Psychometric Evaluation of the Patient's Knee Implant Performance (PKIP) Questionnaire for the Assessment of Primary Total Knee Arthroplasty

Theresa Coles,[1] Kim Dwyer,[2] Margaret Mordin,[1] Valerie Williams,[1] Mark Clatworthy,[3] Piers Yates,[4] William Hamilton[5]

[1]RTI Health Solutions, Research Triangle Park, NC, United States; [2]DePuy Synthes Joint Reconstruction, Warsaw, IN, United States; [3]Ascot Hospital, Auckland, New Zealand; [4]Fremantle Hospital, Fremantle, Australia; [5]Anderson Orthopedic Research Institute, Alexandria, VA, United States

BACKGROUND

PKIP Description

- Newly developed patient-reported outcome (PRO) measure
- Assesses knee implant functional performance
- For use with patients before and after primary total knee arthroplasty (TKA)
- Assesses factors that lead to patient dissatisfaction and describes unmet needs in knee functional performance
- Addresses stability, motion, satisfaction, and confidence with knee implant
- 24 items, each employing a 5-, 6-, or 11-point ordinal response scale, with higher values indicating better knee stability (except for two items that are reverse scored)

PKIP Development

- The PKIP has undergone two phases of development:
  - Phase 1: Conceptual model, literature review, focus groups, and in-depth interviews described in Lewis et al. (2014)[1]
  - Phase 2: Psychometric evaluation

OBJECTIVE

- To evaluate the psychometric properties of the PKIP (phase 2 of the PKIP development)

METHODS

Study Background

- Prospective, multicenter, nonrandomized, noncomparative, longitudinal study to gather clinical and PRO data regarding knee products currently on the market

Participants

- Adults (aged 22 to 80) with noninflammatory degenerative joint disease who were candidates for primary TKA

Design

- 20 international sites (English-speaking countries/patients)
- Patients given one of four total knee prostheses configurations (Table 1)

Table 1. Study Strata and Expected Sample Sizes

	Fixed Bearing	Mobile Bearing	Total
Cruciate retaining	210	210	420
Posterior stabilizing	210	210	420
Total	**420**	**420**	**840**

- The study included five study visits:
  - Visit 1: Presurgery
  - Visit 2: Surgery
  - Visit 3: Less than 1 year following TKA surgery (postsurgery to 10 months)
  - Visit 4: Minimum 1 year following TKA surgery (10-22 months)
  - Visit 5: Minimum 2 years following TKA surgery (22 months or greater)

Sample

- An interim subset of the full study sample: n = 761
- Collected data necessary for the psychometric evaluation at three of the clinic visits: presurgery (visit 1), postsurgery to 10 months (visit 3), and 10-22 months (visit 4).

Measures

- The following measures were administered to patients at all visits except surgery (visit 2).
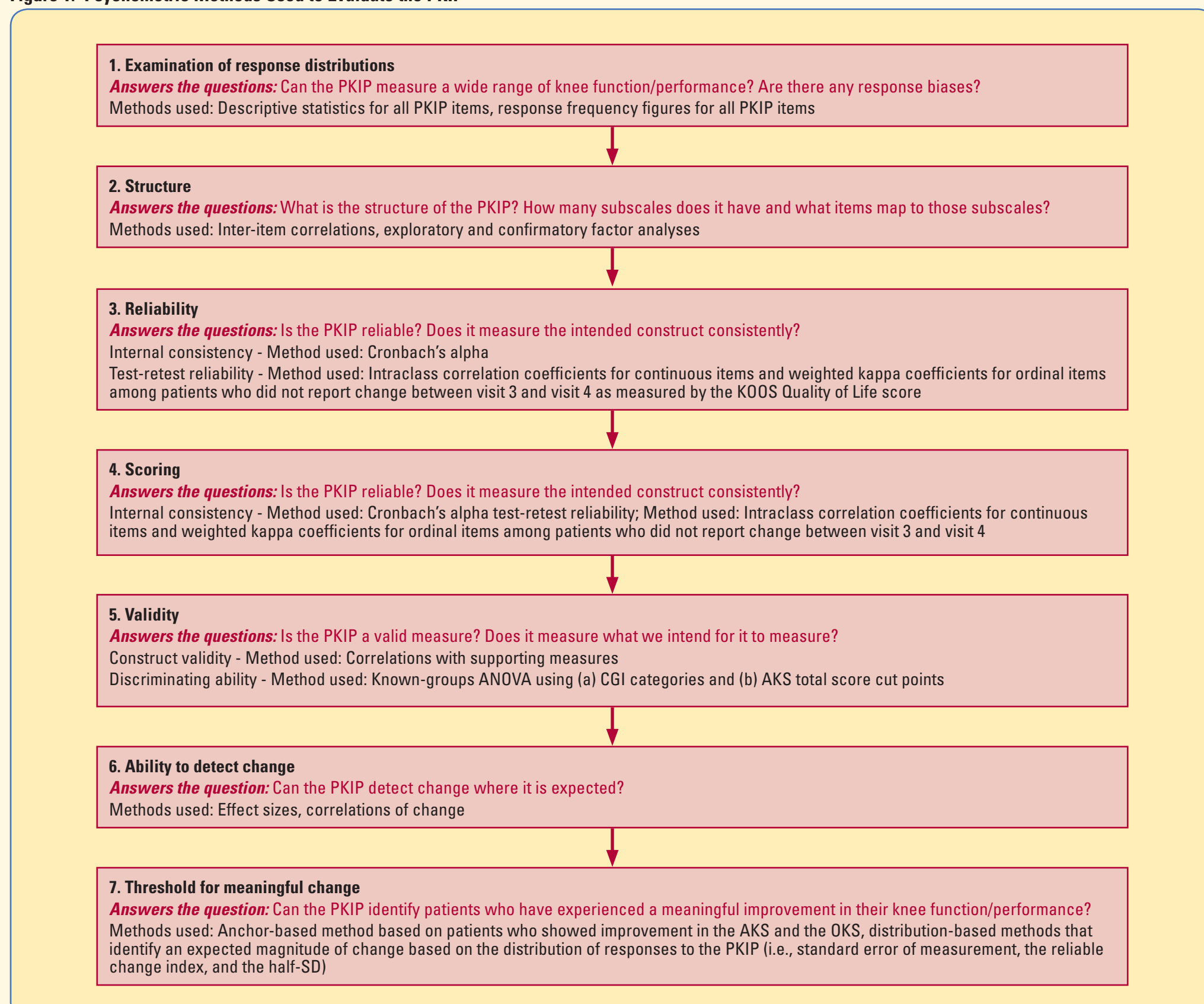
PRO Measures

- EQ-5D-3L[2,3]
- Knee Injury and Osteoarthritis Outcome Score (KOOS) Subscores[4]
- Oxford Knee Score (OKS) Total Score[5]
- Knee Noise and Front of Knee Pain (KNFKP) Items 1-3

Clinical Measures

- American Knee Society scores (AKS)[6]
- Clinical Global Impression score (CGI)[7]

Psychometric Methods (Figure 1)

Figure 1. Psychometric Methods Used to Evaluate the PKIP

1. Examination of response distributions
Answers the questions: Can the PKIP measure a wide range of knee function/performance? Are there any response biases?
Methods used: Descriptive statistics for all PKIP items, response frequency figures for all PKIP items

2. Structure
Answers the questions: What is the structure of the PKIP? How many subscales does it have and what items map to those subscales?
Methods used: Inter-item correlations, exploratory and confirmatory factor analyses

3. Reliability
Answers the questions: Is the PKIP reliable? Does it measure the intended construct consistently?
Internal consistency - Method used: Cronbach's alpha
Test-retest reliability - Method used: Intraclass correlation coefficients for continuous items and weighted kappa coefficients for ordinal items among patients who did not report change between visit 3 and visit 4 as measured by the KOOS Quality of Life score

4. Scoring
Answers the questions: Is the PKIP reliable? Does it measure the intended construct consistently?
Internal consistency - Method used: Cronbach's alpha test-retest reliability; Method used: Intraclass correlation coefficients for continuous items and weighted kappa coefficients for ordinal items among patients who did not report change between visit 3 and visit 4

5. Validity
Answers the questions: Is the PKIP a valid measure? Does it measure what we intend for it to measure?
Construct validity - Method used: Correlations with supporting measures
Discriminating ability - Method used: Known-groups ANOVA using (a) CGI categories and (b) AKS total score cut points

6. Ability to detect change
Answers the question: Can the PKIP detect change where it is expected?
Methods used: Effect sizes, correlations of change

7. Threshold for meaningful change
Answers the question: Can the PKIP identify patients who have experienced a meaningful improvement in their knee function/performance?
Methods used: Anchor-based method based on patients who showed improvement in the AKS and the OKS, distribution-based methods that identify an expected magnitude of change based on the distribution of responses to the PKIP (i.e., standard error of measurement, the reliable change index, and the half-SD)

RESULTS

Participant Characteristics

- A total of 761 patients completed the presurgery visit, 698 (91.7%) the less than 1 year (postsurgery to 10 months) visit, and 463 (60.8%) the less than 1 year (10-22 months) visit at the time of data transfer for this analysis.
- Participants' average age was 65.7 years (standard deviation [SD] = 8.1; range = 33-80 years).
- Female participants outnumbered males (58% vs. 42%).
- The majority of participants were white (72%).
- The average body mass index was 32.1 (SD = 6.6).
- Sample characteristics were similar across knee configurations.

Response Distributions

- Descriptive statistics did not reveal floor or ceiling effects or other response biases. Results were similar across knee configurations.

Structure

- Three main concepts emerged from the inter-item correlations and exploratory factor analysis based on medium to strong inter-item correlations and factor loadings ≥ 0.37: confidence/stability, modify activities, and satisfaction.
- Although the correlational analyses and factor analyses suggested that the confidence and stability items were highly related, qualitative interviews revealed that confidence and stability were distinct concepts. Therefore, the confidence and stability items were separated into two subscales.
- Confirmatory factor analysis supported the structure depicted in Figure 2 (chi-square = 15.94; df = 5; P = 0.007; comparative fit index = 0.98; Tucker-Lewis Index = 0.96; root mean square error of approximation = 0.069).

Figure 2. Structure of the PKIP
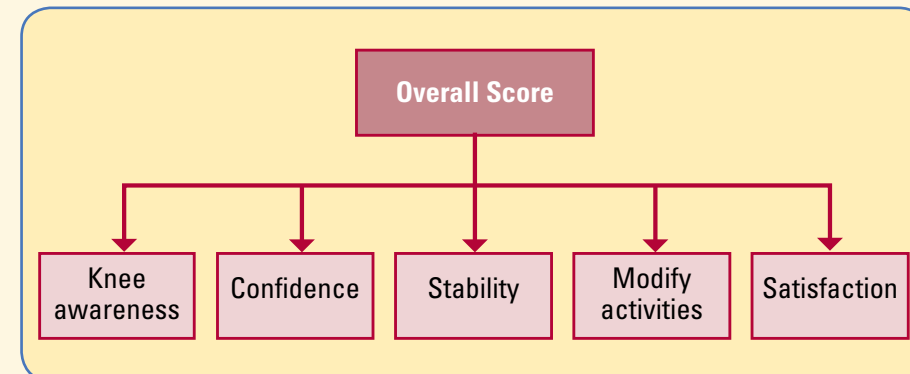
Reliability

- Cronbach's alphas were generally satisfactory across PKIP subscales and knee configuration, almost all attaining the minimum suggested threshold of 0.70.
- Although this study was not designed to assess test-retest reliability, an analysis was undertaken based on a small subgroup of patients (n = 60) who had the same KOOS Quality of Life score at both less than 1 year (postsurgery to 10 months) and minimum 1 year (postsurgery 10-22 months) and were presumed to be stable. The test-retest intraclass correlation coefficient of the PKIP Overall score was 0.77 using a time interval that averaged 256 days (range = 77 to 401 days).

Scoring

- A PKIP scoring algorithm was developed that involved reverse scoring two items, simple unit weighting, summing, and missing rules.
- PKIP subscale scores (Confidence, Stability, Modify Activities, Satisfaction) range from 0 to 10, and the PKIP Overall score ranges from 0 to 100; higher scores indicate better knee functioning.

Construct Validity

- Table 2 summarizes the correlation hypotheses and the actual correlational results for the PKIP Overall score at the minimum 1 year visit (10-22 months).
- Generally, the relationships between the PKIP Overall score and the supporting measures matched their hypothesized strength or were stronger, except for the relationships between the PKIP Overall score and the AKS scores, which were generally weak.

Table 2. Summary of PKIP Validity Correlation Hypotheses and Results for the PKIP Overall Score

Measures	Hypothesized Relationship	Minimum 1 year (Postsurgery 10-22 Months)	Change: Postsurgery 10-22 Months– Presurgery
KOOS Symptoms	Strong	Strong	Strong
KOOS Pain	Strong	Strong	Strong
KOOS Activities of Daily Living	Strong	Strong	Strong
KOOS Sports and Recreation	Moderate	Moderate	Moderate
KOOS Quality of Life	Strong	Strong	Strong
OKS Total Score	Strong	Strong	Strong
AKS Pain	Moderate	Moderate	Moderate
AKS Range	Moderate	Weak	Weak
AKS Stability	Moderate	Weak	Weak
AKS Alignment	Moderate	Weak	Weak
AKS Function	Moderate	Moderate	Moderate
AKS Total Score	Strong	Moderate	Moderate
KNFKP Item 1—hear grinding or clicking	Weak	Weak	Weak
KNFKP Item 2—pain with grinding or clicking	Weak	Moderate	Moderate
KNFKP Item 3—pain in front of knee	Weak	Strong	Moderate
EQ-5D-3L Index	Weak	Moderate	Weak
EQ-5D-3L VAS	Weak	Strong	Moderate
CGI	Weak	Moderate	Weak

VAS = visual analog scale.
Note: weak: $r < 0.30$; moderate: $r = 0.30$ to 0.50; strong: $r > 0.50$.

Discriminating Ability

- Analyses of variance (ANOVAs) were used to explore the discriminating ability of the PKIP by examining mean differences in scores across known subgroups by comparing the following:
  - Patients whose physicians rated them in the top 25% versus the bottom 25% on the CGI, and the top 50% and bottom 50%
  - Patients with AKS total scores ≥ 80 (better knee functioning) versus patients with AKS scores < 80 (worse knee functioning)
- It was hypothesized that patients who were rated in the top 25% or 50% of the CGI would have statistically better functioning than those whose physicians rated them in the bottom 25% or 50%. PKIP score means followed this pattern, and all ANOVA P values were statistically significant except for the Modify Activities score at both postsurgery visits.
- It was hypothesized that patients with better AKS scores would score better on the PKIP. Table 3 presents these PKIP score results.

Table 3. Known-Groups ANOVAS: AKS Scores at Minimum 1 Year (Postsurgery 10-22 Months

PKIP Scale	n	AKS ≥ 80 Mean (SD)	n	AKS < 80 Mean (SD)	F-Statistic (P Value)
PKIP Confidence	371	8.5 (1.6)	80	7.0 (2.1)	51.98 (0.000)
PKIP Stability	373	8.8 (1.6)	80	7.0 (2.3)	63.91 (0.000)
PKIP Modify Activities	373	6.5 (3.4)	81	5.2 (3.0)	10.37 (0.001)
PKIP Satisfaction	373	8.4 (1.7)	81	6.5 (2.6)	65.18 (0.000)
PKIP Overall	371	76.4 (16.7)	80	57.9 (18.5)	77.03 (0.000)

Ability to Detect Change

- Table 4 presents PKIP effect sizes from presurgery to postsurgery 10-22 months. Effect sizes were large in magnitude.

Table 4. PKIP Effect Size Estimates

PKIP Scale	Effect Size Estimates[a] (Presurgery to Postsurgery 10-22 Months)	Interpretation of Effect Size[b]
PKIP Confidence	2.12	Large effect
PKIP Stability	2.27	Large effect
PKIP Modify Activities	0.93	Large effect
PKIP Satisfaction	3.60	Large effect
PKIP Overall	2.96	Large effect

[a] Effect size estimates of change were computed using the mean change divided by the SD of the base-line score.
[b] Cohen (1988)[8] provides a general rule of thumb for the interpretation of effect size estimates: effect sizes of approximately 0.20 represent small effects, those of approximately 0.50 represent moderate effects, and those greater than approximately 0.80 represent large effects.

- Correlations also were calculated to provide evidence of the PKIP to detect change. Correlations between KOOS change scores and change in PKIP Overall scores were moderate to strong ($r = 0.44$ to 0.74). Correlations between AKS change scores and change in PKIP Overall scores were weak to moderate ($r = -0.02$ to 0.40).

Threshold for Meaningful Change

- Anchor-based method:
  - The AKS was considered as an anchor, but its appropriateness of use was not supported by the correlations between PKIP change and the AKS change scores, which were weak to moderate.
  - OKS as an anchor (5-point change) yielded a 23-point change in the PKIP Overall score from presurgery to the minimum 1 year visit (10-22 months).
- Distribution-based methods:
  - Standard error of measurement = 8.43
  - Reliable change index = 11.92
  - Half-SD = 7.5
- Preliminary working value for the responder threshold estimate: 7.5 to 23 points on the 100-point PKIP Overall score scale.

CONCLUSIONS

- This study provides important results regarding the behavior and psychometric properties of the PKIP in a population of patients before and after primary TKA.
- The PKIP is a reliable, valid, useful, and appropriate measure of patients' knee stability and functioning before and after TKA.
- Future studies should monitor the performance of the PKIP in new populations, in other languages, and in treatment comparisons.

REFERENCES

Please see handout for complete reference list.

CONTACT INFORMATION

Theresa Coles, MS
Senior Health Outcomes Analyst

RTI Health Solutions
Phone: +1.919.316.3843; E-mail: tcoles@rti.org