help ensure the scale is appropriate to use across diverse patient populations. Researchers should also confirm the KOOS-JR-6 measure-

ment findings in a cross-validation sample of patients who only respond to those six items to ensure the measurement properties are consistent. Additionally, further psychometric studies should be performed to establish other relevant measurement properties (e.g., MCIDs, responsiveness, internal consistency, reliability, etc.) to inform and guide use of the KOOS-JR and KOOS-JR-6 in research and clinical practice.

CONCLUSIONS

Findings in the present study suggest that the KOOS-JR demonstrates structural validity and can be used to compare patient reported outcomes between sex, age groups (e.g., middle aged vs. older adults), and intervention categories (i.e., arthroplasty vs. non-operative care). However, the KOOS-JR did not demonstrate sound longitudinal measurement invariance; researchers and clinicians who desire to use the scale longitudinally should do so with caution. Longitudinal use of the KOOS-JR should include consideration of how patients conceptualize knee health over time due to differences in patient interpretation of item #1 and its influence on overall KOOS-JR scores. Thus, follow-up

questions of the patient's perception of knee stiffness and its influence on overall knee health is warranted; researchers and clinicians could also choose not to score that item and instead only use the items in the KOOS-JR-6 when examining change in knee health over time. Future research is still needed to establish all the necessary measurement properties for effective use of the KOOS-JR-6 in clinical practice and research.

COMPETING INTERESTS

The authors declare that they have no competing interests.

FUNDING

This publication was supported by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under Grant #P20GM103408 and an Idaho WWAMI Research Training Support Award

Submitted: May 08, 2023 CDT, Accepted: July 19, 2023 CDT



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-NC-4.0). View this license's legal deed at <https://creativecommons.org/licenses/by-nc/4.0> and legal code at <https://creativecommons.org/licenses/by-nc/4.0/legalcode> for more information.

REFERENCES

1. Deshpande BR, Katz JN, Solomon DH, et al. Number of persons with symptomatic knee osteoarthritis in the US: impact of race and ethnicity, age, sex, and obesity. *Arthritis Care Res*. 2016;68(12):1743-1750. [doi:10.1002/acr.22897](https://doi.org/10.1002/acr.22897)
2. Vina ER, Kwok CK. Epidemiology of osteoarthritis: literature update. *Curr Opin Rheumatol*. 2018;30(2):160-167. [doi:10.1097/bor.00000000000000479](https://doi.org/10.1097/bor.00000000000000479)
3. Veronese N, Stubbs B, Solmi M, et al. Association between lower limb osteoarthritis and incidence of depressive symptoms: data from the osteoarthritis initiative. *Age Ageing*. 2017;46(3):470-476. [doi:10.1093/ageing/afw216](https://doi.org/10.1093/ageing/afw216)
4. Schieir O, Tosevski C, Glazier RH, Hogg-Johnson S, Badley EM. Incident myocardial infarction associated with major types of arthritis in the general population: a systematic review and meta-analysis. *Ann Rheum Dis*. 2017;76(8):1396-1404. [doi:10.1136/annrheumdis-2016-210275](https://doi.org/10.1136/annrheumdis-2016-210275)
5. Black N. Patient reported outcome measures could help transform healthcare. *BMJ*. 2013;346(jan28 1):f167-f167. [doi:10.1136/bmj.f167](https://doi.org/10.1136/bmj.f167)
6. Lam KC, Harrington KM, Cameron KL, Valier ARS. Use of patient-reported outcome measures in athletic training: common measures, selection considerations, and practical barriers. *J Athl Train*. 2019;54(4):449-458. [doi:10.4085/1062-6050-108-17](https://doi.org/10.4085/1062-6050-108-17)
7. Jayakumar P, Bozic KJ. Advanced decision-making using patient-reported outcome measures in total joint replacement. *J Orthop Res*. 2020;38(7):1414-1422. [doi:10.1002/jor.24614](https://doi.org/10.1002/jor.24614)
8. Jette AM. Outcomes research: shifting the dominant research paradigm in physical therapy. *Phys Ther*. 1995;75(11):965-970. [doi:10.1093/ptj/75.1.965](https://doi.org/10.1093/ptj/75.1.965)
9. Lyman S, Lee YY, Franklin PD, Li W, Cross MB, Padgett DE. Validation of the KOOS, JR: A Short-form Knee Arthroplasty Outcomes Survey. *Clin Orthop Relat Res*. 2016;474(6):1461-1471. [doi:10.1007/s11999-016-4719-1](https://doi.org/10.1007/s11999-016-4719-1)
10. Lyman S, Lee YY, McLawhorn AS, Islam W, MacLean CH. What are the minimal and substantial improvements in the HOOS and KOOS and JR versions after total joint replacement? *Clin Orthop Relat Res*. 2018;476(12):2432-2441. [doi:10.1097/corr.0000000000000456](https://doi.org/10.1097/corr.0000000000000456)
11. Hung M, Saltzman CL, Greene T, et al. Evaluating instrument responsiveness in joint function: The HOOS JR, the KOOS JR, and the PROMIS PF CAT. *J Orthop Res*. 2018;36(4):1178-1184. [doi:10.1002/jor.23739](https://doi.org/10.1002/jor.23739)
12. Buller LT, McLawhorn AS, Lee YY, Cross M, Haas S, Lyman S. The short form KOOS, JR is valid for revision knee arthroplasty. *J Arthroplasty*. 2020;35(9):2543-2549. [doi:10.1016/j.arth.2020.04.016](https://doi.org/10.1016/j.arth.2020.04.016)
13. Hunnicutt JL, Hand BN, Gregory CM, et al. KOOS-JR demonstrates psychometric limitations in measuring knee health in individuals after ACL reconstruction. *Sports Health*. 2019;11(3):242-246. [doi:10.1177/1941738118812454](https://doi.org/10.1177/1941738118812454)
14. Allred C, Reeves AJ, Casanova MP, Cady AC, Baker RT. Multi-group invariance testing of the knee injury osteoarthritis outcome score for joint replacement scale. *Osteoarthritis Cartilage*. 2022;4(4):100296. [doi:10.1016/j.ocarto.2022.100296](https://doi.org/10.1016/j.ocarto.2022.100296)
15. Eckhard L, Munir S, Wood D, et al. The ceiling effects of patient reported outcome measures for total knee arthroplasty. *Orthop Traumatol Surg Res*. 2021;107(3):102758. [doi:10.1016/j.otsr.2020.102758](https://doi.org/10.1016/j.otsr.2020.102758)
16. Mokkink LB, Prinsen CA, Patrick DL, et al. *COSMIN Study Design Checklist for Patient-Reported Outcome Measurement Instruments*. Amsterdam, The Netherlands; 2019:1-32.
17. Brown T. *Confirmatory Factor Analysis for Applied Research*. 2nd ed. The Guilford Press; 2015.
18. Kline RB. *Principles and Practices of Structural Equation Modelling*. 4th ed. The Guilford Press; 2016.
19. Padilla JA, Rudy HL, Gabor JA, et al. Relationship between the patient-reported outcome measurement information system and traditional patient-reported outcomes for osteoarthritis. *J Arthroplasty*. 2019;34(2):265-272. [doi:10.1016/j.arth.2018.10.012](https://doi.org/10.1016/j.arth.2018.10.012)
20. Surgical Outcome System Website. Published 2023. <https://surgicaloutcomesystem.com/for-patients>
21. Byrne BM. *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming*. 3rd ed. Routledge; 2016. [doi:10.4324/9781315757421](https://doi.org/10.4324/9781315757421)

22. Hu LT, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Model*. 1999;6(1):1-55. [doi:10.1080/10705519909540118](https://doi.org/10.1080/10705519909540118)
23. Shi D, DiStefano C, Maydeu-Olivares A, Lee T. Evaluating SEM model fit with small degrees of freedom. *Multivariate Behav Res*. 2022;57(2-3):179-207. [doi:10.1080/00273171.2020.1868965](https://doi.org/10.1080/00273171.2020.1868965)
24. Cohen J. Statistical power analysis for the behavioral sciences. *Edition*. 1988;2.
25. Paradowski PT, Bergman S, Sundén-Lundius A, Lohmander LS, Roos EM. Knee complaints vary with age and gender in the adult population. Population-based reference data for the Knee injury and Osteoarthritis Outcome Score (KOOS). *BMC Musculoskelet Disord*. 2006;7(1):38. [doi:10.1186/1471-2474-7-38](https://doi.org/10.1186/1471-2474-7-38)
26. Baldwin JN, McKay MJ, Simic M, et al. Self-reported knee pain and disability among healthy individuals: reference data and factors associated with the Knee injury and Osteoarthritis Outcome Score (KOOS) and KOOS-Child. *Osteoarthritis and Cartilage*. 2017;25(8):1282-1290. [doi:10.1016/j.joca.2017.03.007](https://doi.org/10.1016/j.joca.2017.03.007)
27. Jinks C, Jordan K, Croft P. Measuring the population impact of knee pain and disability with the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC). *Pain*. 2002;100(1-2):55-64. [doi:10.1016/s0304-3959\(02\)00239-7](https://doi.org/10.1016/s0304-3959(02)00239-7)
28. Lübbecke A, Zimmermann-Sloutskis D, Stern R, et al. Physical activity before and after primary total hip arthroplasty: a registry-based study. *Arthritis Care Res*. 2014;66(2):277-284. [doi:10.1002/acr.22101](https://doi.org/10.1002/acr.22101)
29. McClung CD, Zahiri CA, Higa JK, Amstutz HC, Schmalzried TP. Relationship between body mass index and activity in hip or knee arthroplasty patients. *J Orthop Res*. 2000;18(1):35-39. [doi:10.1002/jor.1100180106](https://doi.org/10.1002/jor.1100180106)
30. Grossi G, Soares JJF, Lundberg U. Gender differences in coping with musculoskeletal pain. *Int J Behav Med*. 2000;7(4):305-321. [doi:10.1207/s15327558ijbm0704_03](https://doi.org/10.1207/s15327558ijbm0704_03)
31. Rovner GS, Sunnerhagen KS, Björkdahl A, et al. Chronic pain and sex-differences; women accept and move, while men feel blue. *PLoS One*. 2017;12(4):e0175737. [doi:10.1371/journal.pone.0175737](https://doi.org/10.1371/journal.pone.0175737)
32. Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D. Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. *Eur J Pain*. 2006;10(4):287-333. [doi:10.1016/j.ejpain.2005.06.009](https://doi.org/10.1016/j.ejpain.2005.06.009)
33. Gerdle B, Björk J, Henriksson C, Bengtsson A. Prevalence of current and chronic pain and their influences upon work and healthcare-seeking: a population study. *J Rheumatol*. 2004;31(7):1399-1406.
34. Stubbs D, Krebs E, Bair M, et al. Sex differences in pain and pain-related disability among primary care patients with chronic musculoskeletal pain. *Pain Med*. 2010;11(2):232-239. [doi:10.1111/j.1526-4637.2009.00760.x](https://doi.org/10.1111/j.1526-4637.2009.00760.x)
35. Loef M, Gademan MGJ, Latijnhouwers DAJM, et al. Comparison of KOOS scores of middle-aged patients undergoing total knee arthroplasty to the general Dutch population using KOOS percentile curves: The LOAS Study. *J Arthroplasty*. 2021;36(8):2779-2787.e4. [doi:10.1016/j.arth.2021.03.050](https://doi.org/10.1016/j.arth.2021.03.050)
36. Herrlin SV, Wange PO, Lapidus G, Hållander M, Werner S, Weidenhielm L. Is arthroscopic surgery beneficial in treating non-traumatic, degenerative medial meniscal tears? A five year follow-up. *Knee Surg Sports Traumatol Arthrosc*. 2013;21(2):358-364. [doi:10.1007/s00167-012-1960-3](https://doi.org/10.1007/s00167-012-1960-3)