CRITICALLY APPRAISED TOPIC

FOCUSED CLINICAL QUESTION

For a 60 year old patient with severe knee osteoarthritis (P), is the PROMIS tool (I) more accurate than the TUG test (C) for predicting functional mobility after a total knee arthroplasty (TKA) (O)?

AUTHOR

Prepared by	Kaitlyn Sly	**Date**	November 24, 2015
Email address	Kaitlyn_sly@med.unc.edu		

CLINICAL SCENARIO

For adults over 50 years old, osteoarthritis (OA) is the leading cause of knee pain, which can limit function, decrease mobility, and reduce quality of life.[1] Approximately 700,000 adults seek surgical management of knee OA by undergoing a TKA.[2] However, there is little information about expected outcomes in the early post-operative period or how these outcomes relate to long-term prognosis.[2]

The Timed Up and Go (TUG) test is a simple outcome measure commonly used in clinical practice, which has been proven to be reliable for use in patients awaiting a TKA.[2] In addition, literature describes the TUG test's ability to predict falls in community dwelling older adults[3]; similarly, it would be beneficial if the measure had the ability to predict functional mobility following a TKA.

The Patient-Reported Outcomes Measurement Information System (PROMIS) is a relatively new tool developed using psychometric techniques to collect precise information with relative brevity.[4] Therefore, use of the PROMIS tool in clinical practice has the potential to provide prognostic information quickly and accurately.

Ultimately, researchers and clinicians alike are striving to identify measures to better predict outcomes following a TKA in order to establish prognosis, aid with discharge planning, and recognize those at risk for poor outcomes.

SUMMARY OF SEARCH

[Best evidence appraised and key findings]

- Four electronic databases were searched and 10 articles were located that met inclusion/exclusion criteria, including 2 secondary analyses of controlled trials, 4 longitudinal studies, 3 cross-sectional studies, and 1 observational study with repeated measures design. Three studies were reviewed in detail; however, there were no studies identified that directly compared the PROMIS tool to the TUG test.
- More research was available describing the utility of the TUG test for predicting functional mobility after a TKA compared to the PROMIS tool. Poorer performance on the TUG test (>10.1 seconds) pre-operatively was related to poorer performance on the same measure 6 months post-operatively; in addition, pre-operative TUG test time was a strong predictor of 6 MWT performance 6 months after a TKA.
- No research was located describing the predictive ability of the PROMIS tool for determining functional performance following a TKA. However, the PROMIS tool domains of pain intensity, interference due to pain, and fatigue have been validated for use in a patient population with OA. In addition, the PROMIS tool has adequate test-retest reliability for this patient population.

CLINICAL BOTTOM LINE

Though evidence suggests the PROMIS tool is valid for use in a patient population with OA, research demonstrates that the TUG test is useful in predicting post-operative outcomes following a TKA. Ultimately, pre-operative TUG test time is significantly related to post-operative TUG test outcomes, as well as functional mobility for longer ambulatory distances as measured by the 6MWT. By identifying patients with poor TUG test times pre-operatively, which has been described as >10.1 seconds, clinicians may identify those patients that will require more intensive and supervised rehabilitation after a TKA to minimize poor functional outcomes. Further quality research is needed to determine the use of the PROMIS tool in predicting functional mobility following a TKA.

This critically appraised topic has been individually prepared as part of a course requirement and has been peer-reviewed by one other independent course instructor

SEARCH STRATEGY

Terms used to guide the search strategy			
Patient/Client Group	**I**ntervention (or Assessment)	**C**omparison	**O**utcome(s)
Older adult* Aging Elderly Geriatric Knee osteoarthritis [MeSH] Knee pain Osteoarthritis	PROMIS Score "Patient Reported Outcomes Measurement Information System"	TUG Test "Timed Up and Go" "Up and Go"	Walk* Ambulat* Gait Function Mobility Return Knee replacement Joint replacement Total knee replacement TKR Total knee arthroplasty TKA

Final search strategy:

1. Older adult* OR aging OR elderly OR geriatric

2. Knee osteoarthritis [MeSh] OR knee pain OR osteoarthritis

3. PROMIS Score OR "Patient Reported Outcomes Measurement Information System"

4. TUG Test OR "Timed Up and Go" OR "Up and Go"

5. Walk* OR ambulat* OR gait OR function OR mobility OR return

6. Knee replacement OR joint replacement OR total knee replacement OR TKR OR total knee arthroplasty OR TKA

7. #1 AND #2 AND #3 AND #4 AND #5 AND #6 **(n=0)**

8. #2 AND #3 AND #4 **(n=0)**

9. #1 AND #2 AND #3

10. #2 AND #3

11. #1 AND #2 AND #4 AND #5 AND #6

Databases and Sites Searched	**Number of results**	**Limits applied, revised number of results (if applicable)**
PubMed		
Search Strategy Utilized:		
#9	3	
#10	4	
#11	79	**Limits English [Filter], Humans [Filter] (n=77)**
CINAHL		
Search Strategy Utilized:		
#9	0	
#10	3	
#11	4	

Cochrane Library		
Search Strategy Utilized:		
#9	0	
#10	0	
#11	6	**Include Cochrane Reviews (n=1), trials (n=5)**
Web of Science		
Search Strategy Utilized:		
#9	3	
#10	13	
#11	50	**Limits English [Filter](n=50)**

INCLUSION and EXCLUSION CRITERIA

Inclusion Criteria
• Systematic reviews, secondary analyses of randomized controlled trials/clinical trials, cohort studies, case control studies, cross-sectional studies, qualitative research, and observational studies • Published up to September 2015 • A protocol evaluating the reliability, validity, and/or predictive ability of the PROMIS and/or TUG test in determining functional mobility • Study population of older adults with a mean age of participants reported >50 years old • Study participants report a history of osteoarthritis, are planning to have a TKA, and/or have a history of a TKA • Published in English
Exclusion Criteria
• Studies that involve adults with neurological, cardiopulmonary, or autoimmune diagnoses, such as rheumatoid arthritis • Animal research • Case studies or case series • Abstracts, conference proceedings, letters to the editor, dissertations, narrative article reviews • Studies in children and young adults with a mean age of participants reported <50 years old

RESULTS OF SEARCH

A total of 10 relevant studies were located, reviewed, and categorized as shown in the following table. Study quality was scored based on a protocol developed by Downs and Black. Studies were categorized based on level of evidence described in *Guide to Evidence Based Physical Therapy Practice* by Jewell.[5]

Summary of articles retrieved that met inclusion and exclusion criteria

Author (Year)	Study quality score	Level of Evidence	Study design
Bade (2012)[6]	**21**	**1b**	**Secondary Analysis of Clinical Trial *(Decision algorithm in single population)***
Bade (2014)[2]	**20**	**2b**	**Secondary Analysis of 2-Pooled Randomized Controlled Trials**
Broderick (2013)[4]	**22**	**1b**	**Longitudinal, Cohort Design *(>80% follow up)***

Kennedy (2005)[7]	**18**	**2b**	**Observational, Repeated Measures Design *(Retrospective)***
Kennedy (2006)[8]	**18**	**4**	**Longitudinal, Repeated Measures Design *(<80% follow up)***
Ko (2013)[9]	**21**	**1b**	**Cross Sectional, Cohort Design *(>80% follow up)***
Robbins (2014)[10]	**20**	**1b**	**Cross Sectional Design *(>80% follow up)***
Storey (2013)[11]	**19**	**1b**	**Cross Sectional Design *(>80% follow up)***
Swinkels (2013)[12]	**21**	**4**	**Prospective, Longitudinal, Cohort Design *(<80% follow up)***
Zeni (2010)[13]	**17**	**1b**	**Prospective, Longitudinal Design *(>80% follow up)***

BEST EVIDENCE

The following 3 studies were identified as the 'best' evidence and selected for critical appraisal. Reasons for selecting these studies were:

- **Bade (2012)[6]:** This study evaluated patient data (mean age 64.8 years) from a previous clinical trial and created a clinical decision algorithm, qualifying this study as level 1b evidence. This study scored 21/29 when assessed using the Downs and Black checklist. The quality of this article was higher than other studies evaluated. In addition, this research is useful for addressing the clinical question because it utilizes the TUG test to predict function after a TKA. Ultimately, Bade and colleagues created an algorithm to be utilized pre-operatively to assess possible post-operative outcomes in patients 6 months after a TKA.
- **Bade (2014)[2]:** Bade published another relevant article more recently predicting functional performance post-TKA when utilizing pre-operative and acute TUG test data. Like Bade et al (2012), the research provides useful prognostic information in patients (mean age 64.6 years) 6 months after a TKA. This study utilizes data provided from two separate randomized controlled trials and provides follow up analysis, qualifying it as a lower level of evidence (2b) compared to some of the cohort studies evaluated. However, the study scored 20/29 on the Downs and Black checklist, which was higher than most of the studies evaluated.
- **Broderick (2013)[4]:** Search results yielded only one applicable study utilizing the PROMIS score in patients with OA. Though this study does not comment on the predictive value of the PROMIS score, it does provide valuable insight into the tool's validity in a patient population with OA (mean age 56.9 years). Overall, this study qualifies as level 1b evidence because of its cohort design and high follow up rate. In addition, the evidence quality was rated highly at 22/29 on the Downs and Black checklist, which suggests better quality than all the studies evaluated from this search.

SUMMARY OF BEST EVIDENCE

(1) Description and appraisal of *Predicting Poor Physical Performance after Total Knee Arthroplasty* by Bade et al (2012).[6]

Aim/Objective of the Study/Systematic Review:
The objective of the study was to develop a clinical decision algorithm to help guide patients and surgeons in deciding when a total knee arthroplasty (TKA) is warranted by predicating functional performance 6 months after surgery.

Study Design

[e.g., systematic review, cohort, randomised controlled trial, qualitative study, grounded theory. Includes information about study characteristics such as blinding and allocation concealment. When were outcomes measured, if relevant]

Note: For systematic review, use headings 'search strategy', 'selection criteria', 'methods' etc. For qualitative studies, identify data collection/analyses methods.

- The study design was a secondary analysis of an ongoing clinical trial.
    - The clinical trial is entitled *Electrical Stimulation after Total Knee Arthoplasty* (NCT00224913) and is sponsored by the Eunice Kennedy Shriver National Institute of Child Health and Human Development.[14]
    - All of the patients from the clinical trial who were evaluated at the 2 week pre-operative and 6 month post-operative time points were utilized. All subjects received a unilateral TKA in the same hospital with posterior cruciate ligament sacrificing condylar implants with patellar resurfacing. In addition, all subjects participated in standardized inpatient, home care, and outpatient physical therapy.
    - The clinical trial methods describe randomized allocation and single blind masking.
- Primary outcomes were measured at 2 different time points, 2 weeks before and 6 months after surgery. Secondary outcomes were only measured 2 weeks before surgery and were evaluated as predictors of primary outcomes 6 months after surgery.
    - Primary outcomes included functional performance measures. Secondary outcomes included joint performance, anthropomorphic, demographic, and self-report measures.
- The decision algorithm was developed using Classification and Regression Trees.

Setting

[e.g., locations such as hospital, community; rural; metropolitan; country]

All surgeries occurred in the same hospital with subsequent rehabilitation taking place in the hospital, at the patient's home, and at a single outpatient physical therapy clinic. Authors did not specify at which location pre-operative or post-operative data was collected.

Participants

[N, diagnosis, eligibility criteria, how recruited, type of sample (e.g., purposive, random), key demographics such as mean age, gender, duration of illness/disease, and if groups in an RCT were comparable at baseline on key demographic variables; number of dropouts if relevant, number available for follow-up]

Note: This is not a list of the inclusion and exclusion criteria. This is a description of the actual sample that participated in the study. You can find this descriptive information in the text and tables in the article.

- There were 119 subjects recruited for this study. Baseline regressors of patients before surgery included:
    - Sex (% males)=45.4
    - Mean age (SD)=64.8 (9.2) years
    - BMI (SD)=30.4 (4.4) kg/m^2
    - Six minute walk test (SD)=467 (119) m
    - Stair climbing test (SD)=19.9 (9.0) s
    - Timed up and go (SD)=10.0 (2.7) s
    - Knee flexion (SD)=118.0 (13.9)°
    - Knee extension (SD)=4.0 (5.0)°
    - Surgical limb quadriceps strength (SD)=19.3 (7.4) Nm^2/kg
    - Non-surgical limb quadriceps strength (SD)=23.6 (9.1) Nm^2/kg
    - Knee outcome survey (KOS)-ADLS (SD)=50.7 (16.7)%
    - SF-36 Mental component score (MCS) (SD)=55.9 (8.9)
    - SF-36 Physical component score (PCS) (SD)=31.9 (8.1)
    - Global rating of knee function (SD)=54.1 (20.9)
- A convenience sample was collected from participants who were scheduled for a TKA with participating local orthopaedic surgeons and who completed data collection for the clinical trial 2 weeks pre-operatively and 6 months post-operatively.
- Eligibility Criteria:
    - Grade 3-4 osteoarthritis in at least 1 tibiofemoral compartment
    - Scheduled to receive a TKA
- Exclusion Criteria:
    - Maximal pain in non-operative leg >4/10
    - Diagnosis of symptomatic arthritis in any other lower extremity joint
    - Cardiovascular impairment
    - Neurological impairment
    - Uncontrolled hypertension
    - Uncontrolled diabetes
    - BMI>40 kg/m^2

- Living >20 miles from designated outpatient physical therapy clinic
- Dropout rate and number available at follow up were not considered secondary to the study's design.

Intervention Investigated

[Provide details of methods, who provided treatment, when and where, how many hours of treatment provided]

Control

There was no control group for this study, and all subjects received similar surgeries, rehabilitation, and data collection.

Experimental

- Data was collected for all subjects 2 weeks before undergoing a TKA. At this time, performance based testing procedures, self-reported function questionnaires, clinical measures, and anthropometric measures were recorded.
  - Authors did not specify where data was recorded.
- All subjects received a unilateral TKA.
- After surgery, the patients participated in 3-5 days of inpatient post-operative care and 2-3 weeks of home based physical therapy.
  - It is assumed multiple physical therapists were utilized and were following the same post-operative protocol.
- At 3-4 weeks post-TKA, participants received outpatient physical therapy at one designated outpatient physical therapy clinic.
  - Rehabilitation consisted of progressive strengthening, functional retraining, manual therapy to improve ROM, and modalities to reduce pain and inflammation.
  - Subjects attended physical therapy 2-3 times per week for 6-8 weeks, with an average of 17 total treatment sessions.
  - Authors did not state if a single physical therapist or multiple therapists provided outpatient therapy.
- Only primary, performance based outcomes were re-administered at 6 months post-operatively.

Outcome Measures (Primary and Secondary)

[Give details of each measure, maximum possible score and range for each measure, administered by whom, where]

- Performance based measures were considered the primary measures of the study and were administered pre- and post-operatively. The primary measures included:
  - Timed Up and Go (TUG) test
    - Time to rise from a chair, walk 10 feet, and return to sitting. Subjects were permitted to use the arms of the chair for support during transfers from sitting to standing. The average of two trials was taken.
    - The investigator administered the test and excellent test-retest reliability was reported.
    - Data reported by Shumway-Cook et al found a cut-off time >13.5 seconds was indicative of increased risk for falls in community dwelling older adults.[3]
  - 6 Minute Walk Test (6 MWT)
    - The distance a subject walked in 6 minutes when asked to walk as quickly and safely as possible. Assistive devices were permitted. A single trial was performed per testing session.
    - The authors did not report who administered the test, but reported excellent test-retest reliability.
    - Steffen et al reported normative data for community dwelling older adults, with those 60-69 years old walking between 538-572 meters.[15]
  - Stair Climbing Test (SCT)
    - Ability to ascend and descend a set of 12 steps. Steps were 18 cm high and 28 cm deep. Bilateral handrails were available for use. An average of 2 trials was recorded in seconds.
    - A primary investigator administered the test and excellent test-retest reliability was reported.
    - Hinman et al reported adults >65 years old ascend and descend stairs at an average of 1.3 steps per second.[16]
- Self reported function questionnaires and clinical measures were secondary measures, which were only measured 2 weeks pre-TKA. Secondary measures included:
  - KOS-ADLS
    - A set of 14 questions that assess functional ability and symptoms.
    - Scores are percentage based with 100 being the highest score possible.
    - Excellent test-retest reliability was reported and the self report measure has previously demonstrated sensitivity in people with knee pathology.
    - Authors did not mention where the questionnaire was administered.
  - SF-36

        - The 36 item measure is divided into 8 subscales with scores ranging from 0 to 100, with 100 indicating positive health.
        - The average score for the U.S. population is 50.
        - The physical and mental component score of the SF-36 were calculated for this study.
        - Authors did not mention where the questionnaire was administered.
    - Global rating of knee function score
        - A single number in response to, "How would you rate your overall functional ability from 0 to 100, where 0 represents completely disabled and 100 represents normal function?"[6(p1806)]
        - This answer is reported to correlate with the KOS-ADLS and the Lysholm Knee Score.
        - Authors did not mention where the questionnaire was administered.
    - Knee ROM
        - Knee flexion and extension were measured with subject in supine. To measure flexion, subjects were asked to pull their heel towards their buttocks. To measure knee extension, a pad was placed under the heel.
        - Authors did not state who took ROM measurements, but procedure was described as using a long-arm goniometer with the axis placed at the femoral lateral epicondyle.
        - Norkin and White reported normal knee ROM between 0°-140°.[17]
    - Quadriceps strength
        - The peak isometric force produced during a volitional contraction on a Kin-Com dynamometer. The knee was positioned at 75° and subjects extended for 3 seconds.
        - The peak of 3 trials was recorded and normalized based on BMI.
        - Again, authors did not report who administered the test.

Main Findings

[Provide summary of mean scores/mean differences/treatment effect, 95% confidence intervals and p-values etc., where provided; you may calculate your own values if necessary/applicable]

- The mean TUG test for subjects at 6 months after TKA was 7.9 seconds ± 1.8 seconds.
- Using Regression Tree analysis, authors determined that individuals taking >10.1 seconds on the TUG test and were older than 72 years before surgery demonstrated the poorest performance on the TUG test 6 months post-operatively (mean TUG=10.8 seconds, 95% CI=10.0 to 11.6).
- Using Regression Tree analysis, authors determined subjects faster than 7.6 seconds on the TUG test had the best outcome on the TUG test 6 months after surgery (mean TUG=5.8 seconds, 95% CI=5.4 to 6.2).
- The mean 6 MWT for subjects at 6 months after TKA was 543 meters ± 118 meters.
- Using Regression Tree analysis, authors determined that individuals walking <314 meters before surgery demonstrated the poorest performance on the 6 MWT 6 months post-operatively (mean 6 MWT=363 meters, 95% CI=297 to 428).
- Using Regression Tree analysis, authors determined subjects walking >668 meters before surgery had the best outcome on the 6 MWT 6 months after surgery (mean 6 MWT=785 meters, 95% CI=723 to 848).
- The mean performance on the SCT for subjects at 6 months after TKA was 12.9 seconds ± 5.0 seconds.
- Using Regression Tree analysis, authors determined that individuals taking >17 seconds to complete the SCT and those who scored <40 on the SF-36 MCS before surgery demonstrated the poorest performance on the SCT 6 months after surgery (mean SCT=22.4 seconds, 95% CI=12.7 to 32.1).
- Using Regression Tree analysis, authors determined subjects performing the SCT faster than 17 seconds before surgery had the best outcome on the SCT 6 months after surgery (mean SCT= 10.3 seconds, 95% CI=9.5 to 11.1).

Original Authors' Conclusions

[Paraphrase as required. If providing a direct quote, add page number]

Poorer performance on the TUG (>10.1 seconds), 6 MWT (<314 meters), and SCT (>17 seconds) pre-operatively was related to poorer performance on the same measures 6 months post-operatively. Age (>72 years old) and decreased mental health (<40 on SF-36 MCS) before surgery were secondary predictors of poor performance 6 months post-TKA.

Critical Appraisal

Validity

[Identify the strengths and limitations of the study, including potential sources of bias. Comment on the overall methodological quality (including the score) as you determined from your assessment of the article. Comment on anything you believe was missing in the paper.]

- This study received a Downs and Black score of 21/29. This score is likely due to the study design, which utilized a secondary analysis of an ongoing clinical trial. The authors were able to use data of patients that completed the trial. This prevented patients being lost to follow up but makes the study have questionable external validity because the subjects were not representative of an entire study population.
- The sample size used to create the clinical decision algorithm was small, and authors did not report a power

analysis, which is needed to determine an appropriate sample size that will demonstrate an effect. The clinical decision algorithm needs to be validated in a larger sample of individuals to more precisely identify cut-points and better predict risk of poor outcomes following a TKA.
- A further limitation to external validity was the strict exclusion criteria applied to the clinical trial participants. Comorbidities, such as hypertension and high BMI, excluded possible participants from the study. Comorbidities are known to influence functional performance.
- External validity may be limited because the study population was given extensive rehabilitation post-TKA, which others might not receive following surgical intervention. In addition, the study participants were all located within 20 miles of an outpatient physical therapy clinical, which greatly limits external validity. However, internal validity is improved by the subjects attending the same outpatient clinic; this increases consistency of treatment and limits confounding.
- The study only examined outcomes up to a 6 month time point; therefore, results cannot be generalized past this point. However, authors mention that patients typically demonstrate the greatest strength and functional gains following a TKA at the 6 month time point.

Interpretation of Results

[This is YOUR interpretation of the results taking into consideration the strengths and limitations as you discussed above. Please comment on clinical significance of effect size / study findings. Describe in your own words what the results mean.]

The Classification and Regression Trees utilized to generate results for this study did not consider statistical significance; instead, the data from each patient was placed into a function which split once specifications were met. This allowed authors to retrospectively determine cut-points for functional performance tests based on data collected 2 weeks pre-operatively and 6 months post-operatively.

Results specific to the clinical question demonstrated that people who performed the TUG in >10.1 seconds and were >72 years old had the worst performance on the TUG test 6 months after surgery; however, those that completed the TUG in <7.6 seconds had the best outcomes on the TUG test 6 months post-TKA. These results are reasonable considering a cut-point for falls in community dwelling older adults is considered >13.5 seconds, and older adults with lower extremity osteoarthritis are at greater risk for falls compared to those that do not have OA.[3,18] However, the 95% confidence intervals that were reported, show that the individuals considered to have poor outcomes were most likely to complete the TUG test between 10.0 seconds and 11.6 seconds at their 6 month post-operative visit. These results suggest that subjects classified to have the poorest outcomes were not at risk for falls, considering >13.5 seconds as the cut-point for increased falls risk. Therefore, the applicability of these results are questionable based on clinical experience, in which older adults that have poor surgical outcomes ambulate much slower and are at an increased risk for falls compared to individuals that had positive surgical outcomes.

In addition to the cut-points established for the TUG test, results indicate that patients who ambulated <314 meters during the 6 MWT before surgery were likely to have the poorest outcomes on the 6 MWT 6 months after surgery. The 95% confidence interval reported for those considered to have poor outcomes 6 months post-operatively was between 297 meters and 428 meters. These results depict poor outcomes because normative data for healthy, older adults is reported between 538-572 meters on the 6 MWT.

Finally, those completing the stair climbing test in >17 seconds and scoring <40 on the SF-36 MCS demonstrated the poorest outcomes on the stair climbing test 6 months post-operatively. Mean time to complete the SCT in the group with the poorest outcomes 6 months post-operatively was 22.4 seconds with a rather large confidence interval reported between 12.7 seconds to 32.1 seconds. Therefore, it might be important to consider mental health as an important factor determining stair climbing performance, perhaps because patients believe stair climbing is an activity that will be painful following surgery.

(2) Description and appraisal of *Predicting Functional Performance and Range of Motion Outcomes after Total Knee Arthroplasty* by Bade et al (2014).[2]

Aim/Objective of the Study/Systematic Review:

The study's objective was to assess the predictive value of pre-operative and acute functional performance and range of motion (ROM) measures on long-term post-operative outcomes following a TKA.

Study Design

[e.g., systematic review, cohort, randomised controlled trial, qualitative study, grounded theory. Includes information about study characteristics such as blinding and allocation concealment. When were outcomes measured, if relevant]

Note: For systematic review, use headings 'search strategy', 'selection criteria', 'methods' etc. For qualitative studies, identify data collection/analyses methods.

- The study design was a secondary analysis of two pooled prospective randomized controlled trials.
    - Data from the two trials were examined before pooling. The subject's data was gathered from the control groups of two separate trials; the first trial studied the influence of neuromuscular electrical stimulation (NMES) on function after a TKA, while the second trial examined the efficacy of a minimally invasive surgery on function after a TKA.
    - The control groups in both trials were utilized for analysis in this study because both received the same tri-compartmental, cemented TKA with medial parapatellar surgical approach. In addition, all subjects received standardized acute, home, and outpatient physical therapy, as well as a standardized home exercise program.
    - Before pooling, no difference was found in subjects' age, body mass index (BMI), sex, or functional outcomes.
    - The two separate randomized controlled trials that were analysed in this study included *Minimally invasive total knee arthroplasty improves early knee strength but not functional performance: A randomized controlled trial* and *Early neuromuscular stimulation improves quadriceps muscle strength after total knee arthroplasty,* both by Stevens-Lapsley et al.[19,20]
        - Both of these studies' methods describe subject randomization and concealed allocation.
- Outcomes were measured at 3 different time points. Data was collected 1-2 weeks pre-operatively, which was considered the pre-operative time point; 48 hours after TKA, which was considered the acute time point; and 6 months post-operatively, which was considered the long-term time point.
    - Outcomes included active knee flexion and extension, TUG test, and the 6 MWT. The 6 MWT was not assessed at the acute time point.

Setting

[e.g., locations such as hospital, community; rural; metropolitan; country]

The pre-operative data and the long-term post-operative data were collected at the Clinical and Translational Research Center at the University of Colorado Hospital. Acute data was collected on the inpatient orthopaedic floor of the University of Colorado Hospital. All surgical operations were performed at the same hospital.

Participants

[N, diagnosis, eligibility criteria, how recruited, type of sample (e.g., purposive, random), key demographics such as mean age, gender, duration of illness/disease, and if groups in an RCT were comparable at baseline on key demographic variables; number of dropouts if relevant, number available for follow-up]

Note: This is not a list of the inclusion and exclusion criteria. This is a description of the actual sample that participated in the study. You can find this descriptive information in the text and tables in the article.

- There were 64 subjects recruited for this study.
    - Mean age (SD)=64.6 (8.5) years
    - Men=32, Women=32
    - BMI (SD)=30.6 (4.8) kg/m^2
- A convenience sample was collected with subjects recruited from the community with assistance of 3 orthopaedic surgeons.
- Subjects were recruited between June 2006 and June 2010.
- Eligibility Criteria:
    - Between the age of 50 and 85 years old
    - Scheduled to receive a primary unilateral TKA secondary to end-stage knee OA
- Exclusion Criteria:
    - Significant cardiac or neurological impairment
    - Contralateral knee osteoarthritis (defined by pain >5/10 with activity)
    - Other unstable lower extremity orthopaedic conditions
    - BMI >40 kg/m^2
    - Uncontrolled diabetes
    - Uncontrolled hypertension
- Dropout rate and number available at follow up were not considered secondary to the study's design.

Intervention Investigated

[Provide details of methods, who provided treatment, when and where, how many hours of treatment provided]

Control

There was no control group for this study, and all subjects received similar surgeries, rehabilitation, and data collection.

Experimental

- Data was collected for all subjects 1-2 weeks before undergoing a TKA at the University of Colorado

Hospital. At this time, active knee flexion and knee extension were measured, as well as the TUG test and 6 MWT.
- All subjects received a cemented TKA.
- After surgery, the patients participated in a standardized rehabilitation program beginning on post-operative day 1. The patients were seen twice a day while in the acute setting for rehabilitation consisting of patient education, passive and active ROM, gait training, transfer training, and stair training.
    - Rehabilitation was performed by multiple physical therapists. Physical therapists reported treatment on a detailed flow sheet in order for the study team to monitor consistency.
- Active knee flexion and extension, as well as TUG test time was collected 48 hours after surgery.
- The subjects also received 16-18 visits of outpatient or home health physical therapy for a total of 8 weeks after being discharged from the hospital.
- All patients were given a home exercise program to complete twice a day until discharged from therapy.
- Active knee flexion and extension, the TUG test, and the 6 MWT were re-assessed at 6 months post-operatively.

Outcome Measures (Primary and Secondary)

[Give details of each measure, maximum possible score and range for each measure, administered by whom, where]

- Active ROM
    - ROM was measured with a long arm goniometer with the subject in supine.
    - Active knee extension was measured with a block placed under the heel, with cues given to actively extend the knee.
    - Active knee flexion was measured by cueing the subject to maximally flex the knee while keeping the heel on the support surface.
    - Authors did not mention who took ROM measures, but did report previous findings on reliability of goniometry. Presumably, different people collected data at the pre-operative, acute, and long-term time points.
        - Good reliability for active knee flexion in acute (ICC=0.89) and outpatient settings (ICC=0.81-0.87).
        - Moderate reliability for active knee extension in an acute setting (ICC=0.69) and good reliability in an outpatient setting (ICC=0.86-0.87).
        - Norkin and White (2003) reported normal knee ROM between 0°-140°.[17]
- TUG test
    - Time to rise from a chair, walk 10 feet, and return to sitting. The use of an assistive device was permitted if needed.
    - Again, authors did not mention who administered the test, but reliability data was reported as good (ICC=0.75).
    - Data reported by Shumway-Cook et al found a cut-off time >13.5 seconds was indicative of increased risk for falls in community dwelling older adults.[3]
- 6 MWT
    - Subjects completed the test in an indoor hallway with a 100 foot walkway. Subjects were instructed to walk back and forth to cover as much distance as possible in 6 minutes. Rest breaks and use of assistive devices were permitted.
    - The 6 MWT demonstrates good reliability (ICC=0.94).
    - Steffen et al reported normative data for community dwelling older adults, with those 60-69 years old walking between 538-572 meters.[15]
    - This measure was not performed in the acute setting, and only taken at the pre-operative and long-term time points.

Main Findings

[Provide summary of mean scores/mean differences/treatment effect, 95% confidence intervals and p-values etc., where provided; you may calculate your own values if necessary/applicable]

- Pre-operative knee flexion (mean (SD)=119.7° (15.2°)) was predictive of long-term knee flexion (mean (SD)=115.1° (10.5°)) (β=0.44, 95% CI=0.31 to 0.58, r^2=0.42, p<0.001).
- Acute knee flexion (mean (SD)=66.6° (16.7°)) was not related to pre-operative knee flexion (β=0.03, 95% CI=-0.20 to 0.26, r^2=0.001, p=0.80).
- Acute knee flexion was not related to long-term knee flexion (β=0.09, 95% CI=-0.07 to 0.26, r^2=0.005, p=0.26).
- Pre-operative knee extension (mean (SD)=1.5° (5.8°)) was predictive of long term knee extension (mean (SD)=0.8°(4.9°)) (β=0.46, 95% CI=0.27 to 0.64, r^2=0.29, p<0.001).
- Acute knee extension (mean (SD)= 10.4° (4.8°)) demonstrated no relationship with pre-operative knee extension (β=0.05, 95% CI=-0.25 to 0.35, r^2=0.002, p=0.76).
- Acute knee extension demonstrated no relationship with long-term knee extension (β=0.04, 95% CI=-0.22 to 0.30, r^2=0.00242, p=0.76).
- Pre-operative TUG performance (mean (SD)=9.0s (3.5s)) was predictive of long-term 6MWT (mean (SD)=483.3m (95.8m)) performance (β=-211, 95% CI=-255 to -165, r^2=0.60, p<0.001).

- Acute TUG performance (mean (SD)=69.0s (54.3s)) was related to pre-operative 6 MWT (mean (SD)=452.7m (124m)) performance (β=-61, 95% CI=-107 to -14, r^2=0.10, p=0.01).
- Acute TUG performance was related to long-term 6 MWT performance (β=-62, 95% CI=-97 to -28, r^2=0.18, p<0.001).
- A hierarchical linear regression was performed to assess the contribution of changes in TUG performance to changes in long-term 6 MWT performance.
    - Age was significantly related to long-term 6 MWT performance (p=0.04).
    - Age + sex was significantly related to long-term 6MWT performance (p=0.008).
    - Age + sex + acute TUG was significantly related to long-term 6MWT performance (p=0.02).
    - Age + sex + acute TUG + pre-operative TUG was significantly related to long-term 6MWT performance (p<0.001).
    - However, in the overall model once pre-operative TUG was added, the acute TUG time was no longer predictive of long-term 6 MWT performance (p=0.65).

Original Authors' Conclusions

[Paraphrase as required. If providing a direct quote, add page number]

Acute measures of knee ROM are not related to long-term ROM outcomes after a TKA; however, pre-operative ROM did have prognostic value. Both pre-operative TUG and acute TUG measures were predictive of long-term functional performance on the 6 MWT; however, pre-operative TUG time was a stronger predictor of long-term 6 MWT performance compared to acute TUG time.

Critical Appraisal

Validity

[Identify the strengths and limitations of the study, including potential sources of bias. Comment on the overall methodological quality (including the score) as you determined from your assessment of the article. Comment on anything you believe was missing in the paper.]

- This study received a Downs and Black score of 20/29. This score is mostly due to the nature of the study design, using a secondary analysis of two pooled randomized controlled trials. Little information was presented about randomization, and the authors were able to use the data of patients that completed the previous studies. This prevented any patients being lost to follow up but makes the study have questionable external validity because subjects were not representative of an entire study population.
- A further limitation to external validity was the strict exclusion criteria of the randomized controlled trials which the subject data was collected. Exclusion criteria prevented those with comorbidities from participating in the study. Comorbidities are known to influence ROM and functional performance.
- External validity might further be limited by the extensive rehabilitation which subjects received. Others who undergo a TKA might not have access to as extensive physical therapy services.
- Potential confounding was limited because authors determined there was no difference in subject's age, BMI, sex, or functional outcomes before pooling the data from the two separate randomized controlled trials.
- Authors mention limitations during collection of acute ROM data, noting differences in medication dosing and anaesthesia type adding variability between subjects and possibly influencing acute ROM measures.
- Authors did not report a power analysis, which is needed to determine an appropriate sample size that will demonstrate an effect.

Interpretation of Results

[This is YOUR interpretation of the results taking into consideration the strengths and limitations as you discussed above. Please comment on clinical significance of effect size / study findings. Describe in your own words what the results mean.]

Results specific to the clinical question demonstrate that pre-operative TUG performance was predictive of 6 MWT performance 6 months after a TKA. Scores reached statistical significance with a p-value found to be less than 0.001 and the confidence interval did not include 0. In addition, r^2 was found to be 0.60 which indicates that the pre-operative TUG Test accounts for about 60% of the variability in the data. Acute TUG performance was also found to be related to 6 MWT performance with statistical significance reached, and the confidence interval did not include 0; however, this model explained less variability in the data with r^2=0.18. Authors completed a linear regression analysis to assess the contribution of changes in long-term 6 MWT performance; researchers determined that acute TUG time was no longer predictive of long-term 6 MWT performance after pre-operative TUG performance was added into the model; the p-value was found to be 0.65 which is not statistically significant. This indicates that pre-operative TUG performance is a better predictor of long-term 6 MWT outcomes compared to acute TUG performance.

Pre-operative knee flexion and knee extension ROM were predictive of long-term knee ROM with p-values found below 0.001 for both instances, demonstrating statistical significance. Acute ROM measures were not predictive of long-term outcomes with p-values found to be greater than 0.05 and confidence intervals including 0. In

addition, r^2 was found to be less than 0.01 for both acute measures of knee flexion and extension which demonstrates that acute ROM does not explain variability in the data.

(3) Description and appraisal of *Validity and Reliability of Patient-Reported Outcomes Measurement Information System Instruments in Osteoarthritis* by Broderick et al (2013).[4]

Aim/Objective of the Study/Systematic Review:

The study's objective was to examine the known-group validity, ecological validity, and reliability of several PROMIS domains in patients with OA compared to the general population.

Known-group validity was operationally defined as a significant difference between 2 groups that are expected to show a difference. Ecological validity was operationally defined as the degree to which patient reported outcome (PRO) based recall over a period of time corresponded with momentary or daily ratings.

Study Design

[e.g., systematic review, cohort, randomised controlled trial, qualitative study, grounded theory. Includes information about study characteristics such as blinding and allocation concealment. When were outcomes measured, if relevant]

Note: For systematic review, use headings 'search strategy', 'selection criteria', 'methods' etc. For qualitative studies, identify data collection/analyses methods.

- The study was a longitudinal, cohort design.
- Data were collected from subjects with OA and subjects of the general population.
    - Known-group validity was determined by comparing subjects with OA to a sample from the general population.
    - Ecological validity was determined by comparing PROMIS computerized adaptive testing (CAT), which asked subjects to record information about the past 7 days, to short form versions of the PROMIS, which were collected daily. CAT allows for measurement to occur with very few items with subsequent questions based upon the person's response to preceding items.
- Data collection occurred over 4 weeks on a daily basis. Participants completed daily PROMIS short forms for 28 consecutive days. On days 7, 14, 21, and 28 the PROMIS CAT was administered via a free, online data collection tool. Participants were trained over telephone to use the online assessment tool. It is assumed participants were responsible for using a personal or public computer to complete the weekly computerized assessments.
- The PROMIS domains examined for this study included pain intensity, pain interference, fatigue, and physical functioning.
    - The general population was not asked to answer questions pertaining to physical functioning due to response burden.
- No randomization occurred due to the study's design. In addition, it is assumed that participants were not blinded to group placement because subjects with OA were required to provide a doctor confirmed diagnosis of OA.
- Known-group validity was measured using analysis of variance (ANOVA) between subjects with OA and the general population.
- Ecological validity was determined through between-person correlations, separately for the two groups.
- Test-retest reliability was calculated using interclass correlation coefficients (ICC).

Setting

[e.g., locations such as hospital, community; rural; metropolitan; country]

All recruitment and data collection occurred online. Therefore, it is assumed participants completed daily and weekly surveys at home or via a public computer. Participants did have a telephone interaction with study coordinators to be trained to use the online PROMIS Assessment Center and/or if a daily assessment was missed.

Participants

[N, diagnosis, eligibility criteria, how recruited, type of sample (e.g., purposive, random), key demographics such as mean age, gender, duration of illness/disease, and if groups in an RCT were comparable at baseline on key demographic variables; number of dropouts if relevant, number available for follow-up]

Note: This is not a list of the inclusion and exclusion criteria. This is a description of the actual sample that participated in the study. You can find this descriptive information in the text and tables in the article.

- General population sample:
    - N=98
- Osteoarthritis sample:
    - N=98

○ Mean age (SD)=43.9 (14.8) years ○ Arthritis diagnosis=19 ○ Women=50 ○ Race: ▪ White=69 ▪ African American=15 ▪ Asian=6 ▪ American Indian=2 ▪ Other=6 ○ Hispanic=14 ○ Married=45 ○ Education: ▪ <High school=1 ▪ High school graduate=17 ▪ Some college=42 ▪ College graduate=28 ▪ Advanced degree=10 ○ Family income: ▪ $0-$20K=6 ▪ $20K-$35K=22 ▪ $35K-$50K=28 ▪ $50K-$75K=22 ▪ $75K+=20 ○ Employed=70 ○ Disability benefits=10	○ Mean age (SD)=56.9 (10.0) years ○ Arthritis diagnosis=98 ○ Women=59 ○ Race: ▪ White=77 ▪ African American=15 ▪ Asian=1 ▪ American Indian=1 ▪ Other=4 ○ Hispanic=11 ○ Married=46 ○ Education: ▪ <High school=1 ▪ High school graduate=10 ▪ Some college=54 ▪ College graduate=23 ▪ Advanced degree=10 ○ Family income: ▪ $0-$20K=23 ▪ $20K-$35K=28 ▪ $35K-$50K=21 ▪ $50K-$75K=12 ▪ $75K+=13 ○ Employed=31 ○ Disability benefits=30

- Participants in the OA sample were significantly older, had a lower income, and were more likely to be receiving disability benefits.
- Eligibility criteria for both samples:
  - Age 21 years or older
  - Fluency in English
  - Availability for 29-36 days
  - High speed internet access
- Additional eligibility criteria for participants in OA group:
  - Doctor confirmed diagnosis of OA
- No exclusion criteria were specified. Authors stated that sampling of the general population group was structured to match the composition of the US according to the 2009 Census Bureau. For sampling in the OA group, recruitment was structured to approximate the composition based on US prevalence rates for arthritis.
- Both groups were a convenience sample recruited from a national online research panel.
- Participants were contacted via telephone if an assessment was missed and compensated $150 for completion of the study. Missing daily values were replaced with a set of plausible values.
- Four subjects dropped out of the study, 2 from each group. Data from dropouts were not included in analyses.
- Compliance for daily PROMIS short-forms was high with 4% missed in the general population sample and 4.4% missed in the OA sample. PROMIS CAT collected at the end of each week was missed 0.5% of the time in both samples.

Intervention Investigated

[Provide details of methods, who provided treatment, when and where, how many hours of treatment provided]

Control

- A sample of the general population was collected through a national online research panel. Sampling of the general population produced a sample that was demographically comparable to the 2009 US population.
- Each participant completed a daily short form PROMIS for 28 days and a weekly PROMIS CAT on days 7, 14, 21, and 28. In addition, participants provided comorbidity information at enrollment.
- PROMIS domains included pain intensity, pain interference, and fatigue.
- Short forms and the PROMIS CAT were completed online through the PROMIS Assessment Center and participants were trained via telephone on using the online program. It is assumed that participants used their personal computer or a public computer with internet access.
- Authors did not describe the approximate time it took to complete each survey.

Experimental

- A sample of participants with OA was collected through a national online research panel. Sampling characteristics were similar to reported US prevalence rates for arthritis.
- Each participant completed a daily short form PROMIS for 28 days and a weekly PROMIS CAT on days 7,

14, 21, and 28. In addition, participants provided comorbidity information at enrollment.
- PROMIS domains included pain intensity, pain interference, fatigue, and physical functioning.
- Short forms and the PROMIS CAT were completed online through the PROMIS Assessment Center and participants were trained via telephone on using the online program. It is assumed that participants used their personal computer or a public computer with internet access.
- Authors did not describe the approximate time it took to complete each survey.

Outcome Measures (Primary and Secondary)

[Give details of each measure, maximum possible score and range for each measure, administered by whom, where]

- Four PROMIS domains were included in the study:
    - Pain intensity assessed participants' average self reported pain.
    - Pain interference item bank measured consequences of pain on daily life.
    - The fatigue item bank measured the symptoms from mild feelings of tiredness to overwhelming exhaustion.
    - The physical functioning item bank measured self-reported capability of upper extremities, lower extremities, and activities of daily living.
        - The general population sample did not complete this domain due to response burden.
- These 4 domains were measured with a daily PROMIS short form and compared with Computerized Adaptive Testing (CAT) utilizing the PROMIS domains.
    - PROMIS Short Form
        - Completed daily and online using the PROMIS Assessment Center.
        - Questions worded as "In the last day..."
        - The short form consisted of 1 pain intensity item, 6 pain interference items, 7 fatigue items, 10 physical functioning items.
        - The form was static, but scoring was completed using a PROMIS scoring engine to allow direct comparison to the PROMIS CAT.
    - PROMIS CAT
        - Completed online using the PROMIS Assessment Center.
        - Data collected weekly on days 7, 14, 21, and 28.
        - Questions worded as "In the last 7 days..."
        - The CAT instrument was set to administer between 4-12 questions and end once a pre-set score was reached.
        - The instrument was dynamic and questions were selected based on the participant's response to the preceding items.
- Scores are reported as a T-score which is anchored to the distribution of scores in the U.S. general population. Therefore, values range from 0 (low) to 100 (high) with a score of 50 being the expected mean value for the U.S. population for domains of pain interference, fatigue, and physical functioning. Pain intensity is measured from a single value 1-10 on a numeric rating scale, with authors reporting the general population pain average as 2.6 units.

Main Findings

[Provide summary of mean scores/mean differences/treatment effect, 95% confidence intervals and p-values etc., where provided; you may calculate your own values if necessary/applicable]

- Known-group validity was determined based on the following results:
    - Mean pain intensity (SD) of the OA group (weekly=5.51 units (1.9); daily=5.23 units (1.9)) was significantly greater ($p<0.001$) than the general population sample (weekly=2.48 units (2.4); daily=2.07 (2.2)). Large effect size reported (weekly (*d)*=1.42; daily (*d*)=1.53) indicating that OA has a large effect on pain intensity.
    - Mean pain interference (SD) of the OA group (weekly=60.9 units (6.1); daily=58.2 units (6.4)) was significantly greater ($p<0.001$) than the general population sample (weekly=51.3 units (9.0); daily=48.6 (7.7)). Large effect size reported (weekly (*d)*=1.25; daily (*d*)=1.37) indicating that OA has a large effect on pain interference.
    - Mean fatigue (SD) of the OA group (weekly=56.2 units (7.8); daily=52.2 units (9.1)) was significantly greater ($p<0.001$) than the general population sample (weekly=48.8 units (9.6); daily=44.2 (9.7)). Large effect size reported (weekly (*d)*=0.85; daily (*d*)=0.84) indicating that OA has a large effect on fatigue.
    - Mean physical function of the OA group was reported as 37.5 units (6.8) for the weekly CAT and 36.9 units (6.5) for the daily short form. No comparison available to the general population sample.
- Ecological validity between daily PROMIS short forms and weekly PROMIS CATs was determined based on the following results:
    - Strong, positive correlation (95% CI) reported between PROMIS CAT and PROMIS short form for the pain intensity domain in the OA sample (0.94 (0.91-0.96)) and general population sample (0.93 (0.90-0.95)).
    - Strong, positive correlation (95% CI) reported between PROMIS CAT and PROMIS short form for the pain interference domain in the OA sample (0.89 (0.85-0.91)) and general population sample (0.88 (0.83-0.91)).

    - Strong, positive correlation (95% CI) reported between PROMIS CAT and PROMIS short form for the fatigue domain in the OA sample (0.89 (0.86-0.91)) and general population sample (0.86 (0.78-0.91)).
    - Strong, positive correlation (95% CI) reported between PROMIS CAT and PROMIS short form for the physical functioning domain in OA sample (0.90 (0.87-0.92)).
- Authors reported high test-retest reliability between weekly PROMIS CAT scores, as well as high test-retest reliability between aggregated daily PROMIS short form scores.
    - Pain intensity: General population sample (ICC)=0.89-0.93; OA sample (ICC)=0.83-0.84.
    - Pain interference: General population sample (ICC)=0.84-0.87; OA sample (ICC)=0.80-0.83.
    - Fatigue: General population sample (ICC)=0.84-0.86; OA sample (ICC)=0.85.
    - Physical functioning: OA sample (ICC)=0.92-0.95.

Original Authors' Conclusions

[Paraphrase as required. If providing a direct quote, add page number]

The PROMIS CAT and PROMIS short form instruments for pain intensity, interference due to pain, and fatigue demonstrated known-group validity and ecological validity in a comparison of OA patients with a general population sample. Moreover, ecological validity was demonstrated for the physical functioning domain in the subjects with OA but was not measured in the general population sample. In addition, good test-retest reliability was observed for the PROMIS CAT and PROMIS short form domains in a population with OA and a general population sample.

Critical Appraisal

Validity

[Identify the strengths and limitations of the study, including potential sources of bias. Comment on the overall methodological quality (including the score) as you determined from your assessment of the article. Comment on anything you believe was missing in the paper.]

- This study received a Downs and Black score of 22/29. This score reflects the lack of randomization consistent with the study design.
- External validity is questionable because a national internet panel was used to recruit subjects until certain characteristics were met to match the US profiles for the general population and for people with OA. Overall, this allows for subjects to well represent an entire population; however, this method of recruitment limits subjects without internet access and/or without income to afford internet access. In addition, since the study was survey based it limited potential subjects with low education or those that are illiterate from participating.
- Authors assumed the data provided by participants occurred during a period of medical stability. If patients were experiencing increases in pain due to OA or initiated a new treatment during a certain week, the findings may not be valid.
- The study had a low drop-out rate (2%) and a high response rate on daily and weekly surveys (95%); however, for those surveys that were not completed, the examiners entered a potentially plausible value based on the person's previously recorded data set, which could introduce bias in results.
- In addition, it is assumed the study groups were not blinded because the OA group was required to confirm a medical diagnosis of OA. Potentially, this could bias their answers to the survey.
- Authors did not state the time period of recruitment, which could potentially increase selection bias if groups were recruit at different time periods.
- There is uncertainty on reasons authors included the physical functioning domain but did not collect data from the general population sample.
- A power analysis was completed and the study was appropriately powered to demonstrate an effect.

Interpretation of Results

[This is YOUR interpretation of the results taking into consideration the strengths and limitations as you discussed above. Please comment on clinical significance of effect size / study findings. Describe in your own words what the results mean.]

Results specific to the clinical question do not address the ability of the PROMIS to predict clinical outcomes following a TKA. However, the results do demonstrate the validity and reliability of the PROMIS short form and the PROMIS CAT for people with OA. Known-group validity, which is the ability of the instrument to detect change when expected, was confirmed for PROMIS domains of pain intensity, pain interference, and fatigue. For all domains tested in both samples, a statistically significant difference was detected with $p<0.001$. In addition, a large effect size was found between groups, demonstrating that the PROMIS instrument was able to detect the large effect ($d>0.8$) of OA on pain intensity, pain interference, and fatigue compared to the general population.

Ecological validity, which is the ability for daily PROMIS short form scores to correspond with weekly PROMIS CAT scores, was confirmed with strong, positive correlations (0.86-0.94) and narrow confidence intervals

(≥0.78) reported between the 2 assessment methods.

Finally, test-retest reliability was found to be appropriate for weekly PROMIS CAT scores and aggregated daily PROMIS short form scores. In research, test-retest reliability should be >0.70 (ICC), which was the case for this study. Potentially, the PROMIS tool could be utilized clinically and in research for people with OA because it has been shown to be a valid and reliable measure in this patient population.

EVIDENCE SYNTHESIS AND IMPLICATIONS

The total number of TKAs performed in the United States annually is steadily increasing as knee OA is becoming more prevalent in an aging population.[1,2] Though surgical interventions are performed regularly, little is known about prognosis, discharge planning, and/or recognizing people at risk for poor outcomes.[2] Therefore, identifying an outcome measure which has the ability to predict function in patients with OA that opted for a TKA would be useful for therapists, physicians, and surgeons to promote positive patient outcomes. The TUG test is a common outcome measure used by physical therapists as an assessment of functional mobility. Previous research by Shumway-Cook et al determined a cut-off time >13.5 seconds as a risk for falls in community dwelling older adults.[3] Similar attempts to determine cut-off times to predict functional outcomes following a TKA have been made. Bade et al (2012) determined that the TUG test may be used to predict function in patients following a TKA.[6] A TUG test time >10.1 seconds in patients that were >72 years old had the poorest results on the TUG test 6 months after surgery.[6] In addition, Bade et al (2014) further explored the use of the TUG test's ability to predict longer ambulatory distances as measured by the 6 MWT.[2] Bade et al (2014) determined that pre-operative TUG test time was significantly related to distance walked during the 6 MWT at 6 months post-operatively.[2]

The PROMIS tool has recently been developed to track patient reported outcomes using psychometrics to accurately, but briefly collect information using responses from answers to preceding questions.[4] The evidence reviewed in this clinically appraised topic did not identify any relevant evidence to support the use of the PROMIS tool to predict functional mobility following a TKA in patients with knee OA; however, more pertinent literature is starting to be published about the PROMIS tool as clinicians and researchers notice its utility. Broderick et al validates the use of the PROMIS tool in a population with OA compared to a general population sample, validates the use of the computerized version of the PROMIS tool compared to the PROMIS short form, and describes adequate test-retest reliability of the tool.[4]

Implications for Clinical Practice

Physical therapists may utilize this information in clinical practice. The PROMIS tool can be used to quickly and accurately collect patient reported outcomes in those with OA; however, the available research suggests the TUG test has greater use to predict functional mobility following a TKA. By administering the TUG test pre-operatively, therapist may identify those patients at risk for poor outcomes following a TKA. Bade et al (2012) utilized Regression Tree analysis to determine those completing a TUG test >10.1 seconds and >72 years old have greater risk for poor functional outcomes following surgery.[6] In addition, the TUG test is significantly related to walking distance as measured by the 6 MWT.[2] This information may allow therapists to identify those patients that will require more intensive and supervised rehabilitation after a TKA to minimize poor functional outcomes. In addition, it may assist therapists and doctors in an acute care setting with discharge planning following a TKA by identifying patients that might need increased support if returning home or referral to a more supervised rehabilitation setting after discharge from the hospital.

There are several considerations that clinicians should understand when interpreting these results. First, the studies utilizing the TUG test only followed outcomes of participants for 6 months. Therefore, results cannot be generalized past this time point. Second, the exclusion criteria utilized in the 2 articles by Bade et al were relatively strict and did not include patients with common comorbidities, such as hypertension or high BMI, which might decrease the external validity of the results.[2,6] Third, longer ambulatory distances measured by the 6 MWT in research by Bade et al (2014) might not be applicable to functional ambulation, such as on uneven or unstable surfaces, because the test was performed in a flat hallway with little interference.[2] Finally, the clinical decision algorithm described by Bade et al (2012) should be validated in a larger patient population to allow for more precise identification of cut-points and better predict risk for poor outcomes.[6]

Implications for Future Research

Ultimately, the evidence describing the ability of the TUG test and PROMIS tool to predict functional mobility following a TKA is sparse, with no evidence directly comparing the 2 outcome measures. In addition, there is no evidence to describe the ability of the PROMIS tool to predict functional mobility following a TKA. The quality of evidence describing the predictive ability of the TUG test is good, with 9 studies reviewed utilizing study designs appropriate for prognosis. However, these studies lack randomization and/or do not utilize a control group for comparison, which reduces methodological quality. In addition, future research correlating TUG test time to functional activities, such as walking on an unstable surface, up/down inclines, and around obstacles would be beneficial in predicting functional mobility.

There is very little research available for the PROMIS tool as it relates to a patient population with OA. As more domains are developed, validity and reliability will need to be established. As more research continues to be published, the use of the tool may be more widely accepted. Therefore, research of high methodological quality,

utilizing a large sample size should be conducted to determine the predictive ability of the PROMIS tool in patients with OA that are undergoing a TKA. Future research comparing the TUG test and the PROMIS tool would be beneficial to determine an accurate outcome measure which can predict functional mobility following a TKA with the intention of promoting the best patient outcomes following surgery.

REFERENCES

1. Nguyen US, Zhang Y, Zhu Y, Niu J, Zhang B, Felson DT. Increasing prevalence of knee pain and symptomatic knee osteoarthritis: Survey and cohort data. *Ann Intern Med*. 2011;155(11):725-732.

2. Bade MJ, Kittelson JM, Kohrt WM, Stevens-Lapsley JE. Predicting functional performance and range of motion outcomes after total knee arthroplasty. *Am J Phys Med Rehabil*. 2014;93(7):579-585.

3. Shumway-Cook A, Brauer S, Woollacott M. Predicting the probability for falls in community-dwelling older adults using the timed up & go test. *Phys Ther*. 2000;80(9):896-903.

4. Broderick J, E., Schneider S, Junghaenel D, U., Schwartz J, E., Stone A, A. Validity and reliability of patient-reported outcomes measurement information system instruments in osteoarthritis. *ARTHRITIS CARE RES (2151464X)*. 2013;65(10):1625-1633.

5. Jewell DV. Appendix A: Hierarchy of evidence for articles about prognosis. In: *Guide to evidence based physical therapy practice.* Sudbury, Massachusetts: Jones and Bartlett; 2008:423.

6. Bade MJ, Wolfe P, Zeni JA, Stevens-Lapsley JE, Snyder-Mackler L. Predicting poor physical performance after total knee arthroplasty. *J Orthop Res*. 2012;30(11):1805-1810.

7. Kennedy DM, Stratford PW, Wessel J, Gollish JD, Penney D. Assessing stability and change of four performance measures: A longitudinal study evaluating outcome following total hip and knee arthroplasty. *BMC Musculoskelet Disord*. 2005;6:3.

8. Kennedy DM, Hanna SE, Stratford PW, Wessel J, Gollish JD. Preoperative function and gender predict pattern of functional recovery after hip and knee arthroplasty. *J Arthroplasty*. 2006;21(4):559-566.

9. Ko V, Naylor JM, Harris IA, Crosbie J, Yeo AE. The six-minute walk test is an excellent predictor of functional ambulation after total knee arthroplasty. *BMC Musculoskelet Disord*. 2013;14:145.

10. Robbins SM, Rastogi R, McLaughlin TL. Predicting acute recovery of physical function following total knee joint arthroplasty. *J Arthroplasty*. 2014;29(2):299-303.

11. Storey AS, Myrah AM, Bauck RA, Brinkman DM, Friess SN, Webber SC. Indoor and outdoor mobility following total knee arthroplasty. *Physiother Can*. 2013;65(3):279-288.

12. Swinkels A, Allain TJ. Physical performance tests, self-reported outcomes, and accidental falls before and after total knee arthroplasty: An exploratory study. *Physiother Theory Pract*. 2013;29(6):432-442.

13. Zeni JA, Jr., Snyder-Mackler L. Early postoperative measures predict 1-and 2-year outcomes after unilateral total knee arthroplasty: Importance of contralateral limb strength. *Phys Ther*. 2010;90(1):43-54.

14. Eunice Kennedy Shriver National Institute of Child Health and Human Development. Electrical stimulation after total knee arthroplasty. In ClinicalTrials.gov [Internet]. Bethesda (MD): National Library of Medicine (US). Available from: https://clinicaltrials.gov/. NLM Identifier: NCT00224913.

15. Steffen TM, Hacker TA, Mollinger L. Age- and gender-related test performance in community-dwelling elderly people: Six-minute walk test, berg balance scale, timed up & go test, and gait speeds. *Phys Ther*. 2002;82(2):128-137.

16. Hinman MR, O'Connell JK, Dorr M, Hardin R, Tumlinson AB, Varner B. Functional predictors of stair-climbing speed in older adults. *J Geriatr Phys Ther*. 2014;37(1):1-6.

17. Norkin C, White DJ. *Measurement of joint motion.* 4th ed. Philadelphia: F.A. Davis Company; 2009:450.

18. Dore AL, Golightly YM, Mercer VS, et al. Lower-extremity osteoarthritis and the risk of falls in a community-based longitudinal study of adults with and without osteoarthritis. *Arthritis Care Res (Hoboken)*. 2015;67(5):633-639.

19. Stevens-Lapsley JE, Bade MJ, Shulman BC, Kohrt WM, Dayton MR. Minimally invasive total knee arthroplasty improves early knee strength but not functional performance: A randomized controlled trial. *J Arthroplasty*. 2012;27(10):1812-1819.

20. Stevens-Lapsley J, Balter JE, Wolfe P, Eckhoff DG, Kohrt WM. Early neuromuscular electrical stimulation to improve quadriceps muscle strength after total knee arthroplasty: A randomized controlled trial. *Phys Ther*. 2012;92(2):210-226.