Utilization of Computer Adaptive Tests in Patients with Knee Conditions: A Systematic Review

Abbreviations:
PROM: Patient-reported Outcome Measure
CAT: Computer adaptive test
PROMIS: Patient Reported Outcomes Measurement Information System
FOTO Knee FS: Focus On Therapeutic Outcomes Knee Functional Status
OA: Osteoarthritis
PF: Physical Function
PI: Pain Interference
PB: Pain behavior
D: Depression
M: Mobility
HAQ-II: Health Assessment Questionnaire Disability Index
MLQOL: Multi-ligament Quality of Life;
WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index
KOOS: Knee injury and Osteoarthritis Outcome Score
IKDC: International Knee Documentation Committee
MARS: Marx Activity Rating Scale
EQ-5D-5L: European Quality of life-5 Dimension-5 level
SF-36: Short Form 36
GROC: Global Rating Of Change
NPS: Numeric Pain Scale
FTSST: Five Time Sit to Stand Test
TUG: Timed Up and Go
IPAQ: International Physical Activity Questionnaire
SANE: Single Assessment Numeric Evaluation
MLKI: Multi-Ligament Knee Injury
BEMD: Bilateral Extensor Mechanism Disruption
BTP: Bicondylar Tibial Plateau
CR: cruciate retaining
ICC: intraclass correlation coefficient
TKA: Total Knee Arthroplasty

Introduction

Patient-reported outcome measures (PROMs) are widely recognized in the field of orthopedics as important tools to evaluate the value of care and the effectiveness of treatment from the perspective of the patient.[1,2] As healthcare in the United States continues to shift towards value-based reimbursement, data from PROMs increasingly influence patterns of care delivery and clinical decision-making.[2] An abundance of validated PROMs for orthopedic knee conditions are historically used in clinical practice such as the International Knee Documentation Committee (IKDC) subjective knee form, the Knee Injury and Osteoarthritis Outcomes Score (KOOS), and the Lysholm Knee score, among others.[1,3,4] These fixed scale PROMs typically carry a trade-off between test length and score precision and they require responses to most, if not all, of the questions to produce a valid score.[5] Administering these tests burden patients and clinicians by increasing time required for completion, especially when multiple PROMs are indicated.[5,6] User fatigue and respondent burden may preclude the collection of accurate data and render the results less meaningful.

For a PROM to be implemented successfully, it should possess sound psychometric properties, minimize respondent fatigue, and improve the feasibility of scoring and interpretation.[1,3] Computerized adaptive testing (CAT) is a measurement methodology that meets these metrics. In contrast to fixed length PROMs, CATs measure patient-reported outcomes using algorithms based on item response theory (IRT). IRT tailors item delivery to the patient by selecting each successive item from an item bank based on the patient's response to the previous item, in order to generate an estimation of the construct being measured.[5] This allows for precise and accurate data collection with fewer questions.[5] CATs reduce administrative and respondent burden, as well as improve the efficiency of patient-reported data collection and interpretation.[3,5,7] With development of CATs such as the Patient Reported Outcomes Measurement Information System (PROMIS-CAT)[8] and Focus On Therapeutic Outcomes (FOTO)[9], there has been an increase in published literature evaluating the utility of these tests in research and practice. Recent studies have validated the use of CATs in spine surgery[10–12] and upper limb trauma.[13] They reported precision and efficiency of CATs, strong correlations with fixed scales, and adequate psychometric properties compared to fixed length PROMs.[10–14]

For people with orthopedic knee conditions, there are several fixed length legacy PROMs used in research and clinical practice.[1,3,4] However, it is unclear how and to what extent CATs are used in this population, and the utility of CATs compared to legacy PROMs. Therefore, the purpose of this study was to systematically review the use of CATs to measure patient-reported data in people with knee conditions. Specifically, our aim was to (1) assess how and for what purposes CATs have been used, (2) evaluate the psychometric properties of CATs, and (3) determine correlations between legacy PROMs and CATs for people with knee conditions.

Methods

Literature search

This systematic review was performed in compliance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines[15] and the review protocol was registered in PROSPERO. A literature review was performed to identify articles published prior to December 3, 2019. The following electronic databases were systematically searched: PubMed, Sport Discus, EMBASE, CINAHL, and Web of Science. The following search strategy was used to identify relevant articles: ((“computer* adapt* test*” OR “item response theory” OR “PROMIS” OR “patient reported outcomes measurement information system”) AND (“Knee” OR “ACL” OR “cruciate ligament” OR “MCL” OR “LCL” OR “PCL” OR “menis*” OR “patellofemoral” OR “tibiofemoral” OR “patella” OR “tibia” OR femur”)). A step-by-step PubMed search query can be found in Appendix 1. The lead author (T.H.) performed supplemental manual bibliographic reference searches to identify relevant studies missed during the initial database search.

Study Selection and Eligibility Criteria

The title and abstracts for each study were screened independently by the lead author (T.H.) and one other author (E.S., G.D.) in a stepwise manner (figure 1 (PRISMA flow diagram)). The full-text review was performed in the same manner with each study receiving two independent reviews. Disagreements were independently resolved by a third author (L.T.). Our inclusion criteria were as follows: (1) full-text publication (2) level I-IV studies as defined by the Oxford University Centre for Evidence-Based Medicine (OECBM)[16] including randomized controlled trials, cohort studies, case-control studies, cross sectional studies, case series, and clinical measurement studies (3) study sample of adults or children with a knee condition (4) reported patient-reported data collected via CAT. Articles were excluded if they were not published in English, did not include patients with knee conditions, had a study sample of patients with lower limb amputations, or only looked at non-computerized short form versions of a measure that can also be administered as a CAT. Review articles, editorials, conference abstracts, and protocols were excluded. Corresponding authors were contacted by the lead author if CAT utilization was not specified.

Data Extraction and Synthesis

The lead author (T.H.) independently extracted the following data into an electronic database (Microsoft Excel; Microsoft, Redmond, WA): (1) academic journal, study type, study design, sample size, patient demographics (age, sex, BMI, knee-related diagnosis), level of evidence, measurement domain, interventions where applicable, and length of follow-up) (2) psychometric properties including measures of validity, reliability, responsiveness, interpretability, and feasibility (3) correlations between legacy PROMs and CATs.

Studies were classified into 3 categories: development, testing, outcome assessment in the manner described by Jayakumar et al.[13] Studies were classified as combinations of categories where applicable. "Development" entailed studies that utilized methods of developing a CAT.[13] "Testing" was used to describe studies that evaluated psychometric properties of a CAT and "outcome assessment" was used to denote studies that used CATs to measure outcomes alone or in combination with fixed scales.[13] Psychometric properties of CATs were evaluated using previously described criteria and were summarized in table format as suggested by Terwee et al.[17]

Statistical Analysis

When possible, data were pooled for synthesis. Correlation of CATs to legacy PROMs and other outcome measures were interpreted using previously established guidelines—with $0.1 \le |r| < 0.3$, $0.3 \le |r| < 0.5$, and $|r| \ge 0.5$ signifying low, moderate, and strong correlations, respectively.[18]

Risk of Bias Assessment

The Cochrane Risk of Bias 2 tool was used to evaluate risk of bias in randomized controlled trials in this review.[19] This tool was designed to detect risk of bias due to the randomization process, bias due to deviation from intended interventions, bias due to missing data, bias in the measurement of outcomes, and bias in the selection of reported results.[19] Risk of bias for cohort, cross-sectional, and non-randomized studies was evaluated using the Risk of Bias Assessment tool for Non-randomized Studies (RoBANS).[20] The RoBANS tool evaluates 6 domains including selection of participants, confounding variables, measurement of exposure, blinding of outcomes, incomplete outcome data, and selective outcome reporting.[20] Risk of Bias for studies that reported psychometric properties was assessed using the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) Risk of Bias checklist.[21] The COSMIN checklist contains standards for study design and preferred statistical methods of studies that examine measurement properties such as validity, reliability, and responsiveness.[21]

Results

Study Selection

The initial search identified 346 studies that were screened for eligibility. Of these, 278 were excluded during title/abstract screening, leaving 67 studies for full text review. Twenty-seven studies were excluded during the full text review due to: use of fixed length short form CAT only (n = 17), failing to report data from CAT (n = 6), full text unavailable (n = 3), and less than 50% of the patient sample had knee conditions (n = 1). Forty studies met the criteria and were included in this review. (Fig. 1) Three unique CATs were identified from the search: PROMIS, FOTO knee FS, and Osteoarthritis-CAT (OA-CAT). PROMIS CATs, FOTO knee FS, and OA-CAT were used in 30 (75%), 7 (17.5%), and 4 (10%) studies, respectively.

Risk of Bias

Across studies, failure to adjust for confounding variables (33%) and selection bias (20%) were the most frequently implicated domains at risk of bias (Appendix 2, Table 2). Risk of bias for studies assessing psychometric properties of CATs varied. The estimates of reliability and measurement error were rated as doubtful for the OA-CAT subscales, FOTO Knee FS, and PROMIS PF-CAT. The reasons for doubtful ratings included inadequate time intervals between administrations[22–25], recall bias[23,24], multiple testing bias[23,24], and inadequate descriptions of intraclass correlation coefficients (ICC) [22–25] (Appendix 2, Table 3). Risk of bias for each study included in this review can be viewed in detail in Appendix 2.

Publication and Study Characteristics

Study and population characteristics for all included studies can be found in Tables 1 and 2, respectively. Studies that included CATs for patients with knee conditions were published in a variety of different journals with the most common journals being *The Journal of Knee Surgery* (n = 8, 20%), and *The Orthopaedic Journal of Sports Medicine* (n = 5, 12.5%). Cohort (n = 22, 55%) and cross-sectional (n = 9, 22.5%) studies were the most common study designs. Regarding study purpose, studies were classified as development (n = 3, 7.5%), testing (n = 13, 32.5%), outcome assessment (n = 14, 35%), and combination of testing and outcome assessment (n = 10, 25%) (Table 1). There were a variety of domains of health assessed with studies administering a range of 1 to 6 domains (Table 1). The most frequently administered CAT domains were PROMIS Physical Function-CAT (PF-CAT) (n = 30, 75%), PROMIS Pain Interference-CAT (PI-CAT) (n = 15, 37.5%), PROMIS Depression-CAT (D-CAT) (n = 9, 22.5%), and FOTO Knee Functional Status (FS) (n = 7, 17.5%). Additional study characteristics can be found in Appendix 3.

Sample sizes ranged from 30 to 28,320 and were 49% female on average. Mean age and BMI ranged from 24-67 years and 26.3-36.7 kg/m^2, respectively. Mean follow-up among studies that utilized CATs as outcome measures ranged from 6 weeks to 3 years. CATs were used in a wide variety of diagnoses and surgical procedures (Table 2). The most frequently observed populations included TKA (n = 8, 20%), knee impairments receiving rehabilitation services (n = 7, 17.5%), ACL injury and reconstruction (n = 5, 12.5%), and meniscal injury and knee surgery (n = 5, 12.5%).

Psychometric properties

The psychometric properties for each measure and patient population are summarized in Table 3 and can be viewed in greater detail in Appendix 4.

Validity

Content validity and internal consistency were assessed for each domain of the OA-CAT[23,24] and the FOTO Knee FS CAT[22] but not the PROMIS CAT measures. Construct validity was assessed in multiple versions of the PROMIS PF-CAT[7,25–33], the PROMIS PI-CAT[34,35], the OA-CAT-disability[23], and the FOTO Knee FS CAT.[9,22,36,37] Criterion validity of the PROMIS PF-CAT[27,29–31] and PROMIS PI-CAT[35] was established with the IKDC.

Reproducibility

Reliability was assessed in one version of the PROMIS PF-CAT (v1.0)[38] and all domains of the OA-CAT[23,24] and FOTO Knee FS[22] using various methods (Table 3 & Appendix 4).

Responsiveness

Responsiveness was assessed for multiple versions of PROMIS PF-CAT[7,29,30,38,39], PROMIS PI-CAT[34,39], PROMIS D-CAT[39], all domains of the OA-CAT[40], and the FOTO knee FS measure.[9,41] A variety of methods were used to calculate the responsiveness of measures including receiver operating curve area under the curve, effect size, and minimal detectable change. For this reason, pooling of results was not possible.

Interpretability

Minimal Clinical Important Differences (MCID) were calculated using both anchor and distribution-based methodologies. The MCID's for the PROMIS PF- and PI-CAT ranged from 1.97-22.5[30,38,39,42,43] and -3.2-4.43[30,39,43] T-score units, respectively. The MCID for the PROMIS D-CAT was found to be -4.9 T-score units in patients undergoing ACL reconstruction (ACLR).[39] The MCID for FOTO Knee FS ranged from 4-14 FS units.[9,41,44] Table 4 summarizes the MCIDs for CAT domains and patient populations.

Floor-ceiling effects

No studies reported floor or ceiling effects for ≥15% of the study sample. The majority of the studies that evaluated PROMIS CATs reported no floor or ceiling effects.[26,28,32,33,35,45] For the PROMIS PF-CAT, 0.2-0.5% floor effects and 0.4-9% ceiling effects were reported in 3 studies[7,27,31] (Appendix 3). The OA-CAT domains of disability, functional difficulty, and functional pain had ceiling effects of 4.02-6.1%[23,40], 0.6%[24], and 0.6-6%[23,40], respectively. The FOTO Knee FS CAT exhibited minimal floor and ceiling effects of 0.05 to 0.7% and 0.4 to 5%, respectively.[9]

Feasibility

The average number of questions answered on the PROMIS PF-CAT was 4.34 (range, 4-6.59) with an average completion time of 71 seconds (Table 5).[7,25–29,32,33,45–47] Other PROMIS domains (PI[45,46], M[45], and D[46]) averaged 5.1, 5.53, and 6.4 questions answered. Completion times were 43 and 46 seconds for PI and D, respectively.[46] Patients who completed the FOTO knee FS answered an average of 6.43 questions.[9,22] The OA-CAT FD took an average of 85 seconds to answer 5-10 questions[24,25,40] and the OA-CAT FP required an average 52 seconds to answer an average of 5 questions.[40]

Correlations of CATs with fixed length measures

A summary of the correlations of CAT domains to fixed length measures, the OA-CAT, and performance measures can be found in Table 5. Raw values from each included study can be found in Appendix 5.

PROMIS PF-CAT

The PROMIS PF-CAT demonstrated moderate-strong correlations to a majority of the fixed length disease-specific outcome measures including the IKDC[27,29–31], Lysholm[45], KOOS[7,26,32,33] and WOMAC[26] subscales, and KOOS, JR.[28] PROMIS PF-CAT also correlated strongly with the OA-CAT (r = 0.65-0.79).[25] Low-moderate correlations were reported with SF-36 GH (r = 0.12-0.43)[7,32,33] and MARS (r = 0.01-0.42).[7,26,32,33] Low correlations were found with the Press Ganey Medical Outcomes Survey (r = -0.18-0.22).[47]

PROMIS PI-CAT

The PROMIS PI-CAT demonstrated strong correlations with the IKDC[30,35], Tegner ARS[35], MLQOL[45], and NPS knee pain.[35] Moderate correlations were reported with the MARS, IPAQ, and NPS body pain.[35]

PROMIS M-CAT

The PROMIS M-CAT demonstrated strong correlations with the Lysholm and the MLQOL-AL.[45] Moderate correlations were reported with the Tegner ARS and MLQOL-PI.[45]

FOTO knee FS and OA-CAT disability

Low-strong correlations were reported with FOTO knee FS and global rating of change (GROC).[48] The OA-CAT disability correlated strongly the Health Assessment Questionnaire (HAQ-II).[23]

Discussion

The measurement of patient-reported outcomes has become an increasingly important component of establishing value in healthcare. This systematic review identified 3 unique CAT tools used to collect patient-reported data in patients with knee conditions. Both the OA-CAT and PROMIS have multiple subscales that are used to assess different domains of health while the FOTO Knee FS measure is designed only to capture functional status.[9,22,36,37,41,44,48] The OA-CAT measures 3 domains relevant to individuals with OA including disability[23], functional difficulty[24], and functional pain.[24] PROMIS CAT domains covered a broader range of health constructs including physical function[7,25,26,28–35,38,39,42,43,45–47,49–60], pain interference[30,34,35,39,43,45,46,49,51–54,57,58,60], depression[35,39,46,49,53,54,57,58,60], mobility[45], pain behavior[34], social satisfaction[35,57,60], and anxiety.[35,57,58,60]

CATs were administered and used for a variety of purposes. In general, the studies included in this review sought to establish psychometric properties, assess correlations with fixed length PROMs, characterize normal recovery patterns, and determine differences in outcomes between patient groups. The psychometric properties of validity, reliability, responsiveness, clinical feasibility and interpretability were found for all 3 CATs in reference to a variety of knee conditions. The majority of the included publications reported moderate to strong correlations of CATs to a variety of established fixed length PROMs.

Utilization of CATs

Studies varied greatly with regard to patient demographics, sample size, and length of follow-up. This finding is somewhat expected given the broad search strategy that was utilized to identify all pertinent literature. PROMIS measures were the most frequently utilized CATs in patients with knee conditions. No studies detailing the development of PROMIS CAT measures were identified in this review. However, it should be noted that rigorous development and calibration of PROMIS item banks has been previously detailed for the general U.S. population.[61–64] A large majority of the studies sought to establish the psychometric properties of PROMIS CAT measures for specific knee conditions.[7,25–33,35,39,45,47,56,60] Studies testing the psychometric properties of CATs for specific conditions are important catalysts for establishing validity of CATs for use in research and routine clinical practice.[13] Thus, it is reasonable to infer that promising results from psychometric studies have led to an increase in utilization of PROMIS CATs as primary outcome measures and replacing traditional fixed length measures. The PROMIS was developed out of an National Institutes of Health (NIH) initiative for improved patient-reported outcome measurement efficiency, precision, and reduced test burden.[8] Since 2004, the NIH has been developing large item banks to measure relevant symptoms and health constructs designed to be applicable in both research and clinical practice.[8] In 2014, Rose et al.[62] described the development of the physical function item bank and PF-CAT. Recent publication trends suggest that PROMIS CAT use is increasing. Prior to 2018, only 4 studies had been published on PROMIS CATs compared to 26 published between 2018 and 2019, suggesting that there is a move towards embracing these assessments as valid PROMs for patients with knee conditions. It is not surprising then that a trend was observed for PROMIS CAT assessments to be utilized to compare and analyze outcomes between two or more patient groups[34,42,51,52,54,55,57,58], characterize normal recovery patterns after surgery[43], and analyze postoperative variations[49,59] in patients undergoing knee surgery. It is noteworthy that PROMIS CAT usage has primarily been reported in surgical populations. Only one study by Chang et al.[25] utilized the PROMIS PF-CAT as an outcome measure in a sample of patients with knee OA not indicated for surgery.

The increasing trend for CATs to be used as outcome measures was not observed consistently with the OA-CAT scales or FOTO Knee FS. The OA-CAT scales were only reported in 4 publications. Two[23,24] of the 4 studies described the development of 3 different OA-CAT scales and were published in 2009. The other two studies[25,40] published between 2016 and 2018, tested the psychometric properties and used OA-CAT scales as outcome measures. The relatively small number of publications is somewhat unsurprising considering the OA-CAT is intended only for use in patients with lower extremity OA.[23,24] In contrast to the OA-CAT, the FOTO Knee FS measure can be administered to any patient with a knee impairment, broadening its application. The FOTO Knee FS literature was published from 2005-2013 and described the development of the scale and testing of its psychometric properties. To our knowledge, no studies utilizing FOTO Knee FS as an outcome measure have been published in the last 7 years. This is likely because FOTO Knee FS is primarily intended for use as an outcome measure in outpatient physical therapy practice. The current body of literature has validated the use of FOTO Knee FS measure in large populations of patients seeking care for knee impairments in outpatient settings.[9,22,36,37,41,44,48] This is in contrast to PROMIS measures, which were designed to be used in both research and clinical practice.[8,61]

The increased use of PROMIS in comparison to other CATs is also likely related to it being a generic measure of health, whereas the OA-CAT and FOTO Knee FS aim to assess disease- and region-specific areas, respectively. Distinct advantages of generic CAT measures such as PROMIS is that they have wide clinical applicability and allow for comparability across patient conditions and diseases.[61] Jevotovksy et al.[53] demonstrated this feature of PROMIS CATs by comparing preoperative and postoperative scores of an ACLR cohort with a cohort of patients undergoing lumbar discectomy to determine the relative value of the two procedures.

Comparisons of the value of interventions are likely to become more common as PROMIS instruments continue to be used more wisely. Furthermore, the results of these comparative studies can direct future research and clinical practice habits.

Psychometric properties

Validity

Evidence on the psychometric properties of CATs for patients with knee conditions is promising. As mentioned previously, PROMIS measure development occurred in the general population. As such, content validity and internal consistency of PROMIS scales were not addressed in the studies included in this review. These properties have been established for the general U.S. population for which PROMIS was designed to be used and have been described previously.[61,62,64] The results presented in Table 3 reflect the psychometric properties that were tested specifically in patients with knee conditions and not the full body of literature on PROMIS measures. Rose et al.[62] reported the internal consistency of the PROMIS Physical function item bank to be 0.95. The PROMIS PF- and PI-CAT scales appear to be valid instruments for multiple classifications of knee conditions (Table 3). The PROMIS PF- and PI-CAT demonstrated criterion validity with the IKDC in multiple knee conditions, suggesting that they could be used where the IKDC was previously indicated. Construct validity of the PROMIS PI- and PF-CAT was established with a variety of validated disease-specific and generic PROMs across different knee conditions. (Table 5, Appendix 5).

The OA-CAT scales are valid and reliable for patients with lower extremity knee or hip OA (Table 3). Jette and colleagues[23,24] described the development and content validity of the OA-CAT measures in 2009. The construct validity of the OA-CAT disability was established with the Health Assessment Questionnaire-II (HAQ-II).[23] The psychometric properties of the

OA-CAT FD and FP scales were compared to the WOMAC, which is widely used in OA research.[24] However, we were unable to give a positive rating for construct validity because no predefined hypotheses were formed.[17] Future studies utilizing the OA-CAT scales should form specific, predefined hypotheses to avoid bias in interpretation of the results.[17] McDonough et al.[40] conducted further psychometric testing on all 3 OA-CAT scales and found the internal consistency values to be between 0.86 and 0.88. The FOTO Knee FS measure is valid and reliable for patients with knee impairments receiving outpatient physical therapy services. Hart et. al.[22] detailed established the content validity and internal consistency ($\alpha = 0.96$) of the FOTO Knee FS measure in 2005. The FOTO Knee FS measure is based on the 18-item Lower Extremity Functional Scale (LEFS), a measure with wide clinical applicability and sound psychometric properties.[22] Criterion validity was not tested in the OA-CAT scales or the FOTO Knee FS measure (Table 3).

Reproducibility

Acceptable levels of agreement and reliability for PROMIS measures have been established for the general population.[61] As we only intended to capture the literature with regard to patients with knee conditions, this is one potential reason that these properties were rarely described for PROMIS CAT measures in studies included in this review (Table 3). Another potential explanation for the lack of PROMIS publications reporting measures of reproducibility in knee-specific populations could be due to the high number of observational and cross-sectional study designs and limited number of longitudinal studies. A recent review by Jayakumar et al.[13] observed a similar trend for psychometric reporting on CATs in patients with upper limb trauma corroborating the findings of this review that CATs have not yet been fully embraced in orthopedics. Agreement concerns the absolute measurement error and is important when trying to distinguish clinically meaningful change from measurement error.[17] High reliability is important for distinguishing between patients with various levels of a disease or construct.[17] Together, agreement and reliability concern the reproducibility of a measure.[17] In order to be fully integrated into routine clinical use, further testing and improved reporting of these properties in cohorts of patients with knee conditions is warranted.

Values for agreement and reliability were found for the OA-CAT scales and the FOTO Knee FS measure. However, the type of ICC used was often poorly described or unclear. As such, the potential for risk of bias should be considered when interpreting the results from Table 3. The $ICC_{agreement}$ is preferred because systematic differences are considered to a part of measurement error.[17]

Responsiveness and Floor/Ceiling Effects

All 3 CATs demonstrated responsiveness comparable to fixed length measures (Table 3, Appendix 4). Scott et al.[7] assessed the responsiveness of the PROMIS PF-CAT in patients who underwent ACLR up to 6 months after surgery. However, responsiveness in this study was presented as effect sizes. Indices of responsiveness such as effect size and standardized response mean are more representative of treatment effects rather than the properties of the measure in question.[17] Responsiveness can be considered a measure of longitudinal validity, as it concerns the ability of a measure to detect change over time.[17] The responsiveness of a measure as represented by area under (AUC) the receiver operating curve (ROC) can also be used in predictive modeling.[29,39] The PROMIS PF-CAT demonstrated responsiveness in patients with knee impairments, patients undergoing ACLR, and patients undergoing knee surgery.[7,29,30,38,39] Papuga et al.[29] reported preliminary evidence that preoperative and early postoperative PROMIS PF-CAT scores could predict poor outcomes in a cohort of patients who underwent ACLR. Chen et al.[39] determined optimal cutoffs for preoperative PROMIS PF-, PI, and D-CAT scores that could predict the likelihood of achieving an MCID postoperatively. Applications of responsiveness, such as the previous two examples, should be replicated in other patient populations to aid clinicians in stratifying patients and adjusting interventions.

There were very minimal floor and ceiling effects observed across multiple knee conditions (Table 3). We defined floor and ceiling effects as "favorable" when <15% of the respondents scored at the upper or lower limits of the measure.[14,17] All 3 CATs demonstrated favorable floor and ceiling effects in all populations of patients with knee conditions observed in this review, consistent with the findings of a recent review on the floor and ceiling effects of CATs in various orthopedic populations.[14] The highest ceiling effects observed (9%) for the PROMIS PF-CAT v1.2 occurred in a young and active population of patients undergoing ACLR at the 2-year follow-up.[7] This is potentially concerning when considering administration of the PROMIS PF-CAT to younger, active patients. However, it should be noted that the developers have since refined the item bank (PROMIS PF-CAT v2.0) to improve item coverage for higher level patients.[62] Three studies[26,46,59] utilized PROMIS PF-CAT v2.0 and only one study[26] reported on floor and ceiling effects. Carender et al.[26] found no floor or ceiling effects for patients with patellofemoral malalignment and chondral disease undergoing Fulkerton osteotomies and concomitant cartilage procedures. However, these results should be interpreted with caution as the preinjury level of physical function was not reported. Future studies should utilize the PROMIS PF-CAT v2.0 in populations that report higher physical function (i.e. ACLR) to determine if the item coverage is sufficient.

<u>Interpretability</u>

PROMIS CAT measures and the FOTO Knee FS measure also demonstrated properties of clinical interpretability (Table 3). The MCID values for these CATs have been determined for multiple clinical populations including patients with knee pathology and knee impairments (Table 4). Surgical populations, including patients undergoing ACLR, TKA, and knee arthroscopy have also had MCID values calculated (Table 4). The MCID has not yet been reported for the OA-CAT scales.

The MCID is critical to applying meaningful interpretation of scores and guide clinical decision-making.[38] Both anchor-based and distribution-based methods were utilized to derived MCIDs. Anchor-based methods define change using an external criterion while distribution-based methods use statistical characteristics of a sample.[17,38] Hung et al.[60] utilized both methods to determine the MCID for the PROMIS PF-CAT and found a wide range of MCID values (Table 6) and highlighted that there is no correct or incorrect way to calculate MCID. While anchor-based methods are generally preferred, MCID values derived from both methods can provide a reference for clinicians to make individual judgements about interpreting meaningful change.[60] Studies on the FOTO Knee FS measure[9,41,44] also used a combination of methods for deriving the MCID, providing clinical guidance to clinicians that work with patients who have knee impairments.

Clinical Feasibility

Consistent with findings from other reviews,[12–14] this review found that CATs reduce respondent burden by significantly reducing the number of questions asked and the total time required for completion when compared to fixed length PROMs (Table 5). Robins et al.[27] reported average administration times of 55 seconds and 268 seconds for the PROMIS PF-CAT and IKDC, respectively. The ability of CATs to be clinically feasible without compromising measurement precision lend support toward adopting the measures in routine clinical practice. In busy practices where patients may be interrupted while completing an assessment, CATs have the ability to generate a score estimate even if a stopping rule is not met. This unique feature of CATs is likely to mitigate the shortcomings associated with incomplete respondent data from traditional fixed length PROMs. There are some potential limitations for CATs to be feasibly incorporated into clinical practice. Costs associated with startup and maintenance may preclude widespread adoption of CATs. Lizzio et al.[46] determined the initial and maintenance costs of a PROMIS-based registry in an ambulatory sports medicine clinic to be $2,045 and $1000 monthly, respectively. However, without an existing outcome collection database such as REDCap, initial costs can approach $9000.[46] The initial startup costs could be a barrier to implementation for smaller institutions and private clinics. Clinical feasibility may also be affected by an individual's familiarity with using electronic devices. Lizzio et al.[46] recorded compliance rates greater than 80% for patients over the age of 62. However, this was significantly lower than the >95% compliance rates for patients under the age of 62.[46] Hence, some patient populations may be less familiar with using electronic devices which should be taken into consideration when these instruments are delivered in the clinic. In spite of any unfamiliarity with electronic device use, the efficiency of CAT tools is likely to offset the learning curve for electronic administration.

Correlations of CATs to other outcome measures

The PROMIS PF-CAT was strongly correlated with other fixed length assessments that measure physical function such as the IKDC, SF-36 PF, KOOS Sport, KOOS ADL, and WOMAC function, among others. Further, the PROMIS PF-CAT demonstrated low to moderate correlations with measures that do not directly assess physical function such as the SF-36 GH, MARS, and the Press Ganey Outpatient Medical Practice Survey. These correlations maintain the construct validity of the PROMIS PF-CAT. However, strong correlations were observed with the EQ-5D-5L, an instrument designed to measure general health and not physical function. One explanation is that the patients in the 4 studies[7,26,32,33] that administered the EQ-5D-5L were otherwise healthy and more likely to report better general health and physical function than other clinical populations.[32]

The PROMIS PI-CAT was studied less frequently but demonstrated similarly strong correlations to the IKDC in 2 studies. One study found strong (r = 0.58) correlations between the PROMIS PI-CAT and the NPS for body pain and knee pain, respectively. In their analysis, Nadarajah et al. found that pain intensity and PROMIS PI-CAT shared only approximately 25% of the variance. Thus, while these measures were strongly correlated, they may assess different domains of pain. PROMIS PI-CAT may be a more clinically useful assessment as it goes beyond a simple measurement of pain intensity to determine how pain impacts a person's life.

Chang et al.[25] found a strong correlation between the PROMIS PF-CAT and OA-CAT FD scale before and after a 6-week exercise intervention. The OA-CAT was slightly more responsive, but the difference in the effect size was not statistically significant, suggesting that either measure can be used to assess physical function in patients with OA.[25] Chang and colleagues[25] also examined the correlation of the two CATs to the FTSST. Interestingly, only low to moderate negative correlations were observed suggesting that the construct of physical function is not necessarily analogous to improved lower extremity strength. Similarly, Givens et al.[56] observed only moderate correlations between PROMIS PF-CAT and TUG in patients with OA indicated for TKA. This data suggests that the PROMIS PF-CAT should be considered an adjunct and not a replacement to performance-based outcome measures.

A potential explanation is that during development, each CAT tool was compared to well-established fixed length assessments with wide clinical applicability.[22–24] Thus, both the FOTO Knee FS and OA-CAT scales are appropriate for clinical use. However, there are gaps in the current body of literature that could be filled to improve the applicability of these measures. As mentioned previously, the correlation of the OA-CAT FD and FP scales with the WOMAC should be assessed in future studies along with predefined hypotheses about the degree of correlation. Additionally, only 2 studies[25,40] have been published with knee OA-specific cohorts. The OA-CAT scales were intended for use with all patients with lower extremity OA. However, it is important to determine psychometric properties for specific populations as there measurement properties differ between populations.[17] The FOTO Knee FS measure, to our knowledge, has not been compared to any other measures of function with regard to assessing correlations with other measures. It would be useful to determine the degree of correlation between the PROMIS PF-CAT and FOTO Knee FS in patients with knee conditions to determine if they measure the same construct.

Limitations

This review is not without limitations. First, only studies included before December 3, 2019 were included in this review. We acknowledge publication bias, as only full text articles published in English were included. Additionally, we only sought to capture literature on CAT tools used in populations of patients with knee conditions. Hence, literature on the psychometric properties of CAT tools that were developed for the general population could have been missed. However, we sought to mitigate this limitation by referencing previously established properties where applicable. Publication trends since 2017 suggest that the volume of PROMIS CAT literature is increasing year after year. Thus, relevant studies published after database retrieval could have been missed. However, we consider the increase in PROMIS CAT publications to justify the timing of this review. Selective reporting bias is a concern as well; not all studies included in this review reported correlations to fixed length measures or appropriate methods for deriving psychometric properties. We acknowledge potential selection bias as many of the studies were performed at single institutions and excluded non-English speakers. Additionally, younger, high functioning patients could have been underrepresented as only 4 studies[7,29,32,39] had a patient sample with an average age less than 30-years-old. Thus, our finding that CATs have minimal ceiling effects should be interpreted cautiously. A search strategy with broad inclusion and exclusion criteria yielded studies with variability in study design, patient populations, and data reporting. As such, pooling of results was not possible and analyses of subgroups were not performed. However, a sensitive strategy was chosen to ensure that all relevant studies were included to provide a complete review of CAT utilization in the knee literature.

Future Research

Future outcomes research should utilize evidence-based criteria, such as the criteria proposed by Terwee et al.[17] to design and evaluate studies on psychometric properties of PROMs. Future studies evaluating PROMIS CAT measures should examine the reproducibility and longitudinal validity of PROMIS CATs in both surgical and nonsurgical knee-specific populations to further validate PROMIS CAT measures as surrogates to fixed length PROMs. Additionally, PROMIS CAT measures should be used in younger, high-functioning populations to determine measurement properties such as ceiling effects. The OA-CAT scales have potential for use in OA research, however, further establishment of responsiveness and MCID is warranted before being implemented into clinical practice. Additionally, future research should address the longitudinal and construct validity of the OA-CAT in cohorts of patients with knee OA. Knee OA is a lifelong condition. Hence, a measure that possesses longitudinal validity and responsiveness is important to capture change over long periods of time. The FOTO Knee FS measure should be utilized in prospective research designs to assess longitudinal validity. Furthermore, the PROMIS PF-CAT should be compared to the FOTO Knee FS to determine if they measure similar constructs of physical function.

Conclusions

This systematic review aimed to identify how CATs were used in patients with knee conditions as well as examine the psychometric properties and correlations to fixed length measures. There is a growing trend for PROMIS CAT measures to be used in outcomes research, primarily for patients with knee conditions indicated for surgery. PROMIS CATs, OA-CAT scales, and the FOTO Knee FS all demonstrated sound psychometric properties and were more efficient than traditional fixed length PROMs without sacrificing measurement precision. Furthermore, studies on PROMIS CAT measures and the FOTO Knee FS measure have derived clinically relevant values, such as the MCID to assist clinicians and patients in shared decision-making about treatment goals. As such, the PROMIS CAT measures and FOTO Knee FS measure are recommended for clinical use in patients with knee conditions. All 3 measures are appropriate for use in health outcomes research in populations of patients with knee conditions.

Bibliography

1. Beaton DE, Schemitsch E. Measures of health-related quality of life and physical function. *Clin. Orthop. Relat. Res.* 2003;(413):90-105. doi:10.1097/01.blo.0000079772.06654.c8.
2. Ayers DC. Implementation of Patient-reported Outcome Measures in Total Knee Arthroplasty. *J Am Acad Orthop Surg* 2017;25 Suppl 1:S48-S50. doi:10.5435/JAAOS-D-16-00631.
3. Fidai MS, Saltzman BM, Meta F, et al. Patient-Reported Outcomes Measurement Information System and Legacy Patient-Reported Outcome Measures in the Field of Orthopaedics: A Systematic Review. *Arthroscopy* 2018;34(2):605-614. doi:10.1016/j.arthro.2017.07.030.
4. Wang D, Jones MH, Khair MM, Miniaci A. Patient-reported outcome measures for the knee. *J. Knee Surg.* 2010;23(3):137-151. doi:10.1055/s-0030-1268691.
5. Brodke DJ, Saltzman CL, Brodke DS. PROMIS for orthopaedic outcomes measurement. *J Am Acad Orthop Surg* 2016;24(11):744-749. doi:10.5435/JAAOS-D-15-00404.
6. Bodart S, Byrom B, Crescioni M, Eremenco S, Flood E. Perceived Burden of Completion of Patient-Reported Outcome Measures in Clinical Trials:: Results of a Preliminary Study. *Ther. Innov. Regul. Sci.* 2019;53(3):318-323. doi:10.1177/2168479018788053.
7. Scott EJ, Westermann R, Glass NA, Hettrich C, Wolf BR, Bollier MJ. Performance of the PROMIS in patients after anterior cruciate ligament reconstruction. *Orthop. J. Sports Med.* 2018;6(5):2325967118774509. doi:10.1177/2325967118774509.
8. Cella D, Yount S, Rothrock N, et al. The patient-reported outcomes measurement information system (PROMIS): Progress of an NIH roadmap cooperative group during its first two years. *Med. Care* 2007;45(5 Suppl 1):S3-S11. doi:10.1097/01.mlr.0000258615.42478.55.
9. Hart DL, Wang Y-C, Stratford PW, Mioduski JE. Computerized adaptive test for patients with knee impairments produced valid and responsive measures of function. *J. Clin. Epidemiol.* 2008;61(11):1113-1124. doi:10.1016/j.jclinepi.2008.01.005.
10. Bhatt S, Boody BS, Savage JW, Hsu WK, Rothrock NE, Patel AA. Validation of Patient-reported Outcomes Measurement Information System Computer Adaptive Tests in Lumbar Disk Herniation Surgery. *J Am Acad Orthop Surg* 2019;27(3):95-103. doi:10.5435/JAAOS-D-17-00300.
11. Boody BS, Bhatt S, Mazmudar AS, Hsu WK, Rothrock NE, Patel AA. Validation of Patient-Reported Outcomes Measurement Information System (PROMIS) computerized adaptive tests in cervical spine surgery. *J Neurosurg Spine* 2018;28(3):268-279. doi:10.3171/2017.7.SPINE17661.
12. Haws BE, Khechen B, Bawa MS, et al. The Patient-Reported Outcomes Measurement Information System in spine surgery: a systematic review. *J Neurosurg Spine*

2019;30(3):405-413. doi:10.3171/2018.8.SPINE18608.
13. Jayakumar P, Overbeek C, Vranceanu AM, et al. The use of computer adaptive tests in outcome assessments following upper limb trauma: a systematic review. *Bone Joint J.* 2018;100-B(6):693-702. doi:10.1302/0301-620X.100B6.BJJ-2017-1349.R1.
14. Gulledge CM, Lizzio VA, Smith DG, Guo E, Makhni EC. What Are the Floor and Ceiling Effects of Patient-Reported Outcomes Measurement Information System Computer Adaptive Test Domains in Orthopaedic Patients? A Systematic Review. *Arthroscopy* 2020. doi:10.1016/j.arthro.2019.09.022.
15. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 2009;6(7):e1000097. doi:10.1371/journal.pmed.1000097.
16. OCEBM Levels of Evidence - CEBM. Available at: https://www.cebm.net/2016/05/ocebm-levels-of-evidence/. Accessed April 13, 2020.
17. Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J. Clin. Epidemiol.* 2007;60(1):34-42. doi:10.1016/j.jclinepi.2006.03.012.
18. Cohen J 1923-1998. *Statistical Power Analysis For The Behavioral Sciences*. 2nd ed. Hillsdale, N.J: L. Erlbaum Associates; 1988:400.
19. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019;366:l4898. doi:10.1136/bmj.l4898.
20. Kim SY, Park JE, Lee YJ, et al. Testing a tool for assessing the risk of bias for nonrandomized studies showed moderate reliability and promising validity. *J. Clin. Epidemiol.* 2013;66(4):408-414. doi:10.1016/j.jclinepi.2012.09.016.
21. Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Qual. Life Res.* 2018;27(5):1171-1179. doi:10.1007/s11136-017-1765-4.
22. Hart DL, Mioduski JE, Stratford PW. Simulated computerized adaptive tests for measuring functional status were efficient with good discriminant validity in patients with hip, knee, or foot/ankle impairments. *J. Clin. Epidemiol.* 2005;58(6):629-638. doi:10.1016/j.jclinepi.2004.12.004.
23. Jette AM, McDonough CM, Haley SM, et al. A computer-adaptive disability instrument for lower extremity osteoarthritis research demonstrated promising breadth, precision, and reliability. *J. Clin. Epidemiol.* 2009;62(8):807-815. doi:10.1016/j.jclinepi.2008.10.004.
24. Jette AM, McDonough CM, Ni P, et al. A functional difficulty and functional pain instrument for hip and knee osteoarthritis. *Arthritis Res. Ther.* 2009;11(4):R107. doi:10.1186/ar2760.
25. Chang F-H, Jette AM, Slavin MD, Baker K, Ni P, Keysor JJ. Detecting functional change in response to exercise in knee osteoarthritis: a comparison of two computerized adaptive tests. *BMC Musculoskelet. Disord.* 2018;19(1):29. doi:10.1186/s12891-018-1942-9.
26. Carender CN, Bollier MJ, Wolf BR, Duchman KR, An Q, Westermann RW. Preoperative performance of PROMIS in patients with patellofemoral malalignment and chondral disease. *Orthop. J. Sports Med.* 2019;7(7):2325967119855001. doi:10.1177/2325967119855001.
27. Robins RJ, Anderson MB, Zhang Y, Presson AP, Burks RT, Greis PE. Convergent Validity of the Patient-Reported Outcomes Measurement Information System's Physical Function Computerized Adaptive Test for the Knee and Shoulder Injury Sports Medicine

Patient Population. *Arthroscopy* 2017;33(3):608-616. doi:10.1016/j.arthro.2016.08.032.
28. Kortlever JTP, Leyton-Mange A, Keulen MHF, et al. PROMIS Physical Function Correlates with KOOS, JR in Patients with Knee Pain. *J. Knee Surg.* 2019. doi:10.1055/s-0039-1688780.
29. Papuga MO, Beck CA, Kates SL, Schwarz EM, Maloney MD. Validation of GAITRite and PROMIS as high-throughput physical function outcome measures following ACL reconstruction. *J. Orthop. Res.* 2014;32(6):793-801. doi:10.1002/jor.22591.
30. Kenney RJ, Houck J, Giordano BD, Baumhauer JF, Herbert M, Maloney MD. Do Patient Reported Outcome Measurement Information System (PROMIS) Scales Demonstrate Responsiveness as Well as Disease-Specific Scales in Patients Undergoing Knee Arthroscopy? *Am. J. Sports Med.* 2019;47(6):1396-1403. doi:10.1177/0363546519832546.
31. Miles M, Nadarajah V, Jauregui JJ, et al. Evaluation of the PROMIS physical function computer adaptive test in patients undergoing knee surgery. *J. Knee Surg.* 2019. doi:10.1055/s-0039-1688691.
32. Hancock KJ, Glass N, Anthony CA, et al. PROMIS: a valid and efficient outcomes instrument for patients with ACL tears. *Knee Surg. Sports Traumatol. Arthrosc.* 2019;27(1):100-104. doi:10.1007/s00167-018-5034-z.
33. Hancock KJ, Glass N, Anthony CA, et al. Performance of PROMIS for healthy patients undergoing meniscal surgery. *J. Bone Joint Surg. Am.* 2017;99(11):954-958. doi:10.2106/JBJS.16.00848.
34. Edelstein AI, Bhatt S, Wright-Chisem J, Sullivan R, Beal M, Manning DW. The Effect of Implant Design on Sagittal Plane Stability: A Randomized Trial of Medial- versus Posterior-Stabilized Total Knee Arthroplasty. *J. Knee Surg.* 2019. doi:10.1055/s-0039-1678524.
35. Nadarajah V, Glazier E, Miller K, et al. Evaluation of preoperative pain using PROMIS pain interference in knee surgery patients. *J. Knee Surg.* 2019. doi:10.1055/s-0039-1688769.
36. Hart DL, Deutscher D, Crane PK, Wang Y-C. Differential item functioning was negligible in an adaptive test of functional status for patients with knee impairments who spoke English or Hebrew. *Qual. Life Res.* 2009;18(8):1067-1083. doi:10.1007/s11136-009-9517-8.
37. Deutscher D, Hart DL, Crane PK, Dickstein R. Cross-cultural differences in knee functional status outcomes in a polyglot society represented true disparities not biased by differential item functioning. *Phys. Ther.* 2010;90(12):1730-1742. doi:10.2522/ptj.20100107.
38. Hung M, Bounsanga J, Voss MW, Saltzman CL. Establishing minimum clinically important difference values for the Patient-Reported Outcomes Measurement Information System Physical Function, hip disability and osteoarthritis outcome score for joint reconstruction, and knee injury and osteoarthritis outcome score for joint reconstruction in orthopaedics. *World J Orthop* 2018;9(3):41-49. doi:10.5312/wjo.v9.i3.41.
39. Chen RE, Papuga MO, Voloshin I, et al. Preoperative PROMIS scores predict postoperative outcomes after primary ACL reconstruction. *Orthop. J. Sports Med.* 2018;6(5):2325967118771286. doi:10.1177/2325967118771286.
40. McDonough CM, Stoiber E, Tomek IM, et al. Sensitivity to change of a computer adaptive testing instrument for outcome measurement after hip and knee arthroplasty and periacetabular osteotomy. *J. Orthop. Sports Phys. Ther.* 2016;46(9):756-767.

doi:10.2519/jospt.2016.6442.
41. Wang Y-C, Hart DL, Stratford PW, Mioduski JE. Baseline dependency of minimal clinically important improvement. *Phys. Ther.* 2011;91(5):675-688. doi:10.2522/ptj.20100229.
42. Kohring JM, Erickson JA, Anderson MB, Gililland JM, Peters CL, Pelt CE. Treated versus untreated depression in total joint arthroplasty impacts outcomes. *J. Arthroplasty* 2018;33(7S):S81-S85. doi:10.1016/j.arth.2018.01.065.
43. Kagan R, Anderson MB, Christensen JC, Peters CL, Gililland JM, Pelt CE. The Recovery Curve for the Patient-Reported Outcomes Measurement Information System Patient-Reported Physical Function and Pain Interference Computerized Adaptive Tests After Primary Total Knee Arthroplasty. *J. Arthroplasty* 2018;33(8):2471-2474. doi:10.1016/j.arth.2018.03.020.
44. Wang Y-C, Hart DL, Stratford PW, Mioduski JE. Clinical interpretation of computerized adaptive test-generated outcome measures in patients with knee impairments. *Arch. Phys. Med. Rehabil.* 2009;90(8):1340-1348. doi:10.1016/j.apmr.2009.02.008.
45. Trasolini NA, Korber S, Gipsman A, San AE, Weber AE, Hatch GFR. Performance of PROMIS Computer Adaptive Testing As Compared With Established Instruments for Multiple-Ligament Knee Injuries. *Orthop. J. Sports Med.* 2019;7(9):2325967119867419. doi:10.1177/2325967119867419.
46. Lizzio VA, Blanchett J, Borowsky P, et al. Feasibility of PROMIS CAT Administration in the Ambulatory Sports Medicine Clinic With Respect to Cost and Patient Compliance: A Single-Surgeon Experience. *Orthop. J. Sports Med.* 2019;7(1):2325967118821875. doi:10.1177/2325967118821875.
47. Kohring JM, Pelt CE, Anderson MB, Peters CL, Gililland JM. Press Ganey Outpatient Medical Practice Survey Scores Do Not Correlate With Patient-Reported Outcomes After Primary Joint Arthroplasty. *J. Arthroplasty* 2018;33(8):2417-2422. doi:10.1016/j.arth.2018.03.044.
48. Schmitt JS, Abbott JH. Patient global ratings of change did not adequately reflect change over time: a clinical cohort study. *Phys. Ther.* 2014;94(4):534-542. doi:10.2522/ptj.20130162.
49. Bernholt D, Wright RW, Matava MJ, Brophy RH, Bogunovic L, Smith MV. Patient reported outcomes measurement information system scores are responsive to early changes in patient outcomes following arthroscopic partial meniscectomy. *Arthroscopy* 2018;34(4):1113-1117. doi:10.1016/j.arthro.2017.10.047.
50. Robins RJ, Daruwalla JH, Gamradt SC, et al. Return to play after shoulder instability surgery in national collegiate athletic association division I intercollegiate football athletes. *Am. J. Sports Med.* 2017;45(10):2329-2335. doi:10.1177/0363546517705635.
51. Cavallero M, Rosales R, Caballero J, Virkus WW, Kempton LB, Gaski GE. Locking plate fixation in a series of bicondylar tibial plateau fractures raises treatment costs without clinical benefit. *J Orthop Trauma* 2018;32(7):333-337. doi:10.1097/BOT.0000000000001188.
52. Virkus WW, Caballero J, Kempton LB, Cavallero M, Rosales R, Gaski GE. Costs and Complications of Single-Stage Fixation Versus 2-Stage Treatment of Select Bicondylar Tibial Plateau Fractures. *J Orthop Trauma* 2018;32(7):327-332. doi:10.1097/BOT.0000000000001167.
53. Jevotovsky DS, Thirukumaran CP, Rubery PT. Creating value in spine surgery: using

patient reported outcomes to compare the short-term impact of different orthopedic surgical procedures. *Spine J.* 2019;19(11):1850-1857. doi:10.1016/j.spinee.2019.05.595.
54. Seifert C, Vokes J, Roberts A, Gorczyca J, Judd K. Simultaneous bilateral extensor mechanism disruptions: more than double the trouble? *J. Knee Surg.* 2019. doi:10.1055/s-0039-1688779.
55. Christensen JC, Brothers J, Stoddard GJ, et al. Higher Frequency of Reoperation With a New Bicruciate-retaining Total Knee Arthroplasty. *Clin. Orthop. Relat. Res.* 2017;475(1):62-69. doi:10.1007/s11999-016-4812-5.
56. Givens DL, Eskildsen S, Taylor KE, Faldowski RA, Del Gaizo DJ. Timed Up and Go test is predictive of Patient-Reported Outcomes Measurement Information System physical function in patients awaiting total knee arthroplasty. *Arthroplasty Today* 2018;4(4):505-509. doi:10.1016/j.artd.2018.07.010.
57. Meredith SJ, Nadarajah V, Jauregui JJ, et al. Preoperative opioid use in knee surgery patients. *J. Knee Surg.* 2019;32(7):630-636. doi:10.1055/s-0038-1666868.
58. Wojahn RD, Bogunovic L, Brophy RH, et al. Opioid consumption after knee arthroscopy. *J. Bone Joint Surg. Am.* 2018;100(19):1629-1636. doi:10.2106/JBJS.18.00049.
59. Pelt CE, Sandifer PA, Gililland JM, Anderson MB, Peters CL. Mean Three-Year Survivorship of a New Bicruciate-Retaining Total Knee Arthroplasty: Are Revisions Still Higher Than Expected? *J. Arthroplasty* 2019;34(9):1957-1962. doi:10.1016/j.arth.2019.04.030.
60. Stevens KN, Nadarajah V, Jauregui JJ, et al. Preoperative expectations of patients undergoing knee surgery. *J. Knee Surg.* 2019. doi:10.1055/s-0039-1698805.
61. 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(11):1179-1194. doi:10.1016/j.jclinepi.2010.04.011.
62. Rose M, Bjorner JB, Gandek B, Bruce B, Fries JF, Ware JE. The PROMIS Physical Function item bank was calibrated to a standardized metric and shown to improve measurement efficiency. *J. Clin. Epidemiol.* 2014;67(5):516-526. doi:10.1016/j.jclinepi.2013.10.024.
63. Rose M, Bjorner JB, Becker J, Fries JF, Ware JE. Evaluation of a preliminary physical function item bank supported the expected advantages of the Patient-Reported Outcomes Measurement Information System (PROMIS). *J. Clin. Epidemiol.* 2008;61(1):17-33. doi:10.1016/j.jclinepi.2006.06.025.
64. DeWalt DA, Rothrock N, Yount S, Stone AA, PROMIS Cooperative Group. Evaluation of item candidates: the PROMIS qualitative item review. *Med. Care* 2007;45(5 Suppl 1):S12-21. doi:10.1097/01.mlr.0000254567.79743.e2.