

Using gold standard patient-reported outcome measures in clinical practice – A new approach to facilitate their use



Abstract

Introduction: To analyze two gold-standard Patient-Reported Outcome Measures (PROMs) in knee OA (WOMAC and SF-36) and determine which questions are the most reflective of the overall score.

Methods: This was a retrospective study on 4,983 patients with primary knee pain. Patients had WOMAC and SF-36 at two-time points, pre-treatment and after three months of treatment. A decision tree classifier supported with a linear mix model regression was applied to determine, identify, and categorize the most influential questions that determine the overall score in each of the questionnaires.

Result: For SF-36, the most influential items were Q22 (39%), Q32 (24%), Q11 (19%), Q25 (19%). For WOMAC, the most influential predictors were Q14 (39%), Q10 (24%) and Q15 (21%). A significant improvement in WOMAC and SF-36 was seen after three months of treatment ($p < 0.01$). For SF-36, the main predictor items were Q11, Q22 and Q32, Regression model $R^2 = 0.841$, $p < 0.01$, $t[55.62] = 0.001$, Beta for Q22 = 0.409, Q32 = 0.352, Q11 = 0.278. For WOMAC, the main predictor items were Q10 and Q15, Regression model $R^2 = 0.930$, $p < 0.01$, $t[35.4] = 0.001$, Beta for Q15 = 0.548, Q10 = 0.4639.

Conclusion: Two questions from the WOMAC questionnaire predicts 93% of the overall score and four questions from the SF-36 predict 84%. The creation of a clinically meaningful assessment tool based on larger scientifically validated PROMs will help to facilitate its use by clinicians and acceptance by patients in clinical practice.

Keywords: patient reported outcome measures, clinical practice, knee

Abbreviations: PROMs: Patient-Reported Outcome Measures; MSK: Musculoskeletal; VAS: Visual Analogue Scale; WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index; EQ-5D: EuroQol Five Dimensions Questionnaire; SF-12: Short Form-12; SF-36: Short Form-36

Introduction

Patient-Reported Outcome Measures (PROMs) are self-administered questionnaires that are used to assess a patient's health state, quality of life, and functional status associated with their health condition without the interpretation of the physician or anyone else [1,2]. There are growing efforts to shift from using PROMs in health research to implementing them in clinical practice [2-4]. Integrating PROMs in clinical practice can serve the entire health care system, including patients, care providers, insurers, and government regulators, and will enhance high-quality clinical care and improve shared decision-making processes [1,5,6]. From a patient's point of view, this will help to quantify health status, monitor changes over time, help to set up expectations, and increase patient engagement [5,7].

PROMs in Musculoskeletal (MSK) conditions are essential to facilitate patient-clinician communication and improve the shared decision-making process. Adding assessments from the patient's perspective provides a patient centered approach that will help to assess disease severity as well as the effectiveness of treatments [2,8,9]. There are some commonly used disease-specific PROMs in MSK conditions, amongst them are the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) and the Short Form-36 (SF-36) [10,11].

Implementing PROMs in clinical practice is still a challenge [10]. The current integration of PROMs in clinical practice is minimal as they are considered complex and resource-intensive [4,12,13]. In essence, there are several barriers to real-life implementation and the adoption of PROMs in clinical practice. Amongst these are skepticism about the validity and potential

**Vikram Madan^{1*},
Avi Elbaz², Amit Mor²,
Yiftah Beer³, Matthew
N Bartels¹ and Sandeep
Yerra¹**

¹Department Rehabilitation Medicine, Montefiore Medical Center, Albert Einstein College of Medicine, USA

²AposTherapy Research Group, Israel

³Department of Orthopaedic Surgery, Assaf Harofeh Medical Center, Israel.

*Author for correspondence:

vmadan@montefiore.org

utility of PROMs data, unfamiliarity with the interpretation of PROMs information, a paucity of direct face-to-face interaction, cost of data collection, and the need for rapid data manipulation and processing [13-15]. Moreover, since most clinics are usually capacity-driven, adding PROMs (WOMAC and SF-36) into a standard care routine will extend a regular session by 20-30 minutes, which can be a significant barrier for adoption. Therefore, it is clear that there is a need for the creation of PROMs that are primarily designed for clinical practice rather than research (i.e., brief, simple, and easy to interpret). Ideally, these will cover general and disease-specific properties and will apply to a range of common MSK conditions [14]. New tools are emerging to create PROMs that will fit a real-live clinical practice work flow [14,16].

One approach to creating a clinical PROM is to adopt a subset of the larger, scientifically validated PROM in a patient care setting. The purpose of the current work is to analyze the standard PROMs in knee OA that assess symptoms of pain and functional limitations as well as the general quality of life (WOMAC and SF-36), and determine which questions out of the 60 are the most reflective of the overall score and show sensitivity to changes in clinical status.

Methods

This was a retrospective study based on a dataset that belongs to a private medical device company (Apos Medicafl Assets Ltd. AMA, Tel-Aviv, Israel). The company provides non-invasive biomechanical treatment for patients with MSK conditions in Israel, UK and USA. PROMs are an integral part of the company's treatment methodology, hence a large dataset was available for analysis. The majority of patients have knee and back arthritic complaints. The protocol was approved by the Institutional Helsinki Committee Registry (Helsinki registration number 141/08, NIH protocol no. NCT00767780). A search for eligible data was done on patients that were treated between October 2010 and June 2017. All patients with a primary knee condition and PROMs at pre-treatment initiation assessment and after three months of treatment were included in the analysis.

The WOMAC questionnaire (disease-specific questionnaire) and SF-36 (general health questionnaire) were used as PROMs to evaluate

pain, functional limitation, and quality of life perception. The WOMAC questionnaire contains 24 Visual Analogue Scale (VAS) questions. Results range from 0 mm-100 mm, in which 0 mm indicates no pain or limitation in function, and 100 mm indicates the most severe pain or limitation in function. The SF-36 contains 36 questions, seven yes/no questions, ten 3-point Likert scale questions, nine 5-point Likert scale questions, and ten 6-point Likert scale questions indicating quality of life. Questions are scored between 0-100, with 0 indicating the worst quality of life and 100 indicating the best quality of life.

All patients were treated with a non-invasive biomechanical foot-worn device that aims to treat patients with MSK conditions by center of pressure manipulation and perturbation training to challenge and train neuromuscular control [17].

Statistical analysis

The purpose of the current study was to identify and categorize the most influential questions that determine the overall score in each of the questionnaires. For this purpose we have divided our statistical analysis into two stages:

1. Calculate the reliability of WOMAC and SF-36 using the Cronbach's Alpha (α) test. Cronbach's Alpha measures how consistently participants respond to one set of items. As a sort of average of the correlations between items. Cronbach's Alpha ranges from 0.0 to 1.0, where 1.0 indicates a perfect correlation and 0.0 suggests no correlation. For SF-36, we excluded all questions with yes/no response due to low sensitivity (Q13-Q19 all had a low Alpha-Cronbach score) and assessed the reliability of the SF-36 twice—first using all items that have at least three possible answers (3-level questions and more) and second using questions with at least five possible answers (5-level questions and more).

2. A decision tree classifier supported with a linear mix model regression. A decision tree approach was selected as it is commonly used in data mining to create models that predict a target value based on several independent variables. In continue to stepwise variable selection in regression analysis, the decision tree method was used to focus on variables selection that should be used to form decision tree models.

A Chi-Squared Automatic Interaction Detection

(CHAID), an important inducer classification tree algorithm, allowed us to assess the relative importance of each question. CHAID uses F test for the continuous attributes, Pearson chi-square test for nominal and likelihood ratio test for the ordinal attributes. Generally, variable importance is computed based on a reduction in the model accuracy (or in the purities of nodes in the tree) when the variable is removed. In most circumstances the more records a variable influences the greater the importance of the variable. The accuracy of the model was calculated from the tree nodes and is a function of sensitivity and specificity. While specificity measures how well the classifier can recognize negative samples, sensitivity assesses how well the classifier can identify positive samples.

Finally, the decision tree was also used for prediction. Since the tree model is derived from historical data, it's easy to predict the result for future records.

Statistical analyses was performed using IBM SPSS Statistics for Windows, version 26.0 (IBM Corp., Armonk, NY, USA) and IBM Modeler. The primary outcome of the study was WOMAC and SF-36 overall scores. Two-sided Pearson's chi-square tests were used to compare categorical data. The normality of continuous data was examined using the Kolmogorov-Smirnov test values are presented as mean \pm standard deviation p-values <0.05 were defined as statistically significant.

Results

Four thousand nine hundred eighty-three (4,983) patients had WOMAC and SF-36 at two-time points, pre-treatment and after three months of treatment. 58% of the patients were females and the mean (SD) age was 62.8 (9.9). The average (SD) pain and functional disability levels at baseline were 39.2 (21.9) and 35.2 (22.5), respectively, where 0 indicates no pain and 100 indicates worse pain. In addition, the average Physical and Mental quality of life, derivatives of the SF-36 questionnaire were 49.1 (19.8) and 67.2 (19.7), respectively, where 0 indicates poor QoL and 100 indicates the best QoL. 87% of the patients had some form of knee OA, 10% had some form of knee injury (meniscal tear, ligament tear, ligament reconstruction), 1% had some form of dislocation/fracture (patella dislocation, patella fracture, tibial plateau fracture), 1% had spontaneous osteonecrosis of

the knee, 1% were post total knee replacement and $<1\%$ had other knee condition.

■ SF-36 and WOMAC reliability

The reliability of SF-36 3-level questions and more was 0.875 and included the following six items: Q7, Q9, Q10, Q11, Q5, and Q6. The reliability of SF-36 5-level questions and more was 0.868 and included the following 14 items: Q20-Q26, Q28-Q32, Q34, Q36. The reliability of WOMAC was 0.973 and included all 24 items.

■ A decision tree classifier supported with a linear mix model regression

In general, the SF-36 and WOMAC dependant variables of the decision tree were the SF-36 overall score and WOMAC overall score, respectively. For SF-36 3-level questions, the most influential predictors were Q11 (36%), Q4 (26%), Q2 (15%) and Q1 (14%). For SF-36 5-level questions and more, the most influential predictors were Q22 (47%), Q25 (35%), Q32 (10%) and Q28 (5%). We then ran a decision tree on the following items (integration of the most influential of both trees): Q1-Q2, Q4, Q6, Q11, Q22, Q24, Q25, Q28, Q32 and found that the most influential items were Q22 (39%), Q32 (24%), Q11 (19%), Q25 (19%). For WOMAC, the most influential predictors were Q14 (39%), Q10 (24%) and Q15 (21%). **TABLE 1** summarizes the main predictive questions to be used.

A significant improvement in WOMAC and SF-36 was seen after three months of treatment ($p<0.01$). WOMAC overall score improved by 15% from 31.3 (27.1) to 26.6 (25.3). SF-36 overall score improved by 5% from 59.7 (31.6) to 62.4 (30.1). For SF-36, the main predictor items were Q11, Q22 and Q32, Regression model $R^2=0.841$, $p<0.01$, $t[55.62]=0.001$, Beta for Q22=0.409, Q32=0.352, Q11=0.278. For WOMAC, the main predictor items were Q10 and Q15, Regression model $R^2=0.930$, $p<0.01$, $t[35.4]=0.001$, Beta for Q15=0.548, Q10=0.4639 **TABLE 2 and TABLE 3. TABLE 4** summarizes the the changes in WOMAC and SF-36 over time. In summary, for SF-36 using the above mentioned 4 questions will cover 40% of the overall score. For WOMAC questionnaire, using Q10 and Q15 will cover 50% of the total score).

Discussion

Our results showed that the use of WOMAC and SF-36 to assess MSK conditions and treatment effect is reliable, similar to previous recommendations [18,19]. WOMAC and SF-36 measure accurately the patient's condition (pain, function, and quality of life). Moreover, we found some items to be more influential than others and were able to identify six questions instead of 60. Two questions from the WOMAC questionnaire predicts 93% of the overall score and four questions from the SF-36 predict 84%. In clinical practice having six items instead of 60 is far more manageable and can be transformational with regards to PROMs integration and implementation.

This study tries to address and overcome the challenges and lack of adoption of PROMs in real-life clinical practice [2,10,20]. Although previous studies have discussed the challenges in implementing PROMS in clinical practice, to the best of our knowledge, there was no attempt to adjust existing research-based PROMs to real-life settings, which is fundamentally different than in research. We believe that instead of trying to implement PROMs in their current format (i.e., long, time-consuming, difficult to interpret) into the clinic, we should try to adjust PROMs to fit a typical real-life clinical practice work-flow by balancing difficulty in administration with clinical utility. Adjusting PROMs to a shorter version can address concerns of capacity intensity (i.e., extending a session by 20-30 min.), additional costs of data collection, and the need for rapid data manipulation and processing and allow the clinic to become a data-driven, evidence-based, best practice clinical setting. Additionally, using the subset of questions that are validated will allow for more rapid development of specific clinical instruments for clinical use. A strength of the study is that we were able to use thousands of records of patients that completed two gold-standard PROMs (WOMAC and SF-36) and had a known clinical benefit that we could compare our extracted subset of question. Using this method, we identified 6 out of 60 questions as the most influential and predictive items. We believe that using this subset of 6 questions, PROM completion can become a straightforward and practical task for both the patient and the clinic. This is a new approach to the problem that prior studies have demonstrated regarding the lack of guidance

and clarity as to what to measure, which tools to use, and how to efficiently apply this in routine clinical practice [14,16,18].

The results of the study suggested that the six items are in accordance with the predictive items that were found in the regression analysis and correlate to the clinical improvement over time. This is important as it addresses the responsiveness requirements i.e., does the PROM detect change over time that matters to patients (sensitivity to change) [18]. It gives additional credibility to the use of a short form in clinical practice. In the unique setting of a busy clinic, using six questions can significantly reduce the burden for the patient and the clinic staff and facilitate adoption. That being said, more research is needed in order to validate the proposed short form. Ideally, this should be done as an on-going registry program aimed to monitor real-life clinical practice patients with a varied patients population.

This study has some limitations that should be acknowledged. First, some patients' characteristics are missing. Although all patients were with a primary knee condition, the diagnosis is missing. In addition, weight and height are missing. This might limit the ability to generalize the results and we recommend that future studies will validate the outcomes of the study on different ethnicities and populations with varying weight distribution. Secondly, this study proposes a novel 6-item questionnaire that was established from a subset of 60 gold-standard questions by identification of the most influential ones. This new questionnaire, however, is currently not being used elsewhere and requires further validation. Lastly, future studies should also compare the correlation between the 6 item questionnaire and objective outcomes such as computerized gait test, other validated questionnaires, so support its validity.

Conclusion

This study demonstrated that two questions from the WOMAC questionnaire predict 93% of the overall score and four questions from the SF-36 predict 84%. A six questions subset from a total of 60 questions in the WOMAC and SF-36 QOL scales could yield over 50% of the sensitivity of the full surveys at a fraction of the overall burden of time and effort. This potentially allows for the addition of PROM

to clinical practice and is in line with previous studies that have stressed the importance of PROMs selection standardization [18] rather than adding new tools. Our real-life experience in implementing the current available PROMs in clinical practice leads us to think that the creative use of a new questionnaire out of existing PROMs may help patient care in busy clinical settings. Future work should focus on validation and extension of the tool in clinical practice.

Declarations

■ Ethics approval and consent to participate

The protocol was approved by the Institutional Helsinki Committee Registry (Helsinki registration number 141/08, NIH protocol no. NCT00767780).

Consent for publication

N/A

Availability of data and materials

All data requests will be reviewed and addressed by the authors.

Competing interests

Non to declare.

Funding

This study was not funded in any way.

Authors' contributions

All authors take full responsibility for the entire manuscript content, integrity of the data and the accuracy of the data analysis. Study Concept and Design: Vikram Madan, MNB. Acquisition of data: Vikram Madan. Analysis and Interpretation of Data: Vikram Madan, Avi Elbaz, Amit Mor, Yiftah Beer, Matthew N. Bartels. Drafting of the Manuscript: Vikram Madan. Critical Revision of the Manuscript for Important Intellectual Content: Vikram Madan, Avi Elbaz, Amit Mor, Yiftah Beer, Matthew N. Bartels.

Acknowledgements

The authors would like to thank Dr. Ornit Cohen for her statistical analysis support.

Conflict of interests

None to declare.

References

- Weldring T, Smith SM. Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). *Health Serv Insights*. 6: 61-68 (2013).
- Valderas JM, Kotzeva A, Espallargues M, et al. The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. *Qual Life Res*. 17: 179-193 (2008).
- Black N. Patient reported outcome measures could help transform healthcare. *BMJ*. 346: f167 (2013).
- Porter I, Goncalves-Bradley D, Ricci-Cabello I, et al. Framework and guidance for implementing patient-reported outcomes in clinical practice: evidence, challenges and opportunities. *J Comp Eff Res*. 5: 507-519 (2016).
- Boyce MB, Browne JP, Greenhalgh J. The experiences of professionals with using information from patient-reported outcome measures to improve the quality of healthcare: a systematic review of qualitative research. *BMJ Qual Saf*. 23: 508-518 (2014).
- Van Der Wees PJ, Nijhuis-Van Der Sanden MW, Ayanian JZ, et al. Integrating the use of patient-reported outcomes for both clinical practice and performance measurement: views of experts from 3 countries. *Milbank Q*. 92: 754-775 (2014).
- Bozic KJ, Belkora J, Chan V, et al. Shared decision making in patients with osteoarthritis of the hip and knee: results of a randomized controlled trial. *J Bone Joint Surg Am*. 95: 1633-1639 (2013).
- Bijlsma JW, Berenbaum F, Lafeber FP. Osteoarthritis: an update with relevance for clinical practice. *Lancet*. 377: 2115-2126 (2011).
- Altman R, Asch E, Bloch D, et al. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and therapeutic criteria committee of the american rheumatism association. *Arthritis Rheum*. 29: 1039-1049 (1986).
- Sorensen NL, Hammeken LH, Thomsen JL, et al. Implementing patient-reported outcomes in clinical decision-making within knee and hip osteoarthritis: an explorative review. *BMC Musculoskelet Disord*. 20: 230 (2019).
- Fennelly O, Blake C, Desmeules F, Stokes D, Cunningham C. Patient-reported outcome measures in advanced musculoskeletal physiotherapy practice: a systematic review. *Musculoskeletal Care*. 16: 188-208 (2018).
- Lohr KN, Zebrack BJ. Using patient-reported outcomes in clinical practice: challenges and opportunities. *Qual Life Res*. 18: 99-107 (2009).
- Valderas JM, Alonso J, Guyatt GH. Measuring patient-reported outcomes: moving from clinical trials into clinical practice. *Med J Aust*. 189: 93-94 (2008).
- Hill JC, Thomas E, Hill S, et al. Development and validation of the keele Musculoskeletal Patient Reported Outcome Measure (MSK-PROM). *PLoS One*. 10: e0124557 (2015).
- Kamaleri Y, Natvig B, Ihlebaek CM, et al. Localized or widespread musculoskeletal pain: does it matter? *Pain*. 138: 41-46 (2008).
- Zimmermann C. Ultra-short PROMs: clever or not? *Br J Cancer*. 103: 1477-1478 (2010).
- Reichenbach S, Felson DT, Hincapié CA, et al. Effect of biomechanical footwear on knee pain in people with knee osteoarthritis: The biotok randomized clinical trial. *JAMA*. 323: 1802-1812 (2020).
- Haywood KL. Patient-reported outcome II: selecting appropriate measures for musculoskeletal care. *Musculoskeletal Care*. 5: 72-90 (2007).
- Kyte DG, Calvert M, van der Wees PJ, et al. An introduction to patient-reported outcome measures (PROMs) in physiotherapy. *Physiotherapy*. 101: 119-125 (2015).
- Dawson J, Doll H, Fitzpatrick R, et al. The routine use of patient reported outcome measures in healthcare settings. *BMJ*. 340: c186 (2010).