Osteoarthritis and Cartilage Open 5 (2023) 100410

Contents lists available at ScienceDirect

Osteoarthritis and Cartilage Open

journal homepage: www.elsevier.com/journals/osteoarthritis-and-cartilage-open/2665-9131

Virtual performance measure in osteoarthritis: An innovative transformation of patient care

Helen Razmjou [a,b,c,*], Suzanne Denis [a,b], Susan Robarts [a,b], James Falconer [a], Mark Anunciacion [a,b], Andrea Nunn [a,b], Amy Wainwright [a,b], Patricia Dickson [a,b], Roxana Najafi [a,b], John Murnaghan [c,d,e]

[a] Bone and Joint Program, Holland Orthopedic & Arthritic Centre, Sunnybrook Health Sciences Centre, Toronto, Canada
[b] Department of Physical Therapy, Faculty of Medicine, University of Toronto, Toronto, Canada
[c] Sunnybrook Research Institute, Sunnybrook Health Sciences Centre, Toronto, Canada
[d] Division of Orthopedic Surgery, Department of Surgery, Sunnybrook Health Sciences Centre, Toronto, Canada
[e] Division of Orthopaedic Surgery, Department of Surgery, Faculty of Medicine, University of Toronto, Toronto, Canada

ARTICLE INFO

Handling Editor: Professor H Madry

Keywords:
Virtual
Hybrid
Performance measures
Validity
Reliability
Accuracy

ABSTRACT

Objectives: The purpose of this study was to develop and establish reliability and validity of a virtual performance measure (VPM) score that encompassed 10 videos in patients with osteoarthritis of the knee joint. Patients' experience and satisfaction were documented.

Design: Forty videos were chosen for 10 functional tasks, with four videos showing increasing difficulty for each task. Patients were requested to choose the video that best reflected their own situation. Clinical and radiological findings and self-report and performance measures were completed.

Results: Data of 100 patients, 70 (70%) females, mean age: 65 ± 9 were examined. The Cronbach's alpha coefficient that examined internal consistency of the VPM score was 0.92. The intraclass correlation value of 0.82 was obtained for test-retest reliability. Factor analysis showed three distinct domains. There was moderate correlations between the VPM score and the self-report and actual performance measures ranging from r = 0.46 to 0.66. The VPM summated score of 10 activities was able to differentiate between candidates and non-candidates for knee arthroplasty, with the area under the curve value of 0.90 indicating excellent predictive validity. The overall patient experience and satisfaction was positive with 67% of participants feeling that virtual care could have an impact on minimizing physical presence in the clinic or hospital.

Conclusions: The VPM is a reliable and valid outcome measure in patients with osteoarthritis of the knee joint. This digital tool has the potential to transform osteoarthritis care by providing a valid remote measurement of real-life functional limitations and reduce the burden of time consuming in-person tests.

1. Introduction

One essential component of a comprehensive patient examination in osteoarthritis (OA) care is the assessment of limitations in daily physical activities, often measured by Patient-Reported outcome measures (PROMs) [1,2] or performance-based tests, which provide a snap shot of the patient's physical abilities [2,3]. In 2020, the use of performance measures became challenging due to pandemic-related restrictions to in-person visits and the need for alternative ways to assess a patient's function became more obvious. While the utilization of remote care and digital transformation of health care was fueled by the pandemic, virtual care is now a vital component of routine care. Validation of technology-based outcome instruments is critical as they will have an impact on quality of health care delivery at present and in the future.

The virtual performance measure (VPM) is an innovative tool that uses digital technology and computerized videos to assess real-life physical limitations without requiring patients to travel for in-person visits in poor weather, or present to high risk environments. The virtual performance measure combines advantages of the self-report and performance-based measures without having some of their weaknesses. The self-report PROMs are highly subjective to the perception of functional difficulty and rely on patient's experience, recall, and emotional

* Corresponding author. Holland Orthopaedic and Arthritic Centre, 43 Wellesley Street East, Toronto, Ontario, Canada M4Y 1H1.
E-mail address: helen.razmjou@sunnybrook.ca (H. Razmjou).

https://doi.org/10.1016/j.ocarto.2023.100410
Received 12 September 2023; Accepted 19 September 2023

state. In addition, they may be difficult to complete by patients who are not fluent in the language the tool is utilized in. The performance-based tools are dependent on patient motivation to complete the test. They are time consuming and require trained personnel and specific testing tools and facilities. They are subject to observer bias and inadequate instructions could affect their accuracy. A digital tool that can be watched and scored remotely can overcome some of these weaknesses. With the increasing utility of internet and digital technology in the health care system, a hybrid approach to outcome measurement may be a pathway to a more flexible and efficient outcome assessment.

The history of using videos as an outcome instrument is fairly recent, dating back to 2014, when the first generation of video-based virtual measures, the computerized animated activity questionnaire (AAQ) were created by a group of Dutch investigators [4,5]. These researchers used avatars (an animated figure representing a patient in the video) to present different levels of difficulty executing daily tasks (e.g. no difficulty to inability to perform a task) in patients with OA of the hip and knee joints. The use of AAQ has shown promising utility in arthritis care [4–6]. While animated avatar models are used for many purposes in today's technology, human models and particularly, live actors with an appropriate age, may be more acceptable to older persons or those with advanced disabilities who use assistive devices. An avatar may not have the same realistic expressions or credibility as a live actor, especially when an activity or movement is associated with a facial expression or a struggle to perform a task due to pain. In addition, older adults who are new to the digital world may have difficulty identifying with a virtual character that has imperfections or awkward body movements. Most importantly, the cost of avatar videography is high and unless there is in-house expertise in using animation software, organizations may be required to outsource production which presents challenges in updating videos over time and as needs change for accurate representations. In short, virtual videos that use human models may be more relatable to an elderly population and their utility in medical videos warrants further investigation.

The primary objective of this study was to establish reliability (internal consistency and test-retest), validity (convergent validity, known-group validity, and factorial validity) of the VPM total score (summated score of 10 different activities watched on 10 videos) in the North American patients with OA of the knee joint. The secondary objective focused on patient experience and satisfaction in relation to quality and applicability of the videos, and their role in reducing in-person hospital visits. It was hypothesized that the human-model based VPM would provide an accurate assessment of function as compared with traditional self-report and in-person performance-based outcome measures and that the majority of patients would express satisfaction with these videos.

2. Methods

This was a cross-sectional study of patients with knee arthritis who were referred to the Hip/Knee Rapid Access Clinic of an academic tertiary care centre for consideration of joint replacement. Exclusion criteria included inability to read English, lack of access to internet or inability to use online surveys. Patients were contacted before their appointment and after they agreed to participate in the study, they were sent a link to the VPM videos. The study protocol was approved by the Human Ethics Research Board of the Sunnybrook Health Sciences Centre (ID# 3703), and patients provided informed consent for participation in the study.

2.1. Virtual performance measure video development

The videos were developed based on input from physical therapists, occupational therapists, and patient representatives. Three versions of the videos were made, of which 40 videos with an age-appropriate model and a white background were chosen for 10 functional tasks. Four videos represented each functional task with the first video showing normal execution of the task and three videos showing increased difficulty as mild, moderate and severe. Patients were asked to choose the video that best reflected their own situation. Choice of "unable to perform" was given for each activity. The first and second functional tasks were siting and rising from a chair with hand rests. The third series of videos simulated different levels of difficulty putting on and taking off socks. The fourth and fifth set of videos simulated getting into the shower and picking up an item from the floor respectively. The sixth and seventh set of videos demonstrated sitting down and getting up from the floor. The eighth set showed walking on even ground and finally the ninth and tenth set of videos demonstrated ascending and descending stairs. All functional tasks had a "not able to perform" option (Table 1). Patients could watch videos that loaded simultaneously as many times as they wished using their smart phone or any computer/tablet. The right leg was used as representing the affected leg and was labeled with a red tape. Red arrows were used in the videos to highlight subtle body adaptations for each task (e.g. using the upper extremity for help, bending down from the trunk, etc.). The videos were hosted on a temporary webpage of an academic centre; one of the largest hip and knee joint replacement centres in Canada, performing over 2000 procedures annually. Patients did not require any specific training to use the videos. Figs. 1–3 show a selected set of videos.

2.2. Clinical examination

The clinical examination included taking a history, observation, and assessment of range of motion. Clinicians also completed a standardized assessment tool; the osteoarthritis Severity Scoring System to determine appropriateness for surgical consultation. This tool was developed by a skilled team of clinicians as a part of a new model of care at our institute in 2007 and has been widely used in Ontario since 2017. The Severity Scoring System assists clinicians to quantify findings related to arthritis and has demonstrated validity in patients with hip and knee arthritis [7,8]. The score is based on three distinct components, clinical (history and physical examination, and pain intensity, ranging from 0 to 3), functional (scores from self-report and performance measures, ranging from

Table 1
Components and scores of the Virtual Performance Measure (VPM) tool.

Functional task	No difficulty	Minimum difficulty	Moderate difficulty	Severe difficulty	Unable to perform
Sitting down on a chair	8	6	4	2	0
Rising from a chair	8	6	4	2	0
Putting on/taking off socks	8	6	4	2	0
Getting into the shower	8	6	4	2	0
Picking up an object from the floor	8	6	4	2	0
Sitting on the floor	8	6	4	2	0
Rising from the floor	8	6	4	2	0
Walking on a flat surface	8	6	4	2	0
Ascending stairs	8	6	4	2	0
Descending stairs	8	6	4	2	0

The responses to 10 questions are added up with a minimum of zero and maximum of 80. The score out of 80 is then transformed to a score out of 100 to a more clinically meaningful score, using the following formula: Total score = {(score)*100}/80 with the score of 100 corresponding to the best or normal function and lower scores referring to lower levels of functioning with zero being completely disabled.

4. Getting into the shower *

Video 1 Video 2 Video 3 Video 4 Unable to perform

Fig. 1. Video set 4. Getting into the shower.

8. Walking on a flat surface *

Video 1 Video 2 Video 3 Video 4 Unable to perform

Fig. 2. Video set 8. Walking on the flat surface.

0 to 3), and radiological examination (severity of arthritis on imaging, scored from a rating of 0–2: no findings, mild/moderate and marked/advanced). The total severity score is calculated by summing all three scores, ranging between 0 and 8 with higher numbers indicating a higher severity.

2.3. Patient-Reported outcome measures (PROMS)

The PROMS used in this study were the lower extremity functional score (LEFS) [9] and P4 (4-item pain intensity measure) [10]. The LEFS has 20 questions with four responses for each question and the total score ranges from 0 to 80 with the highest numbers indicating better function. The items of the P4 address pain in the morning, afternoon, evening, and with activity. Each item is scored on an 11-point numeric pain scale, with higher numbers indicating higher pain. LEFS and P4 are reported to have sufficient levels of reliability and validity in patients with lower extremity arthritis [9–13].

2.4. In-person performance-based measures

Patients executed two performance-based tests with an experienced clinician: the 30- second Chair Stand Test (CST) [14] that documents the maximum number of chair stand repetitions possible in a 30 s period and the 40-m fast-paced walk test [15]. In this test, subjects walk along a 10 m walkway as quickly and as safely as possible, and then turn around at a cone, return and repeat again for a total distance of 40 m. These performance measures are reported to be valid and reliable in patients with hip/knee arthritis [2,15,16].

2.5. Patient experience and satisfaction

Patient experience/satisfaction was explored anonymously by a 5-point Likert scale in the overall quality of the videos (lighting, angle, and speed), applicability to activity limitations, and the patient's impression of the role of virtual care in minimizing physical presence in

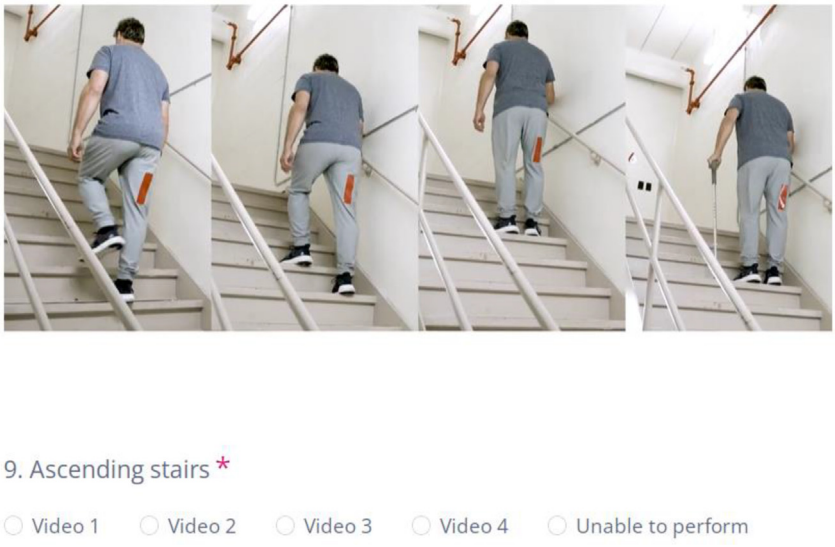


Fig. 3. Video set 9. Ascending stairs.

the hospital. Patients had the option of providing input in free text as well.

2.6. Sample size justification

For test-retest reliability of the continuous data using intraclass correlation (ICC) coefficients, a minimum of 19 patients is required to examine the ICC of 0.70 in the null hypothesis against a higher value of 0.90 in the alternative hypothesis to achieve statistical significance for an alpha 0.05 and power of 80% based on a1-tailed Type I error [17]. For convergent validity, Spearman's correlations (95% CI) was calculated between the total scores of the VPM and the scores of 40-m fast-paced walk and 30 sec-CST. Previous studies [5–8] that have used animated videos have shown correlation Coefficient of >0.60 between the total score of animated physical activities and self-report and performance measures. With the Spearman Coefficient of 0.60, Fisher's desired confidence interval (CI) of 0.30 and $\alpha = 0.5$, a minimum of 86 patients were required. The sample size of the factor analysis was based on prevalent rule-of-thumb of subject to item ratios of 10:1 [18]. We therefore collected data on 100 patients with moderate to severe OA of the knee joint.

2.7. Statistical analysis

Each question of the VPM survey has five options with the "unable to perform" response having a score of zero and no difficulty having a score of 8. The responses to 10 questions are then added up with a minimum of zero and maximum of 80. The score out of 80 is then transformed to a score out of 100 to a more clinically meaningful score, using the following formula: Total score = {(score)*100}/80 with the score of 100 corresponding to the best or normal function and lower scores referring to lower levels of functioning with zero being completely disabled.

2.8. Reliability

As a measure of reliability, internal consistency and test-retest reliability of the single summed total score of the VPM were explored. Internal consistency of the VPM total score was examined using Cronbach's alpha coefficients. This analysis examined the pattern of association between different activities. An inter-item correlation of 0.70 was considered necessary for an acceptable alpha coefficient [19].

For test-retest reliability of the total score of the VPM (continuous data), we examined the reproducibility of the results by ICC statistics [20] on 20 subjects who had completed the videos twice within a day as the disability level may change day to day in arthritis. The values less than 0.5, between 0.5 and 0.75, between 0.75 and 0.9, and greater than 0.90 are indicative of poor, moderate, good, and excellent reliability [21].

2.9. Factorial validity

Factor analysis is a commonly used analysis to establish construct validity of clinically applicable instruments. Exploratory Factor Analysis helps to uncover the underlying latent structure of constructs of an instrument. Factorial validity examined the degree to which covariance among the 10 video responses resembled the covariation of underlying physical activities. We used an oblique rotation method which allows constructs to be correlated.

2.10. Cross-sectional convergent validity

For cross-sectional convergent validity (ability of a measure to align with another measure of related construct), the VPM total score was validated against traditional tools, the LEFS, P4, and actual performance measures (40-m fast-paced walk and CST). Convergent validity was estimated with Spearman's correlations. Correlation coefficients between 0.90 and 1.0 indicate very high correlation. Correlation coefficients whose magnitude are between 0.70 and 0.89, and between 0.4 and 0.69 indicate high and moderate correlations respectively. Correlation coefficients of <0.40 indicate low correlation [22]. We hypothesized a moderate to high correlations between the VPM total score and LEFS and P4.

The correlation between the VPM total score and the osteoarthritis severity scoring system that is based on the overall clinical findings, functional abilities, and radiological findings was examined. We expected that higher VPM score (better function) would have a moderate reverse relationship with the total severity score, in which higher numbers indicate a higher severity.

2.11. Known-group validity

The ability of the VPM total score to differentiate between candidates for knee arthroplasty surgery and those requiring non-surgical management was explored using general linear model (GLM) analysis. We expected that candidates for immediate surgical consultation would have more functional difficulties. It was also expected that the VPM score

would be statistically significantly lower in patients who were using a walking aid (cane, rollator, or wheelchair). In terms of biological differences, we expected that women would have more difficulty than men in performing their functional activities as measured by the total score of the VPM.

The area under the curve (AUC) of the receiver operator characteristic curve was calculated to examine the overall accuracy and predictive ability of the VPM total score [23]. An AUC of 1.0 represents perfect differentiation by the diagnostic instrument, with 100% sensitivity and specificity; an area of 0.90 and higher indicates excellent accuracy and an area of 0.80 and higher indicates good accuracy. The Youden's J (sensitivity-specificty-1) index is often used for selecting an optimal cutoff value and was calculated to identify the optimal VPM score [24,25] for consideration of knee arthroplasty.

2.12. Patient perspective, experience and satisfaction

Patient experience and satisfaction were examined on the descriptive basis, using frequency and percentage.

3. Results

Data of 100 patients, 70 (70%) females, 30 (30%) males, average age: 65 ± 9, min age = 44 and max = 83 years, with OA of the knee joint were examined. The demographic characteristics of the sample is presented in Table 2.

3.1. Reliability

The Cronbach's alpha coefficient that examined internal consistency reliability of the 10 videos was 0.92, indicating high reliability. It is noteworthy that a high value of alpha (>0.90) may suggest redundancies and a potential for shortening of the survey.

The ICC value on a random sample of 20 patients was 0.82 (X^2 = 22.44, $p < 0.0001$, CI $_{95\%}$: 0.75–0.90) for continuous total score of the VPM, which based on the lower confidence interval fell in a good category for test-retest reliability [21].

3.2. Factorial validity

Factor analysis of 10 activities that used oblique rotation, assuming correlation between factors showed three distinct domains (Table 3). The first domain included the four most challenging activities that required squatting or deep loaded bending of the knee: 1) picking up an object from the floor, 2) sitting on the floor, 3) rising from the floor, and 4) putting on/taking off socks. The second factor involved knee flexion while balancing/stabilizing the upper body: 1) sitting down on a chair, 2) rising from a chair, and 3) getting into the shower. The third factor involved upright activities such as 1) walking on a flat surface and navigating the stairs (descending and ascending). There was no cross-loading of the activities at 0.5 significance level, indicating that the three factors were distinct and represented separate concepts. The VPM score contained uniquely reliable variance to warrant their separate interpretation, but when aggregated they formed a coherent single dimension.

3.3. Convergent and known-group validity

The correlations between the VPM total score and the self-report and performance measures were moderate: LEFS ($r = 0.66$, $p < 0.0001$), CST score ($r = 0.60$, $p < 0.0001$), P4 ($r = 0.47$, $p < 0.0001$) and the 40-m fast-paced walk test ($r = 0.46$, $p < 0.0001$). In terms of severity score that incorporated the clinician's impression, the correlation was highest for the functional abilities ($r = -0.61$, $p < 0.0001$) and lowest for radiological findings ($r = -0.25$, $p = 0.01$). The total severity score correlated with the VPM score at −0.42, $p < 0.0001$. The total score of the VPM was able to differentiate between candidates for knee arthroplasty ($F = 30.66$, $p < 0.001$), and those who required a walking aid vs. those who did not use any assistive devices ($F = 30.06$, $p < 0.001$). The VPM score also differentiated between men and women ($F = 4.50$, $p = 0.036$) with women having more functional difficulty (Graphs1–2). The factor of age did not appear to affect the VPM total score ($F = -0.10$, $p = 0.90$) in this sample.

Table 2
Characteristics of sample included (N = 100).

Variables	Number (Percentage/SD)
Age: Mean (SD)	65 ± 9, min 44, max 83
Sex: n (%)	
• Female	70 (70%)
• Male	30 (30%)
Affected knee side: n (%)	
• Left	47 (47%)
• Right	53 (53%)
• Bilateral involvement	46 (46%)
Previous arthroplasty surgery: n (%)	
• No	83 (83%)
• Hip joints	7 (7%)
• Opposite knee	9 (9%)
• Opposite knee and hip	1 (1%)
Walking devices: n (%)	
• Yes	36 (36%)
• No	64 (64%)
Gait: n (%)	
• Normal	16 (16%)
• Abnormal	84 (84%)
Range of motion examination: mean (SD)	
• Flexion	116 (21)
• Flexion contracture (n = 64)	4.2 (4.3)
• Extension	3.6 (7)
• Extension lag (n = 63)	0.81 (2.9)
Severity Score: mean(SD)	
• Clinical (0–3)	1.79 (0.56)
• Functional (0–3)	1.76 (0.65)
• Radiological (0–2)	1.73 (0.46)
• Total score (0–8)	5.3 (1.3)
Outcome measures	
• LEFS	35 (16)
• P4	23 (9)
• Chair Sit Test	11 (5)
• 40 m fast paced walk	34 (9)
Proceeding to surgeon consult	
• No	28 (28%)
• Yes, Candidate for arthroplasty	60 (60%)
• Yes, Second opinion	12 (12%)

SD: Standard Deviation.
LEFS: Lower Extremity functional score.
P4: 4-item pain intensity measure.

Table 3
Factor analysis: Promax oblique rotation.

Rotated Factor Pattern (Standardized Regression Coefficients)			
	Factor1	Factor2	Factor3
Sitting down	.	0.77480	.
Rising from chair	.	0.91533	.
Getting into shower	.	0.75527	.
Walking on flat surface	.	.	0.96377
Descending stairs	.	.	0.67368
Ascending stairs	.	.	0.56198
Rising from floor	0.83365	.	.
Siting on the floor	0.78023	.	.
Putting on/taking off socks	0.67549	.	.
Picking from floor	0.85670	.	.

Values less than 0.5 are not printed.
Promax Rotation, an oblique rotation technique was used to allow factors to be correlated. Factor analysis examined the degree to which covariance among the responses resembled the relationship between physical activities.

The logistic regression that calculated the AUC values showed an AUC of 0.90, indicating excellent predictive validity in relation to candidacy for arthroplasty. The total score of less than 58 out of 100 on VPM score appeared to have the highest sensitivity (0.95) and could indicate the cut off for surgical candidacy.

3.4. Patient perspective, experience, and satisfaction

Satisfaction for the overall quality of the videos in terms of lighting, speed, and angle was reported to be excellent by 64% of the patients, good by 34%; and fair and poor by two patients respectively. Applicability of the videos was rated as excellent by 38%, good by 44%, fair by 14% and poor by 3%, with one missing response. Sixty seven (67%) participants felt that virtual care could have a positive impact on minimizing physical presence in the hospital. A number of patients felt that using a female or obese model will be helpful as gender or certain characteristics may affect the performance of a task.

4. Discussion

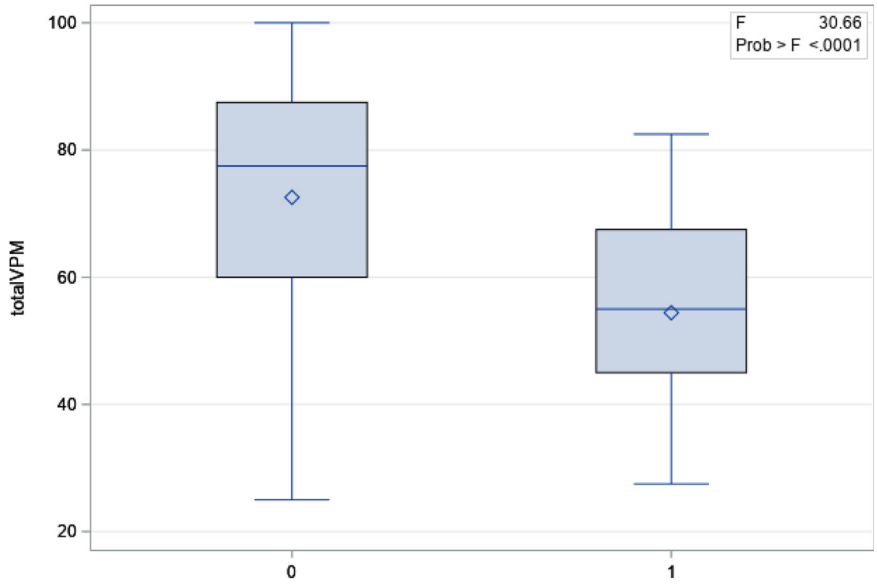
The present study showed that a virtual performance measure that utilized 10 sets of videos using a computer-administrated questionnaire was an accurate and reliable tool for remote assessment of functional abilities of patients with OA of the knee joint. This innovative diagnostic tool does not require any formal training and could facilitate remote assessment of patients with limited access to skilled clinicians and during situations where the presence of the patient in a clinical setting is affected by pandemic, patient distance from the facility, or adverse weather conditions.

To our knowledge, a human model has never been used to represent levels of activity and related difficulty for the OA population. The animated activity questionnaires that have used avatars have proven to be reliable and valid in the osteoarthritis population [4–7,26,27]. Nevertheless, using a live actor within an appropriate age group and proper facial expressions is more acceptable to older persons. In our study, the VPM total score could differentiate between sexes, which may indicate using male and female models may not be necessary, although that may improve accuracy.

Completion of PROMs is inexpensive and convenient. However, they are highly subjective to the perception of “difficulty” or “inability”, are affected by patient's mental and emotional wellbeing and may be difficult to complete by patients with cognitive difficulties or those who speak a different language. The moderate correlation of the VPM as a hybrid instrument with the LEFS, a self-report functional measure and P4 shows that these scales measure a fairly similar concept, despite different methods of application.

The performance measures involve the actual assessment of the person's abilities, nevertheless, they are time consuming (which is a barrier to their use), expensive (require trained personnel, space), and are inconvenient (require physical presence of the patient) [13]. In addition, they capture a snapshot of patient's abilities at the time of assessment, are subject to observer bias, instruction bias, and the motivation of the participant to perform the test [28]. The moderate associations between the VPM and the CST and 40-m fast walk test is promising as the VPM is less costly, does not require an examiner/clinician, and provides a convenient alternative to an in-person visit. Obviously, access to a phone or computer and internet is required to watch the videos.

In terms of structure and factorial validity of the videos, three distinct domains represented three different levels of complexity of functional tasks. In factor analysis, the factor loading of each variable quantifies the extent to which the variable is related to a given factor. The deep bending and/or squatting activities that are affected by pain, reduced range of



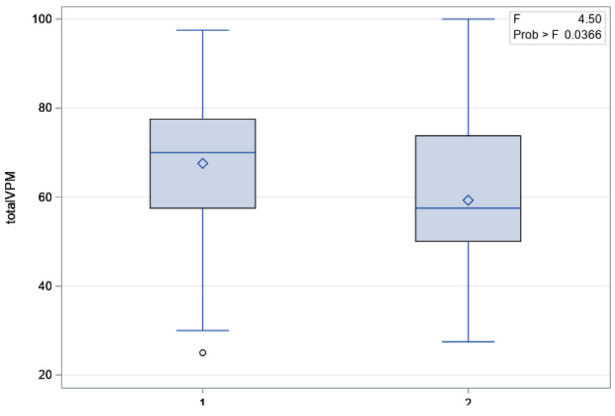
Group 0: Non-candidates for surgery

Group 1: Candidates for knee arthroplasty

Prob: Probability

F: Fisher’s value

Graph 1. Distribution of virtual performance measure total score between candidates and non-candidates for knee arthroplasty.



Group1: Male participants

Group2: Female participants

Prob: Probability

F: Fisher's value

Graph 2. Distribution of virtual performance measure total score between men and women.

motion, and quadriceps weakness were lumped together in one factor. The next domain incorporated sitting and raising from a chair and lifting the leg while getting into a shower, all requiring postural balance and trunk stability, coordination, axial loading, and strength of the lower extremity. Activities in the third domain involved upright activities that required forward locomotion and simultaneous translation of the body, such as walking and stair management.

Research has shown that rising from a chair creates greater torques about the knee joint and higher pressures at the hip joint than activities such as walking and stair-climbing [29,30]. Riley et al. [29] have shown that the transfer phase of the rising, which follows the lift-off from the chair demands a significant degree of coordination. The role of postural balance and increased displacement of the center of mass in the knee OA population during the stand to sit task is highlighted by Fu and colleagues [31] and Gross et al. [30] have proven that chair-rise biomechanics are affected by the reduction in muscle strength of particularly hip musculature. The established relationship between the single-leg standing balance (e.g. while entering the shower) and knee pain, reduced flexion and quadriceps strength may explain the similarity of this task with the stand-sit activities [32].

In summary, the results of the present study highlight the potential utility of the VPM videos to revolutionize patient care and improve the patient experience. A valid and accurate virtual assessment tool could reduce the number of time-consuming in-person performance-based tests and reduce the burden on both the patient and clinician while facilitating the necessary comprehensive assessment for the surgical evaluation of patients with knee OA. Administering virtual outcome measures in an optimal location such as the patient's home will reduce the number and the time constraints of a typical clinical encounter with the orthopedic surgeon or the advanced practice practitioners, improving safety and efficiency of arthritis care.

4.1. Limitations

This study examined the initial stages of the development of a virtual outcome measurement tool in patients diagnosed with arthritis of the knee joint seen at a tertiary care centre. This may affect its generalizability to community clinics or younger patients. Validation is an ongoing process and further assessment of the VPM is required to increase the applicability of this innovative tool in other populations. There is a potential that the survey can be shortened and this warrants further examination in an independent sample.

4.2. Future directions

Considering, measurement properties of diagnostic tests apply to a specific purpose and context in a particular population and are established incrementally as evidence increases, future studies are needed to examine the suitability of the VPM in other lower extremity joints. In addition, longitudinal studies are required to estimate the responsiveness, sensitivity to change, and the minimal clinically important change for making decisions at the individual patient level. Larger studies should examine which factor or domain has the highest ability to differentiate between patients (candidates vs. non-candidates, men vs. women, etc.) for a potential reduction of the number of videos and a shorter version of the VPM.

5. Conclusions

The VPM score showed clinically acceptable reliability and validity in patients with OA of the knee joint. This tool has the potential to transform osteoarthritis care by providing a valid remote measurement of real-life functional limitations.

Author contributions

HR conceptualized the research question and study design, supervised data collection, conducted data analysis, and drafted the manuscript. SD, SR, JM, AW, PD were involved in study design and obtaining grants. SD, JF, and JM assisted with patient requirement and data management. HR, SD, MA, AN, and RN were involved in video development. All authors are accountable for the integrity of the work and have reviewed and approved the final version of the manuscript.

Funding

This study was funded by the Alternative Funding Program Innovation Funds and the Practice Based Research and Innovation of the Sunnybrook Health Sciences Centre.

Declaration of competing interest

The authors have no conflicts of interest to declare.

Acknowledgements

This study was funded by the Alternative Funding Program Innovation Funds, and the Practice Based Research and Innovation funds of the Sunnybrook Health Sciences Centre. The study sponsors had no involvement in publication of the results.

The authors would like to thank Brent Creelman, the former Digital Communications Strategist and Lindsay Pinter, the Digital Communications Specialist at Sunnybrook Health Sciences Centre for their contribution to this study. We thank Keith Breutigam and Paul Dunn, our models of the final and preliminary videos and Perry Walker, the videographer of the VPM tool. The time and contribution of our patient representatives and the rehabilitation team who participated in lengthy discussions is greatly appreciated. We thank the research team of the Sunnybrook Research Institute for their assistance in patient recruitment and acknowledge the contribution of Marni Wilson to data entry and database management.

References

[1] R. Gandhi, D. Tsvetkov, J.R. Davey, K.A. Syed, N.N. Mahomed, Relationship between self-reported and performance-based tests in a hip and knee joint replacement population, Clin. Rheumatol. 28 (3) (2009) 253–257.

[2] P.W. Stratford, D.M. Kennedy, L.J. Woodhouse, Performance measures provide assessments of pain and function in people with advanced osteoarthritis of the hip or knee, Phys. Ther. 86 (11) (2006) 1489–1496.

[3] C.B. Terwee, C. Coopmans, W.F. Peter, L.D. Roorda, R.W. Poolman, V.A. Scholtes, J. Harlaar, H.C. de Vet, Development and validation of the computer-administered animated activity questionnaire to measure physical functioning of patients with hip or knee osteoarthritis, Phys. Ther. 94 (2) (2014) 251–261.
[4] W.F. Peter, H.S. Dagfinrud, N. Osteras, C.B. Terwee, Animated Activity Questionnaire (AAQ), a new method of self-reporting activity limitations in patients with hip and knee osteoarthritis: comparisons with observation by spouses for construct validity, Muscoskel. Care 15 (3) (2017) 263–271.
[5] W.F. Peter, H.C.W. de Vet, C.B. Terwee, Reliability of the Animated Activity Questionnaire for assessing activity limitations of patients with hip and knee osteoarthritis, Muscoskel. Care 16 (3) (2018) 363–369.
[6] W.F. Peter, R.W. Poolman, V.A.B. Scholtes, H.C.W. de Vet, C.B. Terwee, Responsiveness and interpretability of the Animated Activity Questionnaire for assessing activity limitations of patients with hip or knee osteoarthritis, Muscoskel. Care 17 (4) (2019) 327–334.
[7] S. Robarts, V. Palinkas, D. Boljanovic, H. Razmjou, Validity of the virtual severity scoring system in patients with knee osteoarthritis, Orthop Procs 105 (SUPP_8) (2023) 51.
[8] S. Robarts, S. Denis, D. Kennedy, P. Dickson, S. Juma, V. Palinkas, et al., Patient gender does not influence referral to an orthopaedic surgeon by advanced practice orthopaedic providers: a prospective observational study in Canada, BMC Health Serv. Res. 21 (1) (2021 September 11) 952–955.
[9] J.M. Binkley, P.W. Stratford, S.A. Lott, D.L. Riddle, The lower extremity functional scale (LEFS): scale development, measurement properties, and clinical application. North American orthopaedic rehabilitation research network, Phys. Ther. 79 (4) (1999) 371–383.
[10] G.F. Spadoni, P.W. Stratford, P.E. Solomon, L.R. Wishart, The evaluation of change in pain intensity: a comparison of the P4 and single-item numeric pain rating scales, J. Orthop. Sports Phys. Ther. 34 (4) (2004) 187–193.
[11] P.W. Stratford, M. Dogra, L. Woodhouse, D.M. Kennedy, G.F. Spadoni, Validating self-report measures of pain and function in patients undergoing hip or knee arthroplasty, Physiother. Can. 61 (4) (2009) 189–194, discussion 195-6.
[12] D.M. Kennedy, P.W. Stratford, D.L. Riddle, S.E. Hanna, J.D. Gollish, Assessing recovery and establishing prognosis following total knee arthroplasty, Phys. Ther. 88 (1) (2008) 22–32.
[13] S.P. Mehta, A. Fulton, C. Quach, M. Thistle, C. Toledo, N.A. Evans, Measurement properties of the lower extremity functional scale: a systematic review, J. Orthop. Sports Phys. Ther. 46 (3) (2016) 200–216.
[14] C.J. Jones, R.E. Rikli, W.C. Beam, A 30-s chair-stand test as a measure of lower body strength in community-residing older adults, Res. Q. Exerc. Sport 70 (2) (1999) 113–119.
[15] F. Dobson, R.S. Hinman, E.M. Roos, J.H. Abbott, P. Stratford, A.M. Davis, R. Buchbinder, L. Snyder-Mackler, Y. Henrotin, J. Thumboo, P. Hansen, K.L. Bennell, OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis, Osteoarthritis Cartilage 21 (8) (2013) 1042–1052.
[16] H. Master, G. Coleman, F. Dobson, K. Bennell, R.S. Hinman, J.T. Jakiela, D.K. White, A narrative review on measurement properties of fixed-distance walk tests up to 40 meters for adults with knee osteoarthritis, J. Rheumatol. 48 (5) (2021) 638–647.
[17] M.A. Bujang, N. Baharum, A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient: a review, Arch. Orofac. Sci. 12 (1) (2017) 1–11.
[18] A.B. Costello, J. Osborne, W.: best practices in exploratory factor analysis: four recommendations for getting the most from your analysis, Practical Assess. Res. Eval. 10 (7) (2005). Aug 25, 2019, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.110.9154&rep=rep1&type=pdf.
[19] J.C. Nunnally, Psychometric Theory, second ed., McGraw-Hill, New York, 1978.
[20] L.G. Portney, M.P. Watkins, Foundations of Clinical Research: Applications to Practice, Prentice Hall, New Jersey, 2000.
[21] T.K. Koo, M.Y. Li, A guideline of selecting and reporting intraclass correlation coefficients for reliability research, J Chiropr Med 15 (2) (2016) 155–163.
[22] P. Schober, C. Boer, L.A. Schwarte, Correlation coefficients: appropriate use and interpretation, Anesth. Analg. 126 (5) (2018) 1763–1768.
[23] A.K. Akobeng, Understanding diagnostic tests 3: receiver operating characteristic curves, Acta Paediatr. 96 (5) (2007) 644–647.
[24] W.J. Youden, Index for rating diagnostic tests, Cancer 3 (1) (1950) 32–35.
[25] S.I. Bangdiwala, A.S. Haedo, M.L. Natal, A. Villaveces, The agreement chart as an alternative to the receiver-operating characteristic curve for diagnostic tests, J. Clin. Epidemiol. 61 (9) (2008) 866–874.
[26] W.F. Peter, H.C.W. de Vet, M. Boers, J. Harlaar, L.D. Roorda, R.W. Poolman, V.A.B. Scholtes, M. Steultjens, G.J. Hendry, E.M. Roos, F. Guillemin, M.G. Benedetti, L. Cavazzuti, A. Escobar, H. Dagfinrud, C.B. Terwee, Cross-Cultural and construct validity of the animated activity questionnaire, Arthritis Care Res. 69 (9) (2017) 1349–1359.
[27] C.D. do Nascimento, W.F. Peter, I.M. Ribeiro, BdS. Moreira, V.P. Lima, R.N. Kirkwood, AdC. Bastone, Cross-cultural validity of the Animated Activity Questionnaire for patients with hip and knee osteoarthritis: a comparison between The Netherlands and Brazil, Braz. J. Phys. Ther. 25 (6) (2021) 767–774.
[28] H. Wittink, W. Rogers, A. Sukiennik, D.B. Carr, Physical functioning: self-report and performance measures are related but distinct, Spine 28 (20) (2003) 2407–2413.
[29] P.O. Riley, M.L. Schenkman, R.W. Mann, W.A. Hodge, Mechanics of a constrained chair-rise, J. Biomech. 24 (1) (1991) 77–85.
[30] M.M. Gross, P.J. Stevenson, S.L. Charette, G. Pyka, R. Marcus, Effect of muscle strength and movement speed on the biomechanics of rising from a chair in healthy elderly and young women, Gait Posture 8 (3) (1998) 175–185.
[31] S. Fu, T. Duan, M. Hou, F. Yang, Y. Chai, Y. Chen, B. Liu, Y. Ma, A. Liu, X. Wang, L. Chen, Postural balance in individuals with knee osteoarthritis during stand-to-sit task, Front. Hum. Neurosci. 15 (2021) 760960.
[32] M.A. Hunt, F.J. McManus, R.S. Hinman, K.L. Bennell, Predictors of single-leg standing balance in individuals with medial knee osteoarthritis, Arthritis Care Res. 62 (4) (2010) 496–500.