BMJ Open

Study protocol for the validation of a new pictorial Functional Scale in patients with knee osteoarthritis: the Functional Activity Scoring Tool (FAST)

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2023-076947
Article Type:	Protocol
Date Submitted by the Author:	21-Jun-2023
Complete List of Authors:	Tang, Zhi Yin; SingHealth Polyclinics, Department of Physiotherapy Ng, Khim Siong; SingHealth Polyclinics, Department of Physiotherapy Yeung, Meredith T.; Singapore Institute of Technology, Health and Social Sciences Cluster
Keywords:	Patient Reported Outcome Measures, Patient-Centered Care, REHABILITATION MEDICINE, Surveys and Questionnaires

Long Title

Study protocol for the validation of a new pictorial Functional Scale in patients with knee osteoarthritis: the Functional Activity Scoring Tool (FAST)

Short Title

Developing a pictorial FAST scale: study protocol (49 letters with space)

Trial registration number:

NCT05590663 (ClinicalTrials.gov)

Protocol version

16-May-23

Funding:

This work is supported by the SingHealth Polyclinics Research Seed Fund (SHPCG-RSF22-01).

Zhi Yin Tang[1], Khim Siong Ng[1], Meredith T. Yeung[2]

[1]Department of Physiotherapy, Singhealth Polyclinics, Singapore

[2] Health and Social Sciences Cluster, Singapore Institute of Technology, Singapore

Zhi Yin Tang ORCID: 0000-0002-8641-8146

Khim Siong Ng ORCID: 0000-0002-0374-326X

Meredith Yeung ORCID: 0000-0002-4294-4187

ZYT and KSN drafted the manuscript; MTY critically revised the manuscript.

All authors approved the final version of the article.

Conflict of interest:

All authors declare that no conflicts of financial or non-financial competing interests are associated with the current study.

Acknowledgements:

The authors thank and acknowledge SingHealth Polyclinics and Singapore Institute of Technology for supporting this study.

ABSTRACT

Background

Patient-reported outcome measures (PROM) are required for patient-centred care. There are limited PROM with good psychometric properties, and limitations to any language-based scale are often constrained by the written words or numerals used. Therefore, we developed the Functional Activity Scoring tool (FAST), a self-reporting pictorial scale. The FAST measures the impact of knee osteoarthritis on essential activities of daily living (ADL) and the significant changes in the self-perceived functional status over time.

Objectives:

This study aims to: (1) develop the FAST with adaptation from the Wong-Baker FACES pain rating scale; (2) validate the FAST against the Patient-Specific Functional Scale (PSFS) and Knee Injury and Osteoarthritis Outcome Score (KOOS); (3) establish the reliability and validity of FAST in individuals with knee osteoarthritis.

Methods and Analysis

The prospective study protocol investigates the face, content and criterion validity, as well as the intra-, inter-rater, and test-retest reliability. The PSFS and KOOS will be gold standard comparisons. Participant recruitment will occur at four public polyclinics that offer physiotherapy outpatient services in Singapore. Onsite physiotherapists familiar with the study eligibilities will refer potential participants to the investigators after the routine physiotherapy assessment. After providing written consent, eligible participants will complete outcome measurements with the FAST, PSFS and KOOS during baseline and follow-up assessments. The Global Rating of Change (GROC) scale will determine the perceived change in knee osteoarthritis.

Ethics and dissemination

SingHealth-Centralised Institutional Review Board approved the study (CIRB reference number: 2022/2602). The final results will be published via scientific publication. The FAST will benefit the evaluation and management of those who suffer knee osteoarthritis regardless of English proficiency or language barriers.

(269 words)

STRENGTHS AND LIMITATIONS OF THIS STUDY

- The first pictorial patient-specific functional assessment tool - the Functional Activity Scoring Tool (FAST), is developed.
- This multi-site study will validate the novel pictorial scale created and reviewed by an expert panel comprising patients and their families, physiotherapists and family physicians.
- The proposed study aligns with international consensus standards on best practices of instrument development and validation studies—the COnsensus-based Standards for selecting health-status Measurement INstruments (COSMIN).
- Validation of patient-reported outcome measures (PROM) is an iterative process. More testing of its psychometric properties must follow to support its usefulness in patients with other musculoskeletal conditions.
- Although the study protocol will not alter the standardised physiotherapy treatment, we cannot rule out possible confounding variables that may influence the study outcomes.

INTRODUCTION

Healthcare professionals regularly assess the crucial yet trouble-functioning tasks in activities of daily living (ADL). While various condition-specific questionnaires, such as the Roland-Morris Disability Questionnaire (RMQ)[1] or Knee Injury and Osteoarthritis Outcome Score (KOOS)[2], and health status measures, such as the 36-Item Short Form Survey (SF-36)[3] or EuroQol-5D (EQ-5D)[4] exist, unfortunately, limited patient-reported outcome measures (PROM) have been established thus far, especially in the area of osteoarthritis. The application of PROM in orthopaedic is expected to increase.[5] PROM were initially created for research purposes and eventually adopted for clinical management, seeking to determine patients' perceptions of their symptoms, functional status, and health-related quality of life. PROM are frequently unfittingly referred to as "outcome measures," even though they measure health—by comparing a patient's health at different times, the care outcome received can be determined.[6] PROM provide additional 'patient-centred' data that is unique in capturing the patient's perspective on the impact of their disease or disorder and its treatment.[5] These self-reported instruments elicit information about a patient's health status directly from the patient without needing interpretation from a healthcare professional.[6] The approach of gathering patient-centred data is integral in informing clinical care and supplementing measurable clinical improvements in the patients as part of the routine practice. Well-validated PROM assessing functional outcomes is required in the era of patient-centred care for holistic management.

Few osteoarthritis-specific PROM have been developed and extensively studied. A systematic review[7] identified these PROM attempt to measure psychometric properties such as pain, mental functions and moods, physical symptoms such as stiffness and mobility, as well as function in sports and recreation with either the term or subscale level. Overall, the review findings found limited evidence of psychometric properties from these PROM. Concurrently, straightforward tools to report on self-efficacy were limited. Among these, the Patient-Specific Functional Scale (PSFS)[8] was uniquely developed to enable self-reporting of the impact of musculoskeletal conditions on essential ADL and the significant

changes in the self-perceived functional status over time. A recent systematic review[9] of the PSFS against the Consensus-based Standards for the selection of health-status Measurement Instruments (COSMIN) criteria concluded that it is an easy-to-use, reliable and responsive scale in numerous musculoskeletal conditions. The PSFS is applicable across various conditions and body regions as it allows the comparison of functional outcomes across conditions and between studies.[10-12] However, the usefulness of a language-based PROM is restricted if patients and families have limited English proficiency, which is a barrier to healthcare services that is well documented.[13] Singapore is primarily an English-speaking country. However, her geographical location, historical and cultural backgrounds greatly influence the languages used in this city-state. Our anecdotal experience suggests that older patients (especially individuals ≥ 65 years old) have difficulty understanding and accurately completing the PSFS, as English may not necessarily be their primary spoken language. Difficulty with completing forms can occur for many reasons, such as the written words and numerals not being universally understood, and problems with health literacy are common and underestimated.[14-16] Very often, informal interpretation, such as relying on a family member to translate the communication, has shown to be associated with a more significant number of errors.[17] Pictures or pictorial aids are a useful adjunct to medical information and aid the transfer and comprehension of written and spoken information.[18] One good example is using the pictorial scale to measure pain in the Wong-Baker FACES pain rating scale,[19] with proven benefits such as improving adherence to medications and enhancing the understanding of instructions.[20]

As such, there is a growing need for a new PROM that is simple, reliable and responsive, yet minimises the limitations of any word or language-based outcome measure that are currently in use. The Functional Activity Scoring Tool (FAST) has been developed to address this situation. The FAST is a new pictorial-scale PROM measuring function in an individual with osteoarthritis. Several aspects were considered during the conceptualisation of the instrument: the applicability to a broad range of clinical presentations (conditions, limitations and age); simple administration; concise yet effectual for speedy medical documentation; and simple interface in electronic medical record systems.

HYPOTHESIS and AIM

The confidence of a PROM depends on the psychometric evaluation of its measurement properties, and it must be undertaken to satisfy rigorous criteria.[21] These include validity (to what extent does the instrument measure the construct it purports to measure), reliability (the degree to which measurement is free from error) and responsiveness (the ability of an outcome measure to detect change over time in the construct to be measured).[22] The process to assess these measurement properties must be iterative and studied individually. Thus, we hypothesise that the measure of function, an additional dimension to the quality of life, is possible by the same principle. The new FAST scale can be used to measure function and difficulty in performing ADL in patients with osteoarthritis, in an equally valid and reliable manner as the PSFS and KOOS. We aim to provide a standardised tool for gathering and documenting patients' symptoms. With these considerations, we developed the FAST scale. This study aims to: (1) develop the FAST pictorial functional scale with adaptation from the Wong-Baker FACES pain rating scale; (2) validate the FAST against the PSFS and KOOS; (3) establish the reliability and validity of FAST in individuals with knee osteoarthritis.

METHOD AND ANALYSIS

Study design and setting

This study will be a prospective validation study to establish the psychometric properties of a newly-developed PROM. This study is proposed under the recommendation of Basch et al. (2015) [23] methods for developing patient-reported outcome-based performance measures and uses the procedures that De Vet and colleagues [24] advocated for in developing a PROM. This approach provides evidence for developing a PROM that measures the intended context and its use as an outcome measure in clinical practice and research trials. The study will take place in four physiotherapy outpatient clinics in Singapore over 12 months.

Patient and public involvement

Patients and families from physiotherapy outpatient clinics provided input and suggestions to the FAST scale during its conceptualisation and feasibility stage. Hence, their feedback also shaped the scale design, with the pros and cons of the different versions of the FAST scale discussed with patients and/or their families who will not be recruited as study participants.

Development of the FAST

During the feasibility stage, surveys were conducted on patients, families and healthcare professionals to gather feedback on the application of the PSFS. The most prevalent verbatim was "difficulty comprehending PSFS due to its being too lengthy and the lack of pictorial aid to assist patient's comprehension of the scale." Therefore, a prototype of the FAST scale was created and reviewed by an expert panel of academics, researchers and clinicians (n=7) and a series of cognitive interviews with a purposive sample of patients older than 65 (n=12) to elicit feedback on its relevance, clarity and acceptability. The final version of FAST was developed after three revisions. Figure 1 presents the conceptualisation and revision process of the FAST development. The final version of the FAST scale from this revision process will be used to test for reliability and validity in this study protocol. It consists

of a pictorial diagram with seven expression faces corresponding to an 11-point Likert scale, with Face 1 (the saddest expression) on the left of the scale paired with a score "0" and a verbal description of "unable to perform", Face 4 (neutral expression) to agree with score "5" with a description of "moderately difficult" and Face 7 (the happiest expression) on the right of the scale matching with score "10" and a descriptor of "able to perform like before". The red "cross" on the left and green "tick" on the right accentuate the effects of the facial expression and association with the verbal descriptors.

(Insert figure 1 here)

Sample size

The size of the retest sample was estimated based on a method developed to calculate the required number of participants in a reliability study.[25] The probability of type I and type II error were $\alpha = 0.05$ and $\beta = 0.20$, respectively. An interclass correlation coefficient (ICC) value of less than 0.50 indicated poor reliability, whereas values between 0.50 and 0.75 indicated fair to good reliability; an ICC value greater than 0.9 showed excellent reliability.[26] We hypothesised that our findings would be consistent with a minimum coefficient of 0.75. This level of reliability is at least appropriate for person-level comparisons. Following these assumptions, a minimum of 50 participants will be necessary for the test-retest analysis for this study. According to COSMIN guidelines, validity calculations are considered good-excellent if the sample size exceeds 100 (n=100).[27] To allow for a possible attrition rate of 20%, a minimum sample size of 120 will be needed.

Participant recruitment and selection criteria

Participant recruitment will occur at four public polyclinics that offer physiotherapy outpatient services in different districts of Singapore. Onsite physiotherapists familiar with the study protocol will identify eligible participants during the routine initial physiotherapy assessment. Inclusion criteria will be: individuals diagnosed with knee osteoarthritis and referred for physiotherapy care at the polyclinics; age 45 years and above; and proficient in colloquial/conversational English. Potential participants will be

excluded if there are additional underlying medical or trauma condition(s) of the knees (e.g., trauma, fracture, infection, inflammatory disease, tumour), history of knee surgery within the last three months, or clinically recognisable cognitive impairment that inhibits the comprehension and completion of the questionnaires. Participation in the study is strictly voluntary and will not impact the type or quality of the individual's physiotherapy treatments based on prevailing evidence.

Instruments

The self-administered KOOS is a knee-specific instrument developed to assess the patients' opinions about their knees and associated short and long-term problems.[2] It is a validated tool in Singapore for knee osteoarthritis patients.[28] It consists of 42 items in 5 subscales, i.e., pain (9 questions), symptoms (7 questions), activities in daily living (17 questions), sport and recreation function (4 questions), and knee-related quality of life (4 questions). The 5-point Likert scale scoring system ranges from “0” (no problems) to “2” (moderate problem) to “4” (Extreme problem), and the score for each domain is calculated by summing the questions. Scores will be converted to a 0 to 100 scale, with zero representing extreme knee problems and 100 representing no knee problems. The use of the 0 to 100 score is practical as it projects a direct reference to the percentage concept. [2]

The PSFS is a self-reported, patient-specific measure that assesses patients' functional status.[8] Patients are asked to identify three activities most affected by their conditions and then rate their ability on an 11-point Likert 0 to 10 scale for each activity, where “0” is unable to perform the activity, and “10” being able to perform the activity at the same level as before the onset of symptoms. The total score is computed by dividing the sum of the activity scores by the number of activities listed.

The global rating of change scale (GROC) is an outcome measure that assesses patients' self-perception of change in their condition between sessions.[29] The GROC is quantified on a 15-point Likert scale from "-7” (a very great deal worse) to “0” (about the same) to “7” (a very great deal better). The

scale is easy to administer as it requires minimal skills or training, has good reproducibility and is sensitive to changes. [30]

Procedure

Eligible individuals will be informed of the study's purpose and data collection procedures. Written informed consent will be obtained from every participant before data collection commences. Participants' confidentiality and anonymity will be maintained throughout the study process with a unique identifier, and only the study researchers will have access to the data. Participants will receive standardised care, and their participation status will not be shared with the attending physiotherapists apart from the initial identifications for eligibility. All data collection forms will be coded with the same unique identifier, and the study team will not retain any identifiable information. Only anonymised data will be used for data analysis. The project investigators will perform all data collection. Demographic data, clinical characteristics and primary outcome measurements with the FAST, PSFS and KOOS will be collected during baseline assessment (week 0). Follow-up assessment with FAST, PSFS, and KOOS will be scheduled two to three weeks post initial assessment together with the administration of the GROC to evaluate the efficacy of the standard physiotherapy treatment that the participants will be receiving regardless of the participation status in this study. Figure 2 depicts the workflow of the data collection procedures. This study will not require any alteration or deviation from the standard protocol for knee osteoarthritis physiotherapy management.

(Insert figure 2 here)

Validation

Face validity

The qualitative methods used to determine the face validity of FAST involved face-to-face meetings with an expert panel of academics, researchers and clinicians (n=7), and a series of cognitive interviews with patients (n=12). Three essential criteria were determined in establishing the face validity:

Clarity (the extent to which an item is open to more than one possible interpretation), Relevancy (the extent to which an item will be relevant to its component), and Acceptability (the extent to which readers would easily understand an item).

Content validity

The content validity index (CVI) and context validity ratio (CVR) will determine the content validity. [31] CVI is the most widely reported method for determining content validity in instrument development, assessing its relevance and clarity. There are two methods of calculation, namely item-CVI (I-CVI) and scale level-CVI (S-CVI) [32]. This study will use a 4-point Likert scale "1" = unacceptable, "2" = needs some revision, "3" = needs minor revision and "4" = acceptable, for the calculation of the I-CVI from the total rating scores from all panel members. Where I-CVI is greater than 0.79, the item is acceptable; between 0.70 to 0.79, the item will require revision; and when it is less than 0.70, the item will be eliminated. [33] Similarly, the S-CVI will be determined by the number of items in an instrument that receives a "highly acceptable" grade. The Universal Agreement (UA) among the panel members (S-CVI/UA) and the Average CVI (S-CVI/Ave) are two ways of determining S-CVI. [32] S-CVI/UA will be calculated by the sum of all items with I-CVI equal to 1 divided by the total number of items, and S-CVI/Ave is equal to the sum of all the I-CVI divided by the number of items. Content validity is excellent when the S-CVI/UA is more than 0.8 and the S-CVI/Ave is more than 0.90. [33]

CVR quantify the essentiality of an item. [34] CVR ranges from -1 to 1; a higher score represents a greater agreement between panel members. CVR = (Ne – N/2)/(N/2), where Ne is the number of panel members who rated an item as "essential", and N is the total number of panel members. [32] Each element of the FAST scale will be evaluated on a 3-point Likert scale (1 = not essential, 2 = useful but not essential, 3 = essential).

Criterion Validity

The KOOS Singapore English version and PSFS will serve as the criterion for disability in the knee osteoarthritis population. The two validated self-administered questionnaires are specific and sensitive to change over time. The correlations between the FAST, KOOS and PSFS will assess the criterion validity of the FAST scale.

Reliability

A pilot study will be in place to establish intra-rater and inter-rater reliability among the researchers. Inter-rater and intra-rater reliabilities of FAST will be analysed via intraclass correlation coefficients (ICCs) using average measures of two-way mixed effect and two-way random effect models, respectively, and Bland–Altman plots. The test-retest reliability of FAST will be calculated via ICC for absolute agreement using a two-way mixed-effect analysis of the variance model between the scores of two stable assessment periods (i.e. global rating of change less than 3).

Statistical analysis

All statistical analysis will be conducted using SPSS 26 for Windows with the statistical significance set as $p < 0.05$. Descriptive statistics will be used to describe the demographic variables and their corresponding FAST score using the Mann-Whitney U test or Spearman's correlation. Spearman correlation will investigate the criterion validity against PSFS, KOOS and GROC and the measurement of agreement according to the following criteria: high (rho ≥ 0.60); moderate (rho < 0.60 - ≥ 0.30); or low (rho <0.30). The higher the rho, the higher the agreement between the two instruments. The Cronbach alpha will assess the internal consistency. The intraclass correlation coefficient (ICC) will be used to analyse test-retest reliability. An ICC value of less than 0.5 will indicate poor reliability, whereas values between 0.5 and 0.75 will indicate fair to good reliability; an ICC value greater than 0.9 will show excellent reliability. Each facial expression will be compared with the demographics to determine if there are differences in pain associated with characteristics.

Ethics and dissemination

The SingHealth Centralised Institutional Review Board (CIRB) approved this research protocol: (CIRB reference number: 2022/2602). There are no potential risks for participants taking part in this study. All participants will provide written consent to participate and have the right to withdraw from participation in the project at any time without any compromise or disadvantage to them in any form. All participants will be assigned a unique de-identified code to protect the confidentiality of the participants. Access to the data is restricted to the project investigators, and only anonymised data will be used during data analysis. All investigators declare no financial or other competing interests at all study sites. This study will validate the new pictorial functional scale (FAST) in patients with knee osteoarthritis and hope to investigate if the new scale correlates with similar existing PROM with good validity and reliability. The final results and establishment of the new PROM will be published via scientific publication. This will be advantageous to healthcare professionals in evaluating functional status changes in individuals with osteoarthritis regardless of English proficiency or language barriers.

Trial status

The study is at its pilot trial stage at the time of submission of this study protocol.

(2991 words)

References

1. Roland M, Morris R. A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. *Spine (Phila Pa 1976)* 1983;8(2):141-4. doi: 10.1097/00007632-198303000-00004

2. Roos EM, Roos HP, Lohmander LS, et al. Knee Injury and Osteoarthritis Outcome Score (KOOS)—Development of a Self-Administered Outcome Measure. *Journal of Orthopaedic & Sports Physical Therapy* 1998;28(2):88-96. doi: 10.2519/jospt.1998.28.2.88

3. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30(6):473-83.

4. EuroQol-Group. EuroQol--a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16(3):199-208. doi: 10.1016/0168-8510(90)90421-9

5. Kyte DG, Calvert M, van der Wees PJ, et al. An introduction to patient-reported outcome measures (PROMs) in physiotherapy. *Physiotherapy* 2015;101(2):119-25. doi: 10.1016/j.physio.2014.11.003 [published Online First: 20141128]

6. Black N. Patient reported outcome measures could help transform healthcare. *Bmj* 2013;346:f167. doi: 10.1136/bmj.f167 [published Online First: 20130128]

7. Lundgren-Nilsson Å, Dencker A, Palstam A, et al. Patient-reported outcome measures in osteoarthritis: a systematic search and review of their use and psychometric properties. *RMD Open* 2018;4(2):e000715. doi: 10.1136/rmdopen-2018-000715

8. Stratford P, Gill C, Westaway M, et al. Assessing Disability and Change on Individual Patients: A Report of a Patient Specific Measure. *Physiotherapy Canada* 1995;47(4):258-63. doi: 10.3138/ptc.47.4.258

9. Pathak A, Wilson R, Sharma S, et al. Measurement Properties of the Patient-Specific Functional Scale and Its Current Uses: An Updated Systematic Review of 57 Studies Using COSMIN Guidelines. *Journal of Orthopaedic & Sports Physical Therapy* 2022;52(5):262-75. doi: 10.2519/jospt.2022.10727

10. Chatman AB, Hyams SP, Neel JM, et al. The Patient-Specific Functional Scale: Measurement Properties in Patients With Knee Dysfunction. *Physical Therapy* 1997;77(8):820-29. doi: 10.1093/ptj/77.8.820

11. Hefford C, Abbott JH, Arnold R, et al. The Patient-Specific Functional Scale: Validity, Reliability, and Responsiveness in Patients With Upper Extremity Musculoskeletal Problems. *Journal of Orthopaedic & Sports Physical Therapy* 2012;42(2):56-65. doi: 10.2519/jospt.2012.3953

12. Westaway MD, Stratford PW, Binkley JM. The Patient-Specific Functional Scale: Validation of Its Use in Persons With Neck Dysfunction. *Journal of Orthopaedic & Sports Physical Therapy* 1998;27(5):331-38. doi: 10.2519/jospt.1998.27.5.331

13. Espinoza J, Derrington S. How Should Clinicians Respond to Language Barriers That Exacerbate Health Inequity? *AMA Journal of Ethics* 2021;23(2):E109-16. doi: 10.1001/amajethics.2021.109

14. Williams MV, Parker RM, Baker DW, et al. Inadequate Functional Health Literacy Among Patients at Two Public Hospitals. *JAMA* 1995;274(21):1677-82. doi: 10.1001/jama.1995.03530210031026

15. Baker DW, Gazmararian JA, Sudano J, et al. The Association Between Age and Health Literacy Among Elderly Persons. *The Journals of Gerontology: Series B* 2000;55(6):S368-S74. doi: 10.1093/geronb/55.6.S368

16. Williams MV, Davis T, Parker RM, et al. The role of health literacy in patient-physician communication. *Fam Med* 2002;34(5):383-9.

17. Whitaker KL, Krystallidou D, Williams ED, et al. Addressing language as a barrier to healthcare access and quality. *British Journal of General Practice* 2022;72(714):4-5. doi: 10.3399/bjgp22x718013

18. Ghiassi R, Murphy K, Cummin AR, et al. Developing a pictorial Epworth Sleepiness Scale. *Thorax* 2011;66(2):97-100. doi: 10.1136/thx.2010.136879

19. Garra G, Singer AJ, Taira BR, et al. Validation of the Wong-Baker FACES Pain Rating Scale in Pediatric Emergency Department Patients. *Academic Emergency Medicine* 2010;17(1):50-54. doi: 10.1111/j.1553-2712.2009.00620.x

20. Houts PS, Doak CC, Doak LG, et al. The role of pictures in improving health communication: A review of research on attention, comprehension, recall, and adherence. *Patient Education and Counseling* 2006;61(2):173-90. doi: https://doi.org/10.1016/j.pec.2005.05.004

21. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of Life Research* 2010;19(4):539-49. doi: 10.1007/s11136-010-9606-8

22. Mokkink LB, De Vet HCW, Prinsen CAC, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Quality of Life Research* 2018;27(5):1171-79. doi: 10.1007/s11136-017-1765-4

23. Basch E, Spertus J, Adams Dudley R, et al. Methods for Developing Patient-Reported Outcome-Based Performance Measures (PRO-PMs). *Value in Health* 2015;18(4):493-504. doi: 10.1016/j.jval.2015.02.018

24. de Vet HCW, Terwee CB, Mokkink LB, et al. Measurement in Medicine: A Practical Guide. Cambridge: Cambridge University Press 2011.

25. Bonett DG. Sample size requirements for estimating intraclass correlations with desired precision. *Stat Med* 2002;21(9):1331-5. doi: 10.1002/sim.1108

26. Portney LG, Watkins MP. Foundations of Clinical Research: Applications to Practice: Pearson/Prentice Hall 2009.

27. Gagnier JJ, Lai J, Mokkink LB, et al. COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. *Quality of Life Research* 2021;30(8):2197-218. doi: 10.1007/s11136-021-02822-4

28. Xie F, Li S-C, Roos EM, et al. Cross-cultural adaptation and validation of Singapore English and Chinese versions of the Knee injury and Osteoarthritis Outcome Score (KOOS) in Asians with knee osteoarthritis in Singapore. *Osteoarthritis and Cartilage* 2006;14(11):1098-103. doi: 10.1016/j.joca.2006.05.005

29. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: Ascertaining the minimal clinically important difference. *Controlled Clinical Trials* 1989;10(4):407-15. doi: http://dx.doi.org/10.1016/0197-2456(89)90005-6

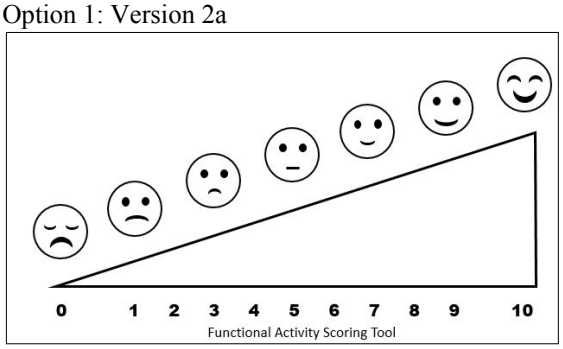
30. Kamper SJ, Maher CG, Mackay G. Global Rating of Change Scales: A Review of Strengths and Weaknesses and Considerations for Design. *Journal of Manual & Manipulative Therapy* 2009;17(3):163-70. doi: 10.1179/jmt.2009.17.3.163

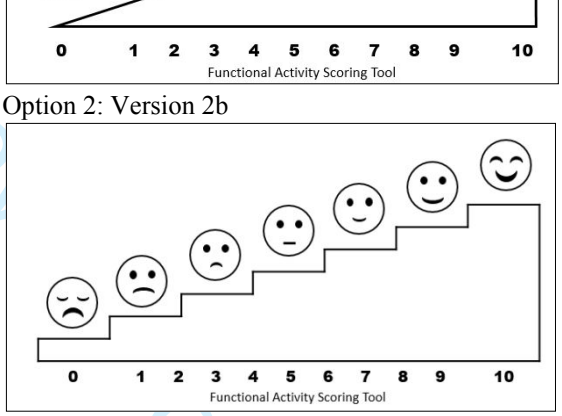
31. Waltz CF, Bausell BR. Nursing research: design statistics and computer analysis: Davis Fa 1981.

32. Zamanzadeh V, Ghahramanian A, Rassouli M, et al. Design and Implementation Content Validity Study: Development of an instrument for measuring Patient-Centered Communication. *Journal of Caring Sciences* 2015;4(2):165-78. doi: 10.15171/jcs.2015.017

33. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Res Nurs Health* 2007;30(4):459-67. doi: 10.1002/nur.20199

34. Yamada J, Stevens B, Sidani S, et al. Content Validity of a Process Evaluation Checklist to Measure Intervention Implementation Fidelity of the EPIC Intervention. *Worldviews on Evidence-Based Nursing* 2010;7(3):158-64. doi: 10.1111/j.1741-6787.2010.00182.x

<table>
<tr>
<td>Original: Version 1
• Inspired by pictographs and Wong-Baker scale
Inspired by pictographs and Wong-Baker scale
Inspired by pictographs and Wong-Baker scale</td>
<td>First Revision: Version 2a & 2b
• The expert review panel recommended delineation of 0 and 10 as stand-alone categories to imply 'inability to perform function' and 'no difficulty to perform function', respectively
• Corresponding faces to score:
- Face 1: Score 0
- Face 2: Score 1
- Face 3: Score 3
- Face 4: Score 5
- Face 5: Score 7
- Face 6: Score 9
- Face 7: Score10
• Pictorial enhancement using a gradient to show decreasing difficulty. Options include slope and step ladder.</td>
</tr>
<tr>
<td>
</td>
<td>Option 1: Version 2a
 

Option 2: Version 2b
 

*Version 2b (Step ladder) was preferred by the majority (6 out of 7; 85%)</td>
</tr>
<tr>
<td>Second Revision: Version 3
• Cognitive interviews with elderly patients using version 2b.
• Common feedback was received to add some wording to improve the clarity of pictorial aids.
• Developers added short descriptors, i.e. "unable to perform, extremely difficult, etc." and redesigned the font display "Functional Activity Scoring Tool" as the questionnaire title after discussion with the expert panel.
• Corresponding faces to score is determined as below:
- Face 1: Score 0
- Face 2: Score 1
- Face 3: Score 3
- Face 4: Score 5
- Face 5: Score 7
- Face 6: Score 9</td>
<td>Third Revision: Finalised Version
• Cognitive interviews with elderly patients using version 3.
• Common feedback received to add short instructions, statements or questions to improve the relevance of use.
• One Patient suggested adding a tick and X at both ends to improve the efficiency of use.
• After discussion with the expert panel, visual enhancement with eye-catching symbols, green tick and red X at both ends were included, and a simple question, "How difficult is it to perform your activity?" was added.</td>
</tr>
</table>

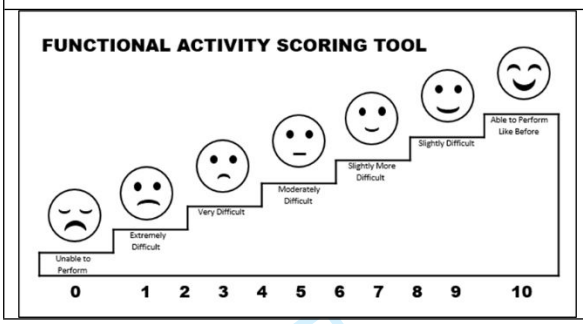
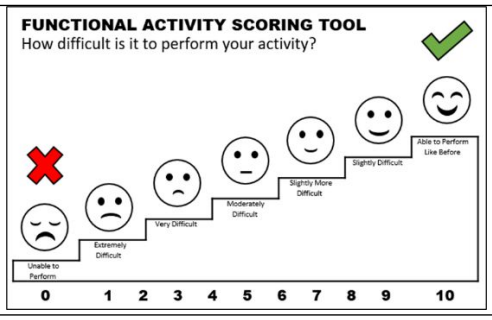
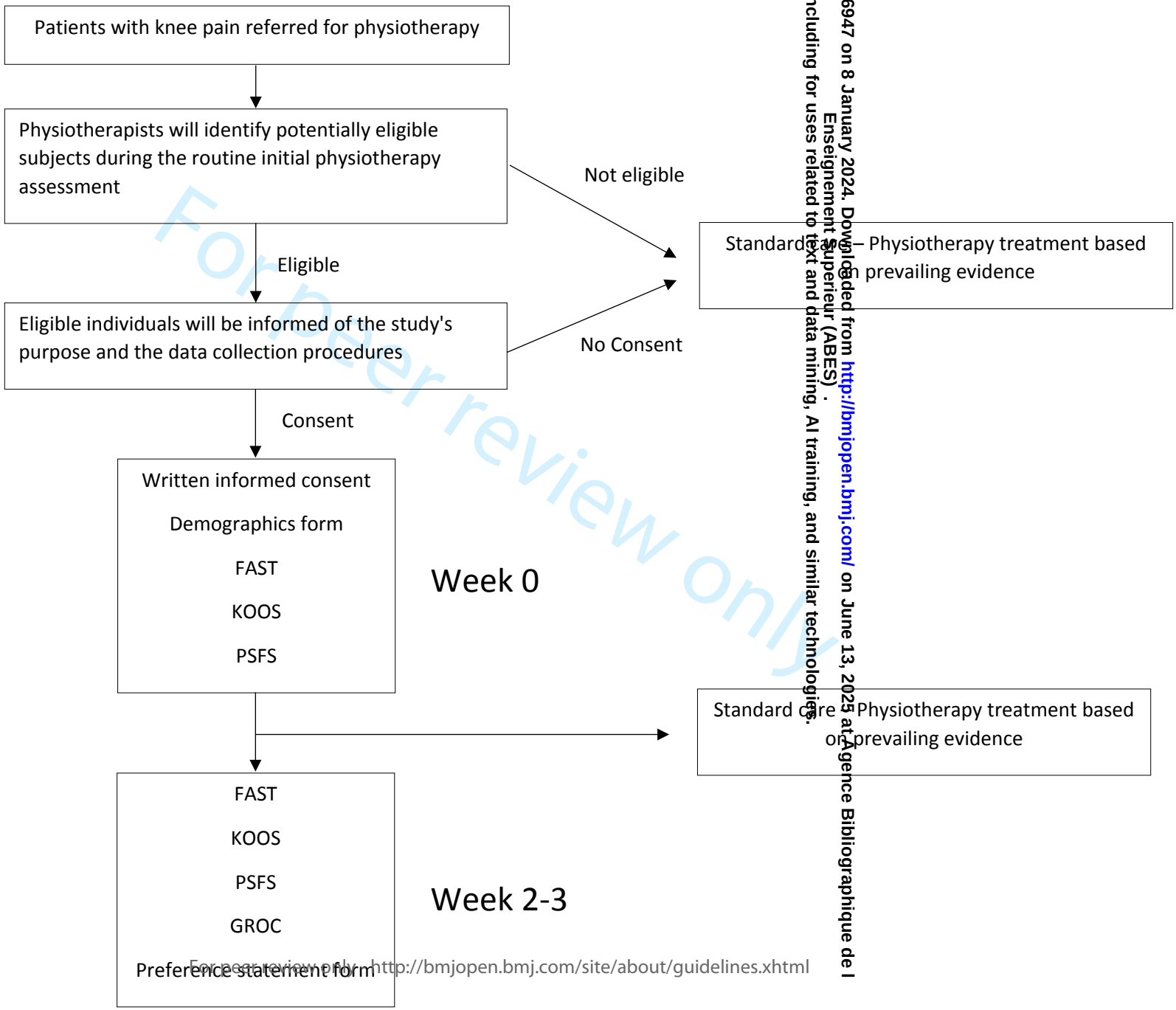
- Face 7: Score10 - Scores 2, 4, 6, and 8 are for patients who indicate the level of difficulty between two faces	
	

Figure 1 Functional Activity Scoring Tool (FAST) Conceptualisation process: Versions & Revisions



Clinical Review & Education

JAMA | Special Communication

Guidelines for Inclusion of Patient-Reported Outcomes in Clinical Trial Protocols
The SPIRIT-PRO Extension

Melanie Calvert, PhD; Derek Kyte, PhD; Rebecca Mercieca-Bebber, PhD; Anita Slade, PhD; An-Wen Chan, MD, DPhil; Madeleine T. King, PhD; and the SPIRIT-PRO Group

IMPORTANCE Patient-reported outcome (PRO) data from clinical trials can provide valuable evidence to inform shared decision making, labeling claims, clinical guidelines, and health policy; however, the PRO content of clinical trial protocols is often suboptimal. The SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) statement was published in 2013 and aims to improve the completeness of trial protocols by providing evidence-based recommendations for the minimum set of items to be addressed, but it does not provide PRO-specific guidance.

OBJECTIVE To develop international, consensus-based, PRO-specific protocol guidance (the SPIRIT-PRO Extension).

DESIGN, SETTING, AND PARTICIPANTS The SPIRIT-PRO Extension was developed following the Enhancing Quality and Transparency of Health Research (EQUATOR) Network's methodological framework for guideline development. This included (1) a systematic review of existing PRO-specific protocol guidance to generate a list of potential PRO-specific protocol items (published in 2014); (2) refinements to the list and removal of duplicate items by the International Society for Quality of Life Research (ISOQOL) Protocol Checklist Taskforce; (3) an international stakeholder survey of clinical trial research personnel, PRO methodologists, health economists, psychometricians, patient advocates, funders, industry representatives, journal editors, policy makers, ethicists, and researchers responsible for evidence synthesis (distributed by 38 international partner organizations in October 2016); (4) an international Delphi exercise (n = 137 invited; October 2016 to February 2017); and (5) consensus meeting (n = 30 invited; May 2017). Prior to voting, consensus meeting participants were informed of the results of the Delphi exercise and given data from structured reviews evaluating the PRO protocol content of 3 defined samples of trial protocols.

RESULTS The systematic review identified 162 PRO-specific protocol recommendations from 54 sources. The ISOQOL Taskforce (n = 21) reduced this to 56 items, which were considered by 138 international stakeholder survey participants and 99 Delphi panelists. The final wording of the SPIRIT-PRO Extension was agreed on at a consensus meeting (n = 29 participants) and reviewed by external group of experts during a consultation period. Eleven extensions and 5 elaborations to the SPIRIT 2013 checklist were recommended for inclusion in clinical trial protocols in which PROs are a primary or key secondary outcome. Extension items focused on PRO-specific issues relating to the trial rationale, objectives, eligibility criteria, concepts used to evaluate the intervention, time points for assessment, PRO instrument selection and measurement properties, data collection plan, translation to other languages, proxy completion, strategies to minimize missing data, and whether PRO data will be monitored during the study to inform clinical care.

CONCLUSIONS AND RELEVANCE The SPIRIT-PRO guidelines provide recommendations for items that should be addressed and included in clinical trial protocols in which PROs are a primary or key secondary outcome. Improved design of clinical trials including PROs could help ensure high-quality data that may inform patient-centered care.

JAMA. 2018;319(5):483-494. doi:10.1001/jama.2017.21903

Editorial page 450

Supplemental content

CME Quiz at jamanetwork.com/learning

Author Affiliations: Centre for Patient Reported Outcome Research, Institute of Applied Health Research, University of Birmingham, Birmingham, England (Calvert, Kyte, Slade); Sydney Medical School, Faculty of Medicine and School of Psychology, Faculty of Science, University of Sydney, Sydney, Australia (Mercieca-Bebber, King); Women's College Research Institute, Department of Medicine, University of Toronto, Toronto, Ontario, Canada (Chan).

Authors/Group Information: The SPIRIT-PRO Group authors and members are listed at the end of this article.

Corresponding Author: Melanie Calvert, PhD, Centre for Patient Reported Outcome Research, Institute of Applied Health Research, University of Birmingham, Edgbaston, Birmingham B15 2TT, England (m.calvert@bham.ac.uk).

Clinical trial protocols are essential documents that describe the study design and conduct. A protocol should provide sufficient detail to enable funders, reviewers, and ethics committees to appraise the scientific, methodological, and ethical rigor of the trial and for the research team to conduct a high-quality study.[1,2] Although trial protocols serve as the foundation for study planning, conduct, reporting, and appraisal, they vary greatly in content and quality.[1,2] To address this issue, the SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) statement was published in 2013.[1,2] SPIRIT provides an evidence-based list of items recommended for inclusion in trial protocols. It does not, however, provide specific guidance on protocol content relating to patient-reported outcomes (PROs), such as health-related quality of life or patient-reported symptoms.

The importance of PROs has been recognized by major international health policy and regulatory authorities and patients.[3-5] Patient-reported outcome results of trials, if captured in a scientifically rigorous way, may inform clinical decision making,[6] pharmaceutical labeling claims,[4,5] and product reimbursement and influence health care policy.[6] Despite this, the quality of PRO content in many protocols is often suboptimal, regardless of the degree of adherence to SPIRIT.[7-9] Because PROs are intrinsically subjective and require completion by patients within a specific time frame, they present a range of scientific and logistical challenges for researchers.[10-12] Comprehensive planning and instruction in the protocol can mitigate many PRO-specific issues through trial conduct and subsequent analysis and reporting. Protocol developers, particularly those not familiar with PRO methodology, may benefit from explanation of PRO-specific aspects to facilitate improvements in content.

The aim of this international project was to develop an evidence-based extension of the SPIRIT 2013 statement, identifying additional PRO items recommended for inclusion in clinical trial protocols (extensions), and to elaborate on the existing SPIRIT 2013 statement specifically as applied to PROs (elaborations).[13] This Special Communication describes the methods used to gain consensus on each additional SPIRIT-PRO extension/elaboration, provides a brief explanatory rationale, and includes PRO-specific items that may be included in supplemental trial documents.

Key Points

Question What information should be included in a clinical trial protocol when a patient-reported outcome (PRO) is a primary or key secondary outcome?

Findings Following an international consensus development process using the Enhancing Quality and Transparency of Health Research (EQUATOR) methodology, 16 PRO-specific items were recommended for inclusion in clinical trial protocols.

Meaning Inclusion of these items in clinical trial protocols may help improve the quality of PRO data.

SPIRIT-PRO Development Methods

The SPIRIT-PRO Extension was developed according to the Enhancing Quality and Transparency of Health Research (EQUATOR) Network's methodological framework for guideline development (eFigure 1 in Supplement 1).[14] This included a systematic review of existing PRO-specific protocol guidance,[15] a stakeholder survey of a group of international experts, and a Delphi exercise and consensus meeting, followed by consultation on the final SPIRIT-PRO Extension.[14] The systematic review comprised a search of the MEDLINE, EMBASE, CINHAL, and Cochrane Library databases (inception to February 2013) using the key words *patient-reported outcomes* or *health-related quality of life* in combination with *guidance*, *guidelines*, or *checklist*. Further guidance documents were identified via Google, Google Scholar, requests to members of the UK Clinical Research Collaboration registered clinical trials units, international experts, and citation and reference searches of included articles. Articles were deemed eligible if they contained guidance, a checklist, or both regarding PRO-related trial protocol content.[15]

eFigure 1 in Supplement 1 summarizes the methods and participants involved in the development of SPIRIT-PRO, the numbers of candidate items considered at each step, and the flow toward the final set of items included in SPIRIT-PRO. The eTable in Supplement 1 outlines the participant characteristics. Patient partners contributed to the co-design of the research and grant application and have provided input throughout the study.

Ethical Review of Study

Ethical approval was provided by the University of Birmingham Ethical Review Board (ERN_16-0819). Participant information was provided to potential participants electronically prior to survey completion and in advance of the consensus meeting. Survey participants provided electronic informed consent, and written consent was provided by the consensus meeting participants.

Systematic Review of Existing PRO-Specific Protocol Guidance and Development of the Delphi and Stakeholder Survey

Our systematic review of existing PRO protocol guidance identified 162 PRO-specific protocol recommendations from 54 sources, such as the need to specify the timing of PRO assessment, the provision of PRO data collection and analyses plans, and specification/justification for the chosen PRO questionnaire.[15] The International Society for Quality of Life Research (ISOQOL) Protocol Checklist Taskforce comprising international experts in PROs research and clinical trials (eTable and eAppendix in Supplement 1) reduced this list to 56 candidate items by removing or merging duplicate items, meaning that 56 items were included in the subsequent identical stakeholder and Delphi surveys. Survey participants were asked to rate the importance of including each of the 56 candidate items in the final SPIRIT-PRO Extension using a 9-point scale ranging from not important (1-3) to important but not critical (4-6) and critical (7-9). Respondents provided separate ratings according to whether a PRO was included as a primary vs secondary outcome in a trial.

International Stakeholder Survey

In 2016, an anonymized online international stakeholder survey was conducted targeting clinical trial research personnel, PRO

Box. Glossary

SPIRIT: Standard Protocol Items: Recommendations for Interventional Trials[1,2]

SPIRIT-PRO Extension item: an additional checklist item describing PRO protocol content to address an aspect of PRO assessment that is not adequately covered by SPIRIT, as judged by available evidence and expert opinion

SPIRIT Elaboration item: an elaboration of an existing SPIRIT item as applied to a specific context; in this instance, as applied to clinical trials assessing PROs

Patient-reported outcome (PRO): an outcome reported directly by patients themselves and not interpreted by an observer; PROs may include patient assessments of health status, quality of life, or symptoms[17]

Proxy-reported outcome: "a measurement based on a report by someone other than the patient reporting as if he or she is the patient"[4]

Health-related quality of life: "a multidimensional concept that usually includes self-report of the way in which physical, emotional, social, or other domains of well-being are affected by a disease or its treatment"[17]

Primary outcome/end point: the most important outcome in a trial, providing the most clinically relevant evidence directly related to the primary objective of the trial

Secondary outcomes/end point(s): outcomes prespecified in the protocol to assess additional effects of the intervention; some PROs may be identified as important or key secondary outcomes

Important or key secondary PROs/end points: Some PRO measures (particularly health-related quality-of-life measures) are multidimensional, producing several domain-specific outcome scales; eg, pain, fatigue, physical function, psychological distress. For any particular trial, it is likely that a particular PRO or PRO domain(s) will be more relevant than others, reflecting the expected effect(s) of the trial intervention(s) in the target patient population. These relevant PRO(s) and/or domain(s) may additionally constitute the important or key secondary PROs (identified a priori and specified as such in the trial protocol and statistical analysis plan) and will be the focus of hypothesis testing. In a regulatory environment, these outcomes may support a labeling claim. Because these outcomes are linked with hypotheses (CONSORT PRO Extension 2b),[17] they may be subject to *P*-value adjustment (or "α spending"). Patient-reported outcomes not only may provide evidence of efficacy/effectiveness but also may be intended to capture and provide evidence of safety and tolerability (eg, PRO-CTCAE).[18]

Concept: "The specific measurement goal (ie, the thing that is to be measured by a PRO instrument). In clinical trials, a PRO instrument can be used to measure the effect of a medical intervention on one or more concepts. PRO concepts represent aspects of how patients function or feel related to a health condition or its treatment."[4]

Domain: "A subconcept represented by a score of an instrument that measures a larger concept comprised of multiple domains. For example, psychological function is the larger concept containing the domains subdivided into items describing emotional function and cognitive function."[4]

Instrument: "A means to capture data (eg, a questionnaire) plus all the information and documentation that supports its use. Generally, that includes clearly defined methods and instructions for administration or responding, a standard format for data collection, and well-documented methods for scoring, analysis, and interpretation of results in the target patient population."[4]

Item: "an individual question, statement, or task (and its standardized response options) that is evaluated by the patient to address a particular concept"[4]

Time window: a predefined time frame before and after the protocol-specified PRO assessment time point whereby the result would still be deemed to be clinically relevant[19]

methodologists, health economists, psychometricians, patient advocates, funders, industry representatives, journal editors, policy makers, ethicists, and researchers responsible for evidence synthesis. Respondents were self-selected volunteers from a sample of eligible individuals recruited via 38 international partner organizations (listed in the eAppendix in Supplement 1). From these organizations, 138 participants provided anonymized survey results, which informed round 2 of the Delphi panel exercise.

International Delphi Exercise

In parallel with the international stakeholder survey, 114 key experts from the ISOQOL Protocol Checklist Taskforce, international partner organizations, and other experts known or recommended to the SPIRIT-PRO Executive (eAppendix in Supplement 1) were invited to join an international, multidisciplinary expert Delphi panel. Delphi panelists were advised not to complete the stakeholder survey to avoid double counting of results. Ninety-nine Delphi panelists completed 2 rounds of online surveys, and results informed the subsequent international consensus meeting. Data collected from the stakeholder and round 1 Delphi surveys were anonymized, and the item-level results were provided to the Delphi panel for consideration prior to voting in Delphi round 2. Further details and the results of the Delphi and stakeholder surveys are available on the study website.[16]

Consensus Meeting

Using the results from the stakeholder survey and Delphi process, the SPIRIT-PRO Operations Team (M.C., D.K., R.M.B., A.S., and M.K.) mapped the 56 candidate SPIRIT-PRO items to corresponding SPIRIT-2013 items, revising wording as needed to address stakeholder/Delphi panelist comments. For each candidate SPIRIT-PRO item, the Operations Team presented the consensus meeting delegates with recommendations for SPIRIT elaborations and extensions (see **Box** for definitions) based on a decision tree (eFigure 2 in Supplement 1) that incorporated information drawn from the Delphi survey and 3 separate reviews of PRO protocol content (n = 207 protocols): protocols from the UK National Institute for Health Research (NIHR) Health Technology Assessment program[7]; cancer trial protocols from the NIHR[8]; and international ovarian cancer protocols.[9] Twenty-nine participants purposively sampled from the Delphi panel attended the 2-day consensus meeting hosted by the University of Birmingham in May 2017 (eTable in Supplement 1). The meeting was designed to

seek consensus on the content of the SPIRIT-PRO Extension. Meeting participants were invited to consider the focus of the guidance and agreed that it should apply to trials in which PROs are a primary or key secondary outcome (as defined in the glossary [Box]). Delegates anonymously voted using Turning Point)/ Responseware software, version 5.1 (Turning Technologies LLC), to (1) include the candidate item as recommended; (2) exclude the item; (3) or initiate further discussion. Key research evidence (round 2 Delphi survey results and systematic review data) presented to meeting participants is available in Supplement 2. Consensus meeting participants were also invited to review Delphi results for recommendation on where to include each of the candidate items in addition to or instead of the trial protocol (eg, guidance/training for trial staff, information/guidance for study participants, or the statistical analysis plan).

Final Consultation

Following the consensus meeting, attendees commented on wording and agreed on the penultimate SPIRIT-PRO Extension content. Broader feedback on the final guidance was sought from the Delphi panel and international partners during a 3-week consultation period. Final edits in response to feedback were made by the Operations Team and agreed on by the SPIRIT-PRO Group.

Results

SPIRIT-PRO Checklist Items and Explanation

The final SPIRIT-PRO Extension recommends that, in conjunction with existing SPIRIT 2013 items, 16 items (11 extensions and 5 elaborations) should be routinely addressed in all clinical trial protocols in which PROs are a primary or key secondary outcome. Further information regarding the SPIRIT 2013 items has been published by Chan et al.[1,2] The **Table** lists the items of the SPIRIT 2013 checklist and the SPIRIT-PRO extensions and elaborations. In total, the 11 extensions and 5 elaborations incorporated content from 34 of the original 56 candidate items, comprising 27 items that were merged during the consensus meeting and a further 7 items that remained unchanged. One new item, SPIRIT-18a(iii)-PRO Extension, was generated through discussion. Definitions of key terms are contained in the glossary (Box). A brief explanation for each PRO extension or elaboration is included herein, with references to supporting empirical evidence when available (items 6a through 22). Item 5a was not supported by empirical evidence but was supported by expert opinion drawn from our systematic review of PRO protocol guidance[15] and, in line with the development of the original SPIRIT statement,[1,2] was underpinned by a strong pragmatic rationale.

Administrative Information

SPIRIT-5a-PRO Elaboration: Specify the individual(s) responsible for the PRO content of the trial protocol.

Explanation: Providing information (eg, name, affiliation, contact details) on who wrote the PRO-specific aspects of the trial protocol promotes transparency and accountability and identifies the appropriate point of contact for resolution of any PRO-specific queries. When patients have actively contributed to this process, this should be documented as per recent guidance for the reporting of patient and public involvement.[21]

SPIRIT-6a-PRO Extension: Describe the PRO-specific research question and rationale for PRO assessment and summarize PRO findings in relevant studies.

Explanation: Inclusion of PROs in a trial requires careful consideration and planning. A clearly defined question helps with selection of measures and specification of hypotheses and analyses. Evidence suggests that many trials include PROs without specifying the PRO-specific research question and without a rationale or any reference to PROs in related studies.[7-9] Consequently, staff and patients may not understand why PROs are being assessed, and missing data may result.[7-12] When the PRO is a secondary outcome, a brief rationale may be adequate.

SPIRIT-7-PRO Extension: State specific PRO objectives or hypotheses (including relevant PRO concepts/domains).

Explanation: PRO measures may be multidimensional (eg, health-related quality of life) or unidimensional (eg, specific symptoms such as pain), and assessments may be scheduled at several time points during a trial. Prespecification of objectives and hypotheses encourages identification of key PRO domains and time points, reducing the risk of multiple statistical testing and selective reporting of PROs based on statistically significant results (see also PRO elaboration 20a below).[4]

Methods: Participants, Interventions, and Outcomes

SPIRIT-10-PRO Extension: Specify any PRO-specific eligibility criteria (eg, language/reading requirements or prerandomization[22] completion of PRO). If PROs will not be collected from the entire study sample, provide a rationale and describe the method for obtaining the PRO subsample.

Explanation: Any PRO-specific eligibility criteria should be considered at the design stage of the trial and clearly specified in the protocol. In large trials, sufficient power may be achieved by collecting PROs from a representative subset of participants, while in some trials it may not be possible to collect PROs in the entire population (eg, because of nonavailability of validated questionnaires in all languages)[8]; in such instances, the rationale for the sampling method should be described.

SPIRIT-12-PRO Extension: Specify the PRO concepts/domains used to evaluate the intervention (eg, overall health-related quality of life, specific domain, specific symptom) and, for each one, the analysis metric (eg, change from baseline, final value, time to event) and the principal time point or period of interest.

Explanation: The PRO concepts/domains and time points for assessment should closely align with the trial objectives and hypotheses. Because of the risk of multiple statistical testing, the domain(s) and principal time point(s) for analyses should be specified a priori.[4,23]

SPIRIT-13-PRO Extension: Include a schedule of PRO assessments, providing a rationale for the time points, and justifying if the initial assessment is not prerandomization. Specify time windows, whether PRO collection is prior to clinical assessments, and, if using multiple questionnaires, whether order of administration will be standardized.

Explanation: Provision of an easy-to-follow schedule will assist staff and may help reduce missing data.[22] Collecting PRO data prior

Table. SPIRIT 2013 and SPIRIT-PRO Extension Checklist: Recommended Items to Address in a Clinical Trial Protocol[a]

SPIRIT Section	SPIRIT Item No.	SPIRIT Item Description	SPIRIT-PRO Item No.	SPIRIT-PRO Extension or Elaboration Item Description	Addressed on Page No.[b]
Administrative Information					
Title	1	Descriptive title identifying the study design, population, interventions, and, if applicable, trial acronym			
Trial registration	2a	Trial identifier and registry name (if not yet registered, name of intended registry)			
	2b	All items from the World Health Organization Trial Registration Data Set			
Protocol version	3	Date and version identifier			
Funding	4	Sources and types of financial, material, and other support			
Roles and responsibilities	5a	Names, affiliations, and roles of protocol contributors	SPIRIT-5a-PRO Elaboration	Specify the individual(s) responsible for the PRO content of the trial protocol.	
	5b	Name and contact information for the trial sponsor			
	5c	Role of study sponsor and funders, if any, in study design; collection, management, analysis, and interpretation of data; writing of the report; and the decision to submit the report for publication, including whether they will have ultimate authority over any of these activities			
	5d	Composition, roles, and responsibilities of the coordinating center, steering committee, end-point adjudication committee, data management team, and other individuals or groups overseeing the trial, if applicable (see item 21a for data monitoring committee)			
Introduction					
Background and rationale	6a	Description of research question and justification for undertaking the trial, including summary of relevant studies (published and unpublished) examining benefits and harms for each intervention	SPIRIT-6a-PRO Extension	Describe the PRO-specific research question and rationale for PRO assessment and summarize PRO findings in relevant studies.	
	6b	Explanation for choice of comparators			
Objectives	7	Specific objectives or hypotheses	SPIRIT-7-PRO Extension	State specific PRO objectives or hypotheses (including relevant PRO concepts/domains).	
Trial design	8	Description of trial design, including type of trial (eg, parallel group, crossover, factorial, single group), allocation ratio, and framework (eg, superiority, equivalence, noninferiority, exploratory)			
Methods: Participants, Interventions, and Outcomes					
Study setting	9	Description of study settings (eg, community clinic, academic hospital) and list of countries where data will be collected; reference to where list of study sites can be obtained			
Eligibility criteria	10	Inclusion and exclusion criteria for participants; if applicable, eligibility criteria for study centers and individuals who will perform the interventions (eg, surgeons, psychotherapists)	SPIRIT-10-PRO Extension	Specify any PRO-specific eligibility criteria (eg, language/reading requirements or prerandomization completion of PRO). If PROs will not be collected from the entire study sample, provide a rationale and describe the method for obtaining the PRO subsample.	

(continued)

Table. SPIRIT 2013 and SPIRIT-PRO Extension Checklist: Recommended Items to Address in a Clinical Trial Protocol[a] (continued)

SPIRIT Section	SPIRIT Item No.	SPIRIT Item Description	SPIRIT-PRO Item No.	SPIRIT-PRO Extension or Elaboration Item Description	Addressed on Page No.[b]
Interventions	11a	Interventions for each group with sufficient detail to allow replication, including how and when they will be administered			
	11b	Criteria for discontinuing or modifying allocated interventions for a given trial participant (eg, drug dose change in response to harms, participant request, or improving/worsening disease)			
	11c	Strategies to improve adherence to intervention protocols and any procedures for monitoring adherence (eg, drug tablet return, laboratory tests)			
	11d	Relevant concomitant care and interventions that are permitted or prohibited during the trial			
Outcomes	12	Primary, secondary, and other outcomes, including the specific measurement variable (eg, systolic blood pressure), analysis metric (eg, change from baseline, final value, time to event), method of aggregation (eg, median, proportion), and time point for each outcome; explanation of the clinical relevance of chosen efficacy and harm outcomes is strongly recommended	SPIRIT-12-PRO Extension	Specify the PRO concepts/domains used to evaluate the intervention (eg, overall health-related quality of life, specific domain, specific symptom) and, for each one, the analysis metric (eg, change from baseline, final value, time to event) and the principal time point or period of interest.	
Participant timeline	13	Time schedule of enrollment, interventions (including any run-ins and washouts), assessments, and visits for participants; a schematic diagram is highly recommended (see figure in Chan et al[1,2])	SPIRIT-13-PRO Extension	Include a schedule of PRO assessments, providing a rationale for the time points, and justifying if the initial assessment is not prerandomization. Specify time windows, whether PRO collection is prior to clinical assessments, and, if using multiple questionnaires, whether order of administration will be standardized.	
Sample size	14	Estimated number of participants needed to achieve study objectives and how it was determined, including clinical and statistical assumptions supporting any sample size calculations	SPIRIT-14-PRO Elaboration	When a PRO is the primary end point, state the required sample size (and how it was determined) and recruitment target (accounting for expected loss to follow-up). If sample size is not established based on the PRO end point, then discuss the power of the principal PRO analyses.	
Recruitment	15	Strategies for achieving adequate participant enrollment to reach target sample size			
Methods: Assignment of Interventions (for Clinical Trials)					
Allocation					
Sequence generation	16a	Method of generating the allocation sequence (eg, computer-generated random numbers), and list of any factors for stratification. To reduce predictability of a random sequence, details of any planned restriction (eg, blocking) should be provided in a separate document that is unavailable to those who enroll participants or assign interventions.			
Allocation concealment mechanism	16b	Mechanism of implementing the allocation sequence (eg, central telephone; sequentially numbered, opaque, sealed envelopes), describing any steps to conceal the sequence until interventions are assigned			
Implementation	16c	Who will generate the allocation sequence, who will enroll participants, and who will assign participants to interventions			
Blinding (masking)	17a	Who will be blinded after assignment to interventions (eg, trial participants, care providers, outcome assessors, data analysts) and how			
	17b	If blinded, circumstances under which unblinding is permissible and procedure for revealing a participant's allocated intervention during the trial			

(continued)

Table. SPIRIT 2013 and SPIRIT-PRO Extension Checklist: Recommended Items to Address in a Clinical Trial Protocol[a] (continued)

SPIRIT Section	SPIRIT Item No.	SPIRIT Item Description	SPIRIT-PRO Item No.	SPIRIT-PRO Extension or Elaboration Item Description	Addressed on Page No.[b]
Methods: Data Collection, Management, and Analysis					
Data collection methods	18a	Plans for assessment and collection of outcome, baseline, and other trial data, including any related processes to promote data quality (eg, duplicate measurements, training of assessors) and description of study instruments (eg, questionnaires, laboratory tests) along with their reliability and validity, if known; reference to where data collection forms can be found, if not in the protocol	SPIRIT-18a (i)-PRO Extension	Justify the PRO instrument to be used and describe domains, number of items, recall period, and instrument scaling and scoring (eg, range and direction of scores indicating a good or poor outcome). Evidence of PRO instrument measurement properties, interpretation guidelines, and patient acceptability and burden should be provided or cited if available, ideally in the population of interest. State whether the measure will be used in accordance with any user manual and specify and justify deviations if planned.	
			SPIRIT-18a (ii)-PRO Extension	Include a data collection plan outlining the permitted mode(s) of administration (eg, paper, telephone, electronic, other) and setting (eg, clinic, home, other).	
			SPIRIT-18a (iii)-PRO Extension	Specify whether more than 1 language version will be used and state whether translated versions have been developed using currently recommended methods.	
			SPIRIT-18a (iv)-PRO Extension	When the trial context requires someone other than a trial participant to answer on his or her behalf (a proxy-reported outcome), state and justify the use of a proxy respondent. Provide or cite evidence of the validity of proxy assessment if available.	
	18b	Plans to promote participant retention and complete follow-up, including list of any outcome data to be collected for participants who discontinue or deviate from intervention protocols	SPIRIT-18b (i)-PRO Extension	Specify PRO data collection and management strategies for minimizing avoidable missing data.	
			SPIRIT-18b (ii)-PRO Elaboration	Describe the process of PRO assessment for participants who discontinue or deviate from the assigned intervention protocol.	
Data management	19	Plans for data entry, coding, security, and storage, including any related processes to promote data quality (eg, double data entry; range checks for data values); reference to where details of data management procedures can be found, if not in the protocol			
Statistical methods	20a	Statistical methods for analyzing primary and secondary outcomes. Reference to where other details of the statistical analysis plan can be found, if not in the protocol	SPIRIT-20a-PRO Elaboration	State PRO analysis methods, including any plans for addressing multiplicity/type I (α) error.	
	20b	Methods for any additional analyses (eg, subgroup and adjusted analyses)			
	20c	Definition of analysis population relating to protocol nonadherence (eg, as randomized analysis) and any statistical methods to handle missing data (eg, multiple imputation)	SPIRIT-20c-PRO Elaboration	State how missing data will be described and outline the methods for handling missing items or entire assessments (eg, approach to imputation and sensitivity analyses).	
Methods: Monitoring					
	21a	Composition of data monitoring committee; summary of its role and reporting structure; statement of whether it is independent from the sponsor and competing interests; and reference to where further details about its charter can be found, if not in the protocol (alternatively, an explanation of why a data monitoring committee is not needed)			
	21b	Description of any interim analyses and stopping guidelines, including who will have access to these interim results and make the final decision to terminate the trial			

(continued)

Table. SPIRIT 2013 and SPIRIT-PRO Extension Checklist: Recommended Items to Address in a Clinical Trial Protocol[a] (continued)

SPIRIT Section	SPIRIT Item No.	SPIRIT Item Description	SPIRIT-PRO Item No.	SPIRIT-PRO Extension or Elaboration Item Description	Addressed on Page No.[b]
Harms	22	Plans for collecting, assessing, reporting, and managing solicited and spontaneously reported adverse events and other unintended effects of trial interventions or trial conduct	SPIRIT-22-PRO Extension	State whether or not PRO data will be monitored during the study to inform the clinical care of individual trial participants and, if so, how this will be managed in a standardized way. Describe how this process will be explained to participants; eg, in the participant information sheet and consent form.	
Auditing	23	Frequency and procedures for auditing trial conduct, if any, and whether the process will be independent from investigators and sponsor(s)			
Ethics and Dissemination					
Research ethics approval	24	Plans for seeking research ethics committee/institutional review board approval			
Protocol amendments	25	Plans for communicating important protocol modifications (eg, changes to eligibility criteria, outcomes, analyses) to relevant parties (eg, investigators, research ethics committees/institutional review boards, trial participants, trial registries, journals, regulators)			
Consent or assent	26a	Who will obtain informed consent or assent from potential trial participants or authorized surrogates and how (see item 32)			
	26b	Additional consent provisions for collection and use of participant data and biological specimens in ancillary studies, if applicable			
Confidentiality	27	How personal information about potential and enrolled participants will be collected, shared, and maintained to protect confidentiality before, during, and after the trial			
Declaration of interests	28	Financial and other competing interests for principal investigators for the overall trial and each study site			
Access to data	29	Statement of who will have access to the final trial data set and disclosure of contractual agreements that limit such access for investigators			
Ancillary and posttrial care	30	Provisions, if any, for ancillary and posttrial care and for compensation to those who are harmed by trial participation			
Dissemination policy	31a	Plans for investigators and sponsor(s) to communicate trial results to participants, health care professionals, the public, and other relevant groups (eg, via publication, reporting in results databases, or other data-sharing arrangements), including any publication restrictions			
	31b	Authorship eligibility guidelines and any intended use of professional writers			
	31c	Plans, if any, for granting public access to the full protocol, participant-level data set, and statistical code			
Appendixes					
Informed consent materials	32	Model consent form and other related documentation given to participants and authorized surrogates			
Biological specimens	33	Plans for collection, laboratory evaluation, and storage of biological specimens for genetic or molecular analysis in the current trial and for future use in ancillary studies, if applicable			

Abbreviations: SPIRIT, Standard Protocol Items: Recommendations for Interventional Trials; PRO, patient-reported outcome.

[a] It is strongly recommended that this checklist be read in conjunction with the SPIRIT 2013 Explanation & Elaboration for important clarification on the items.[1,2,20] Amendments to the protocol should be tracked and dated.

[b] Indicates page numbers to be completed by authors during protocol development.

to randomization helps ensure an unbiased baseline assessment, and if specified as an eligibility criterion, ensures data completeness. This is important because baseline PRO data are often used as a covariate in analyses and are essential to calculating change from baseline. Completion of PROs prior to clinical assessments (as these may influence patient responses) and standardization of the order of questionnaire administration are advised to help reduce measurement error.[24] Allowable time windows for each scheduled PRO assessment should be specified to ensure that PRO data collection captures the effect of the clinical event(s) of interest.

SPIRIT-14-PRO Elaboration: When a PRO is the primary end point, state the required sample size (and how it was determined) and recruitment target (accounting for expected loss to follow-up). If sample size is not established based on the PRO end point, then discuss the power of the principal PRO analyses.

Explanation: In studies in which PROs are the primary outcome or end point, the target sample size will generally be based on an a priori sample size calculation for that end point.[23] Ideally, the criteria for clinical significance (eg, minimal important difference, responder definition) should be specified when known.[25,26] If PROs are a secondary end point, researchers should specify whether the sample size provides sufficient power to test the principal PRO hypotheses.[23]

Methods: Data Collection, Management, and Analysis

SPIRIT-18a(i)-PRO Extension: Justify the PRO instrument to be used and describe domains, number of items, recall period, instrument scaling and scoring (eg, range and direction of scores indicating a good or poor outcome). Evidence of PRO instrument measurement properties, interpretation guidelines, and patient acceptability and burden should be provided or cited if available, ideally in the population of interest. State whether the measure will be used in accordance with any user manual and specify and justify deviations if planned.

Explanation: The selection of PROs to be used in a clinical trial requires careful consideration. Ideally, the measure should be validated in the target population.[27] Consideration should be given to the number of questionnaires to be used, acceptability of the questions, and the likely patient burden (eg, time taken for completion, cognitive burden, emotional burden). Justification for the measures selected will help trial personnel understand why specific measures are being used.[10] Questionnaires should be used in accordance with any existing user manuals to promote data quality and ensure standardized scoring, and any deviations should be described.

SPIRIT-18a(ii)-PRO Extension: Include a data collection plan outlining the permitted mode(s) of administration (eg, paper, telephone, electronic, other) and setting (eg, clinic, home, other).

Explanation: It is important that both research personnel and trial participants understand how, when, and where PRO data will be collected in the study. Increasingly, electronic PRO assessment is undertaken in trials, so evidence of equivalence between different modes of administration should be considered.[28] If electronic PRO measures contain only minor modifications with respect to the paper-based versions, usability testing and cognitive debriefing may provide sufficient evidence of equivalence.[28,29] The setting for PRO data collection should be described and standardized across trial intervention groups and sites.

SPIRIT-18a(iii)-PRO Extension: Specify whether more than 1 language version will be used and state whether translated versions have been developed using currently recommended methods.

Explanation: Multinational trials, or national trials involving participants with different languages, require measures that have been translated and culturally adapted where needed using appropriate methodology.[11,30] This may influence the selection of measure to be used because inclusion of a wide range of participants can help ensure the generalizability of trial results. Plans to use translated versions should be specified in the protocol, citing references when available.

SPIRIT-18a(iv)-PRO Extension: When the trial context requires someone other than a trial participant to answer on his or her behalf (a proxy-reported outcome), state and justify the use of a proxy respondent. Provide or cite evidence of the validity of proxy assessment if available.

Explanation: In some contexts, such as trials involving young children or cognitively impaired participants, it may be necessary for someone other than a trial participant to respond on that participant's behalf. Clear justification and specification of proxy reporting in the protocol allows external reviewers to assess potential bias and facilitates trial reporting in accordance with CONSORT-PRO.[17] Evidence of the size and direction of proxy bias is a key aspect of the validity of proxy versions of PRO measures, informing valid interpretation, and comparison of results. The European Medicines Agency states that "in general proxy reporting should be avoided, unless the use of such proxy raters may be the only effective means of obtaining information that might otherwise be lost."[5] The US Food and Drug Administration also discourages the use of proxy-reported outcomes to inform labeling claims, recommending observer reports instead.[4]

SPIRIT-18b(i)-PRO Extension: Specify PRO data collection and management strategies for minimizing avoidable missing data.

Explanation: Missing data are a particular problem for PROs because participants with the poorest outcomes in a trial often are those who do not complete planned PRO assessments, and data cannot be obtained retrospectively beyond the time frame of interest or from medical records. This is a potentially significant source of bias and may reduce trial power.[31] It is important to note that not all missing PRO data are avoidable: patients have the right to decide not to complete questionnaires. Common reasons for avoidable missing PRO data are administrative errors, lack of explanation of the importance of PRO data, and overly burdensome questionnaires. Addressing these in the protocol should help minimize avoidable missing data. A recent systematic review provides a range of design, implementation, and reporting strategies to help minimize and address missing PRO data.[22] Examples of protocol content include ensuring that PRO end points and hypotheses are clearly defined and scientifically compelling, providing a rationale for PRO assessment, clearly specifying the PRO assessment time points, defining acceptable PRO assessment time windows, aligning PRO assessment time points to clinic visits (if clinically informative), minimizing patient burden, and specifying the importance of complete PRO data.[22]

SPIRIT-18b(ii)-PRO Elaboration: Describe the process of PRO assessment for participants who discontinue or deviate from the assigned intervention protocol.

Explanation: A clear plan for collection of PROs for trial participants who withdraw early from a study or who discontinue the intervention helps minimize bias,[32] ensures that staff collect all required PRO data in a standardized and timely way, and may assist ethical appraisal of the study.

SPIRIT-20a-PRO Elaboration: State PRO analysis methods, including any plans for addressing multiplicity/type I (α) error.

Explanation: Many questionnaires, such as health-related quality-of-life measures, are multidimensional and therefore may yield several summary scores (eg, multiple domains and an overall score). Furthermore, PROs are usually assessed at multiple time points. Statistical analysis of all domains and time points implies multiple hypothesis testing, which inflates the probability of false-positive results (type I error).[23] This can be contained by prespecifying the key PRO domain(s) or overall score of interest and the principal time point(s). Any plans to address multiplicity, such as stepwise or sequential analyses, whereby multiple end points are tested in a defined sequence that contains the overall type I error to the desired level, or conventional nonhierarchical methods (eg, Bonferroni correction), should be specified a priori.[4] The protocol should either fully address these issues or provide a summary with reference to where full details can be found (eg, in the statistical analysis plan).

SPIRIT-20c-PRO Elaboration: State how missing data will be described and outline the methods for handling missing items or entire assessments (eg, approach to imputation and sensitivity analyses).

Explanation: There are 2 levels of missing PRO data: (1) patient completion of some but not all items within an instrument and (2) absence of the entire PRO assessment. Whether and how missing items should be imputed is usually specified in an instrument's scoring algorithm. When entire PRO assessments are missed, analysis requires assumptions about why those data were missing (ie, the missing data mechanism). There are a range of statistical approaches, each with specific assumptions. Common methods include complete case analysis, imputation (various approaches), a range of maximum likelihood modeling approaches, and sensitivity analysis.[32] Inappropriate method selection may lead to potentially biased and misleading results.[22,32] The protocol should acknowledge and summarize these issues, with full details provided in the statistical analysis plan.

Methods: Monitoring

SPIRIT-22-PRO Extension: State whether or not PRO data will be monitored during the study to inform the clinical care of individual trial participants and, if so, how this will be managed in a standardized way. Describe how this process will be explained to participants; eg, in the participant information sheet and consent form.

Explanation: Evidence suggests that monitoring and management of PRO alerts (psychological distress or physical symptoms evident from PRO responses that may require an immediate response) vary across and within trials.[10,11,33] To protect the interests of trial participants and minimize potential bias, it is important to specify plans for monitoring.[34] If monitoring is not planned (for example, in a low-risk study in which alerts are not anticipated), this should also be briefly stated in the protocol, the participant information sheet, and the consent form. Alternative support mechanisms for patients should be outlined.

Supplementary Trial Documents

Supplement 3 outlines additional items recommended for inclusion in other trial documentation, such as the statistical analysis plan, participant information sheet, and training and guidance documents for staff.

Discussion

The SPIRIT-PRO Extension provides international consensus-based guidance on PRO-specific information that should be included in clinical trial protocols. It comprises 16 items: 5 elaborations to existing SPIRIT 2013 items in the context of PROs and 11 new extensions for use alongside the existing SPIRIT 2013 guidance.[1,2] It is important to note that these are minimum requirements and that there may be value in including additional items in the protocols, in supplementary information, or in both, as outlined in Supplement 3. Although this guidance has been developed for trials for which PROs are a primary or key secondary outcome, research groups that create protocols are encouraged to consider use of this guidance in all trials or clinical research studies in which PROs are collected, including if PROs are exploratory end points. The guidance does not aim to be prescriptive regarding how information should be included, as this may vary depending on the research setting and local requirements. Further details of empirical evidence underpinning the SPIRIT-PRO items and examples for implementation will be provided in a future publication on the PROlearn[35] and SPIRIT Initiative[20] websites and will be facilitated through further development of the SPIRIT 2013 implementation tool SEPTRE[20] (SPIRIT Electronic Protocol Tool and Resource) and through dissemination via international partners (eAppendix in Supplement 1). Inclusion of PRO-specific protocol content will facilitate appraisal of the PRO elements by funders, reviewers, research ethics committees, and patient partners. The SPIRIT-PRO Extension is intended to encourage and facilitate careful planning of PRO components of trials and thereby improve PRO trial design. Consequently, this is expected to help staff and patients understand the rationale for PRO assessment, improve PRO data completeness and quality, facilitate high-quality analysis and reporting, and ultimately improve the quality of the global PRO evidence base.

To maximize the benefit of PRO data in policy and in practice, it is recommended that careful consideration be given to the selection of outcomes and measures,[36,37] analysis of PRO data,[4,5,38] and transparent reporting in accordance with CONSORT-PRO.[17] Patient and public involvement in all of these aspects can help ensure that PRO selection and application is transparent, relevant, and acceptable.[39,40] Consistent with this philosophy, patient partners have been involved in all aspects of the development of the SPIRIT-PRO Extension.[39,40] Ultimately, high-quality PRO trial results will help ensure that patients' voices are central to informing shared decision making, labeling claims, clinical guidelines, and health policy, making patient-centered care a reality.

This study has several limitations. First, as the international stakeholder survey included an anonymized nonprobability sample, we were unable to determine either the level or characteristics of nonresponders, meaning that the results of the survey could be affected by nonresponse bias. Second, respondents to the stakeholder survey were self-selecting and Delphi and

consensus meeting participants were purposively sampled based on their roles and expertise relating to PROs. Participants are therefore more likely to have more knowledge relating to PROs than broader research personnel. Third, the systematic review underpinning the process was conducted in 2013; however, throughout the guideline development process, the expert Delphi and consensus meeting participants are encouraged to highlight any additional relevant publications.

Conclusions

The SPIRIT-PRO guidelines provide recommendations for items that should be addressed and included in clinical trial protocols in which PROs are a primary or key secondary outcome. Improved design of clinical trials including PROs could help ensure high-quality data that may inform patient-centered care.

ARTICLE INFORMATION

Accepted for Publication: December 26, 2018.

Authors/SPIRIT-PRO Group Members: Amanda Hunn, MA; Andrew Bottomley, PhD; Antoine Regnault, PhD; An-Wen Chan, MD; Carolyn Ells, PhD; Daniel O'Connor, PhD; Dennis Revicki, PhD; Donald Patrick, PhD; Doug Altman, DSc; Ethan Basch, MD; Galina Velikova, PhD; Gary Price; Heather Draper, PhD; Jane Blazeby, MD; Jane Scott, PhD; Joanna Coast, PhD; Josephine Norquist, PhD; Julia Brown, MSc; Kirstie Haywood, PhD; Laura Lee Johnson, PhD; Lisa Campbell, MD; Lori Frank, PhD; Maria von Hildebrand; Michael Brundage, MD; Michael Palmer, MSc; Paul Kluetz, MD; Richard Stephens, MA; Robert M. Golub, MD; Sandra Mitchell, PhD; Trish Groves, MRCPsych.

Affiliations of Authors/SPIRIT-PRO Group Members: Health Research Authority, London, England (Hunn); European Organization for Research and Treatment of Cancer, Brussels, Belgium (Bottomley); Modus Outcomes, Lyon, France (Regnault); University of Toronto, Toronto, Ontario, Canada (Chan); Panel on Research Ethics, Toronto, Ontario, Canada (Ells); Medicines and Healthcare Products Regulatory Agency, London, England (O'Connor, Campbell); Evidera, Bethesda, Maryland (Revicki); University of Washington, Seattle (Patrick); University of Oxford, Oxford, England (Altman); University of North Carolina, Chapel Hill (Basch); University of Leeds, Leeds, England (Velikova, Brown); Patient Partner, Centre for Patient Reported Outcome Research, Institute of Applied Health Research, University of Birmingham, Birmingham, England (Price, von Hildebrand); University of Warwick, Coventry, England (Draper, Haywood); University of Bristol, Bristol, England (Blazeby, Coast); Janssen Global Services, Johnson and Johnson, High Wycombe, England (Scott); Merck Sharp & Dohme Corporation, Whitehouse Station, New Jersey (Norquist); US Food and Drug Administration, Silver Spring, Maryland (Johnson, Kluetz); Patient-Centered Outcomes Research Institute, Washington, DC (Frank); Queen's University, Kingston, Ontario, Canada (Brundage, Palmer); National Cancer Research Institute Consumer Forum, London, England (Stephens); The JAMA Network, Chicago, Illinois (Golub); National Cancer Institute, Rockville, Maryland (Mitchell); *British Medical Journal*, London, England (Groves).

Author Contributions: Drs Calvert and King had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Drs Calvert and King are cochairs of the SPIRIT-PRO Group.
Concept and design: All authors.
Acquisition, analysis, or interpretation of data: All authors.
Drafting of the manuscript: Calvert, Kyte, Mercieca-Bebber, Slade, Chan, King.
Critical revision of the manuscript for important intellectual content: All authors.
Statistical analysis: Slade.
Obtained funding: Calvert, Kyte, King, Altman, Blazeby, Brown, Brundage, Coast, Draper, von Hildebrand, Ives, Mercieca-Bebber, Price, Roberts, Slade.
Supervision: Calvert, Kyte, Mercieca-Bebber, Slade, King.

Conflict of Interest Disclosures: All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Drs Calvert, Kyte, King, Altman, Blazeby, Brown, Brundage, Coast, Draper, von Hildebrand, Mercieca-Bebber, Price, Roberts, and Slade report receipt of a grant from Macmillan Cancer Support. Dr Calvert reports receipt of personal fees from Astellas and grants from the NIHR. Dr Velikova reports that she was past president of ISOQOL, was past chair of the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Group, and has received consultancy fees from Roche, Eisai, and Novartis related to her clinical work in breast cancer and grants from the NIHR, Yorkshire Cancer Research, Breast Cancer Now, and the EORTC Quality of Life Group. Dr Scott reports participation in the SPIRIT-PRO Delphi panel and consensus meeting as part of her employment with Janssen-Cilag UK. She also holds Johnson & Johnson stock. Dr Golub reports receipt of stipends for teaching from Northwestern University Feinberg School of Medicine. Dr Kyte reports receipt of grants from the NIHR. No other disclosures were reported.

Funding/Support: This work was funded by Macmillan Cancer Support (grant 5592105) and the University of Birmingham and was sponsored by the University of Birmingham. Drs Calvert, Kyte, and Slade are funded by the NIHR Birmingham Biomedical Research Centre and the NIHR Surgical Reconstruction and Microbiology Research Centre. Dr King is supported by the Australian government through Cancer Australia. Dr Blazeby is partially supported by the MRC ConDuCT-II Hub for Trials Methodology Research.

Role of the Funder/Sponsor: The study funders/sponsors had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Disclaimer: The views expressed in this publication are those of the authors and not necessarily those of the National Health Service, the NIHR, or the Department of Health. Views of authors, Delphi participants, and stakeholder participants are individual views and may not represent the views of the broader stakeholder group or host institution. This work reflects the views of the authors and should not be construed to represent the views or policies of the US Food and Drug Administration. Dr Golub, a *JAMA* deputy editor, was not involved in the review of or decision to publish this article.

Additional Contributions: The SPIRIT-PRO Group gratefully acknowledge the additional contributions (eAppendix in Supplement 1) made by the SPIRIT-PRO Executive, the ISOQOL Best Practices for PROs in Randomized Clinical Trials Protocol Checklist Taskforce, the international stakeholders responsible for stakeholder survey distribution and stakeholders who completed the stakeholder survey, the Delphi panelists, the SPIRIT-PRO International Consensus Meeting Participants, and Anita Walker, University of Birmingham, for administrative support.

REFERENCES

1. Chan A-W, Tetzlaff JM, Altman DG, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. *Ann Intern Med*. 2013;158(3):200-207.

2. Chan A-W, Tetzlaff JM, Gøtzsche PC, et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. *BMJ*. 2013;346:e7586.

3. Doward LC, Gnanasakthy A, Baker MG. Patient reported outcomes: looking beyond the label claim. *Health Qual Life Outcomes*. 2010;8(1):89.

4. US Food and Drug Administration. *Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims*. Silver Spring, MD: US Food and Drug Administration; 2009.

5. European Medicines Agency Committee for Medicinal Products for Human Use. *Appendix 2 to the Guideline on the Evaluation of Anticancer Medicinal Products in Man: The Use of Patient-Reported Outcome (PRO) Measures in Oncology Studies EMA/CHMP/292464/2014*. London, England: European Medicines Agency; 2016.

6. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *JAMA*. 2003;290(12):1624-1632.

7. Kyte D, Duffy H, Fletcher B, et al. Systematic evaluation of the patient-reported outcome (PRO) content of clinical trial protocols. *PLoS One*. 2014;9(10):e110229.

8. Kyte D, Retzer A, Keeley T, et al. Systematic evaluation of patient-reported outcome (PRO) protocol content and reporting in cancer clinical trials: the EPIC study. *Qual Life Res*. 2017;26:55.

9. Mercieca-Bebber R, Friedlander M, Kok P-S, et al. The patient-reported outcome content of international ovarian cancer randomised controlled trial protocols. *Qual Life Res*. 2016;25(10):2457-2465.

10. Kyte D, Ives J, Draper H, Keeley T, Calvert M. Inconsistencies in quality of life data collection in clinical trials: a potential source of bias? interviews

with research nurses and trialists. *PLoS One*. 2013;8(10):e76625.

11. Mercieca-Bebber R, Calvert M, Kyte D, Stockler M, King MT. The administration of patient-reported outcome questionnaires in cancer trials: interviews with trial coordinators regarding their roles, experiences, challenges and training. *Contemp Clin Trials Commun*. 2018;9(suppl C):23-32.

12. Kyte D, Ives J, Draper H, Calvert M. Current practices in patient-reported outcome (PRO) data collection in clinical trials: a cross-sectional survey of UK trial staff and management. *BMJ Open*. 2016;6(10):e012281.

13. Calvert M, Kyte D, von Hildebrand M, King M, Moher D. Putting patients at the heart of health-care research. *Lancet*. 2015;385(9973):1073-1074.

14. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. *PLoS Med*. 2010;7(2):e1000217.

15. Calvert M, Kyte D, Duffy H, et al. Patient-reported outcome (PRO) assessment in clinical trials: a systematic review of guidance for trial protocol writers. *PLoS One*. 2014;9(10):e110216.

16. Centre for Patient Reported Outcome Research. International consensus guidance on the inclusion of patient-reported outcomes (PROs) in clinical trial protocols: the SPIRIT-PRO Extension. 2017.https://www.birmingham.ac.uk/research/activity/mds/centres/cpror/research/spirit-pro.aspx. Accessed August 21, 2017.

17. Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD; CONSORT PRO Group. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO Extension. *JAMA*. 2013;309(8):814-822.

18. Kluetz PG, Chingos DT, Basch EM, Mitchell SA. Patient-reported outcomes in cancer clinical trials: measuring symptomatic adverse events with the National Cancer Institute's Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). *Am Soc Clin Oncol Educ Book*. 2016;35:67-73.

19. Fayers PM, Hopwood P, Harvey A, Girling DJ, Machin D, Stephens R; MRC Cancer Trials Office. Quality of life assessment in clinical trials—guidelines and a checklist for protocol writers: the UK Medical Research Council experience. *Eur J Cancer*. 1997;33(1):20-28.

20. SPIRIT Initiative. http://www.spirit-statement.org. Accessed June 23, 2017.

21. Staniszewska S, Brett J, Simera I, et al. GRIPP2 reporting checklists: tools to improve reporting of patient and public involvement in research. *BMJ*. 2017;358:j3453.

22. Mercieca-Bebber R, Palmer MJ, Brundage M, Calvert M, Stockler MR, King MT. Design, implementation and reporting strategies to reduce the instance and impact of missing patient-reported outcome (PRO) data: a systematic review. *BMJ Open*. 2016;6(6):e010938.

23. Sloan JA, Dueck AC, Erickson PA, Guess H, Revicki DA, Santanello NC; Mayo/FDA Patient-Reported Outcomes Consensus Meeting Group. Analysis and interpretation of results based on patient-reported outcomes. *Value Health*. 2007;10(2)(suppl 2):S106-S115.

24. Leidy NK, Revicki DA, Genesté B. Recommendations for evaluating the validity of quality of life claims for labeling and promotion. *Value Health*. 1999;2(2):113-127.

25. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol*. 2008;61(2):102-109.

26. King MT. A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res*. 2011;11(2):171-184.

27. Rothrock NE, Kaiser KA, Cella D. Developing a valid patient-reported outcome measure. *Clin Pharmacol Ther*. 2011;90(5):737-742.

28. Coons SJ, Gwaltney CJ, Hays RD, et al; ISPOR ePRO Task Force. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. *Value Health*. 2009;12(4):419-429.

29. Rutherford C, Costa D, Mercieca-Bebber R, Rice H, Gabb L, King M. Mode of administration does not cause bias in patient-reported outcome results: a meta-analysis. *Qual Life Res*. 2016;25(3):559-574.

30. Wild D, Grove A, Martin M, et al; ISPOR Task Force for Translation and Cultural Adaptation. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value Health*. 2005;8(2):94-104.

31. Bernhard J, Cella DF, Coates AS, et al. Missing quality of life data in cancer clinical trials: serious problems and challenges. *Stat Med*. 1998;17(5-7):517-532.

32. Fairclough DL. *Design and Analysis of Quality of Life Studies in Clinical Trials*. 2nd ed. Boca Raton, FL: Chapman & Hall/CRC; 2010.

33. Kyte D, Ives J, Draper H, Calvert M. Management of patient-reported outcome (PRO) alerts in clinical trials: a cross sectional survey. *PLoS One*. 2016;11(1):e0144658.

34. Kyte D, Draper H, Calvert M. Patient-reported outcome alerts: ethical and logistical considerations in clinical trials. *JAMA*. 2013;310(12):1229-1230.

35. University of Birmingham. PROlearn. https://www.birmingham.ac.uk/prolearn. Accessed June 23, 2017.

36. Williamson PR, Altman DG, Blazeby JM, et al. Developing core outcome sets for clinical trials: issues to consider. *Trials*. 2012;13:132.

37. Prinsen CA, Vohra S, Rose MR, et al. How to select outcome measurement instruments for outcomes included in a "core outcome set"—a practical guideline. *Trials*. 2016;17(1):449.

38. Bottomley A, Pe M, Sloan J, et al; Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints Data Consortium. Analysing data from patient-reported outcome and quality of life endpoints for cancer clinical trials: a start in setting international standards. *Lancet Oncol*. 2016;17(11):e510-e514.

39. Haywood K, Lyddiatt A, Brace-McDonnell SJ, Staniszewska S, Salek S. Establishing the values for patient engagement (PE) in health-related quality of life (HRQoL) research: an international, multiple-stakeholder perspective. *Qual Life Res*. 2017;26(6):1393-1404.

40. Frank L, Basch E, Selby JV; Patient-Centered Outcomes Research Institute. The PCORI perspective on patient-centered outcomes research. *JAMA*. 2014;312(15):1513-1514.

BMJ Open

Study protocol for the validation of a new pictorial Functional Scale in patients with knee osteoarthritis: the Functional Activity Scoring Tool (FAST)

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2023-076947.R1
Article Type:	Protocol
Date Submitted by the Author:	21-Nov-2023
Complete List of Authors:	Tang, Zhi Yin; SingHealth Polyclinics, Department of Physiotherapy Ng, Khim Siong; SingHealth Polyclinics, Department of Physiotherapy Koh, Yi Ling Eileen; SingHealth Polyclinics Yeung, Meredith T.; Singapore Institute of Technology, Health and Social Sciences Cluster
Primary Subject Heading:	Rehabilitation medicine
Secondary Subject Heading:	Patient-centred medicine
Keywords:	Patient Reported Outcome Measures, Patient-Centered Care, REHABILITATION MEDICINE, Surveys and Questionnaires

Long Title

Study protocol for the validation of a new pictorial Functional Scale in patients with knee osteoarthritis: the Functional Activity Scoring Tool (FAST)

Short Title

Validating a new pictorial FAST scale: study protocol (53 letters with space)

Trial registration number:

NCT05590663 (ClinicalTrials.gov)

Protocol version

16-Nov-23

Funding:

This work is supported by the SingHealth Polyclinics Research Seed Fund (SHPCG-RSF22-01).

Zhi Yin Tang[1], Khim Siong Ng[1], Eileen Yi Ling Koh[2], Meredith T. Yeung[3]

[1]Department of Physiotherapy, Singhealth Polyclinics, Singapore

[2] Department of Research, SingHealth Polyclinics, Singapore

[3] Health and Social Sciences Cluster, Singapore Institute of Technology, Singapore

Zhi Yin Tang ORCID: 0000-0002-8641-8146

Khim Siong Ng ORCID: 0000-0002-0374-326X

Eileen Yi Ling Koh ORCID: 0000-0001-9596-7158

Meredith Yeung ORCID: 0000-0002-4294-4187

ZYT and KSN drafted the manuscript; EYLK and MTY critically revised the manuscript.

All authors approved the final version of the article.

Correspondence:

Tang Zhi Yin

tang.zhi.yin@singhealth.com.sg

Conflict of interest:

All authors declare that no conflicts of financial or non-financial competing interests are associated with the current study.

Acknowledgements:

The authors thank and acknowledge SingHealth Polyclinics and Singapore Institute of Technology for supporting this study.

ABSTRACT

Background

Patient-reported outcome measures (PROM) are required for patient-centred care. There are limited PROM with good psychometric properties, and limitations to any language-based scale are often constrained by the written words or numerals used. Therefore, we developed the Functional Activity Scoring tool (FAST), a self-reporting pictorial scale. The FAST measures the impact of knee osteoarthritis on essential activities of daily living (ADL) and the significant changes in the self-perceived functional status over time.

Objectives:

This study aims to: (1) develop the FAST with adaptation from the Wong-Baker FACES pain rating scale; (2) validate the FAST against the Patient-Specific Functional Scale (PSFS) and Knee Injury and Osteoarthritis Outcome Score (KOOS); (3) establish the reliability, validity and responsiveness of FAST in individuals with knee osteoarthritis.

Methods and Analysis

The prospective study protocol investigates the validity, responsiveness and reliability of FAST. The PSFS and KOOS will be gold standard comparisons. Participant recruitment will occur at four public polyclinics that offer physiotherapy outpatient services in Singapore. Onsite physiotherapists familiar with the study eligibilities will refer potential participants to the investigators after the routine physiotherapy assessment. After providing written consent, eligible participants will complete outcome measurements with the FAST, PSFS and KOOS during baseline and follow-up assessments. The Global Rating of Change (GROC) scale will determine how the participant's knee status was changed compared to the beginning of the physiotherapy intervention.

Ethics and dissemination

SingHealth-Centralised Institutional Review Board approved the study (CIRB reference number: 2022/2602). The final results will be published via scientific publication. The FAST will benefit the evaluation and management of those who suffer knee osteoarthritis regardless of English proficiency or language barriers.

(270 words)

STRENGTHS AND LIMITATIONS OF THIS STUDY

- The first pictorial patient-specific functional assessment tool - the Functional Activity Scoring Tool (FAST), is developed.
- This multi-site study will validate the novel pictorial scale created and reviewed by an expert panel comprising patients and their families, physiotherapists and family physicians.
- The proposed study aligns with international consensus standards on best practices of instrument development and validation studies—the COnsensus-based Standards for selecting health-status Measurement INstruments (COSMIN).
- Validation of patient-reported outcome measures (PROM) is an iterative process. More testing of its psychometric properties must follow to support its usefulness in patients with other musculoskeletal conditions.
- Although the study protocol will not alter the standardised physiotherapy treatment, we cannot rule out possible confounding variables that may influence the study outcomes.

INTRODUCTION

Healthcare professionals regularly assess the crucial yet trouble-functioning tasks in activities of daily living (ADL). While various condition-specific questionnaires, such as the Roland-Morris Disability Questionnaire (RMQ) [1] or Knee Injury and Osteoarthritis Outcome Score (KOOS) [2], and health status measures, such as the 36-Item Short Form Survey (SF-36) [3] or EuroQol-5D (EQ-5D) [4] exist, unfortunately, limited patient-reported outcome measures (PROM) have been established thus far, especially in the area of osteoarthritis. The application of PROM in orthopaedic is expected to increase [5]. PROM were initially created for research purposes and eventually adopted for clinical management, seeking to determine patients' perceptions of their symptoms, functional status, and health-related quality of life. PROM are frequently unfittingly referred to as "outcome measures," even though they measure health—by comparing a patient's health at different times, the care outcome received can be determined [6]. PROM provide additional 'patient-centred' data that is unique in capturing the patient's perspective on the impact of their disease or disorder and its treatment [5]. These self-reported instruments elicit information about a patient's health status directly from the patient without needing interpretation from a healthcare professional [6]. The approach of gathering patient-centred data is integral in informing clinical care and supplementing measurable clinical improvements in the patients as part of the routine practice. Well-validated PROM assessing functional outcomes is required in the era of patient-centred care for holistic management.

Few osteoarthritis-specific PROM have been developed and extensively studied. A systematic review [7] identified these PROM attempt to measure psychometric properties such as pain, mental functions and moods, physical symptoms such as stiffness and mobility, as well as function in sports and recreation with either the term or subscale level. Overall, the review findings found limited evidence of psychometric properties from these PROM. Concurrently, straightforward tools to report on self-efficacy were limited. Among these, the Patient-Specific Functional Scale (PSFS) [8] was uniquely developed to enable self-reporting of the impact of musculoskeletal conditions on essential ADL and the significant

changes in the self-perceived functional status over time. A recent systematic review [9] of the PSFS against the Consensus-based Standards for the selection of health-status Measurement Instruments (COSMIN) criteria concluded that it is an easy-to-use, reliable and responsive scale in numerous musculoskeletal conditions. The PSFS is applicable across various conditions and body regions as it allows the comparison of functional outcomes across conditions and between studies [10-12]. However, the usefulness of a language-based PROM is restricted if patients and families have limited English proficiency, which is a barrier to healthcare services that is well documented [13]. Singapore is primarily an English-speaking country. However, her geographical location, historical and cultural backgrounds greatly influence the languages used in this city-state. Our anecdotal experience suggests that older patients (especially individuals ≥ 65 years old) have difficulty understanding and accurately completing the PSFS, as English may not necessarily be their primary spoken language. Difficulty with completing forms can occur for many reasons, such as the written words and numerals not being universally understood, and problems with health literacy are common and underestimated [14-16]. Very often, informal interpretation, such as relying on a family member to translate the communication, has shown to be associated with a more significant number of errors [17]. Pictures or pictorial aids are a useful adjunct to medical information and aid the transfer and comprehension of written and spoken information [18]. One good example is using the pictorial scale to measure pain in the Wong-Baker FACES pain rating scale [19], with proven benefits such as improving adherence to medications and enhancing the understanding of instructions [20].

As such, there is a growing need for a new PROM that is simple, reliable and responsive, yet minimises the limitations of any word or language-based outcome measure that are currently in use. The Functional Activity Scoring Tool (FAST) has been developed to address this situation. The FAST is a new pictorial-scale PROM measuring function in an individual with osteoarthritis. Several aspects were considered during the conceptualisation of the instrument: the applicability to a broad range of clinical

presentations (conditions, limitations and age); simple administration; concise yet effectual for speedy medical documentation; and simple interface in electronic medical record systems.

HYPOTHESIS and AIM

The confidence of a PROM depends on the psychometric evaluation of its measurement properties, and it must be undertaken to satisfy rigorous criteria [21]. These include validity (to what extent does the instrument measure the construct it purports to measure), reliability (the degree to which measurement is free from error) and responsiveness (the ability of an outcome measure to detect change over time in the construct to be measured) [22]. The process to assess these measurement properties must be iterative and studied individually. Thus, we hypothesise that the measure of function, an additional dimension to the quality of life, is possible by the same principle. The new FAST scale can be used to measure function and difficulty in performing ADL in patients with knee osteoarthritis, in an equally valid and reliable manner as the PSFS and KOOS. We aim to provide a standardised tool for gathering and documenting patients' symptoms. With these considerations, we developed the FAST scale. This study aims to: (1) develop the FAST pictorial functional scale with adaptation from the Wong-Baker FACES pain rating scale; (2) validate the FAST against the PSFS and KOOS; (3) establish the reliability, validity and responsiveness of FAST in individuals with knee osteoarthritis.

METHOD AND ANALYSIS

Study design and setting

This study will be a prospective validation study to establish the psychometric properties of a newly developed PROM. This study is proposed under the recommendation of Basch et al. (2015) [23] methods for developing patient-reported outcome-based performance measures and uses the procedures that De Vet and colleagues [24] advocated for in developing a PROM. This approach provides evidence for developing a PROM that measures the intended context and its use as an outcome measure in clinical practice and research trials. The study will take place in four physiotherapy outpatient clinics in Singapore over 12 months.

Patient and public involvement

Patients and families from physiotherapy outpatient clinics provided input and suggestions to the FAST scale during its conceptualisation and feasibility stage. Hence, their feedback also shaped the scale design, with the pros and cons of the different versions of the FAST scale discussed with patients and/or their families who will not be recruited as study participants.

Development of the FAST

During the feasibility stage, surveys were conducted on patients, families and healthcare professionals to gather feedback on the application of the PSFS. The most prevalent verbatim was "difficulty comprehending PSFS due to its being too lengthy and the lack of pictorial aid to assist patient's comprehension of the scale." Therefore, a prototype of the FAST scale was created and reviewed by an expert panel of academics, researchers and clinicians (n=7) and a series of cognitive interviews with a purposive sample of patients older than 65 (n=12) to elicit feedback on its relevance, clarity and acceptability. The final version of FAST was developed after three revisions. Figure 1 presents the conceptualisation and revision process of the FAST development. The final version of the FAST scale from this revision process will be used to test for reliability and validity in this study protocol. It consists

of a pictorial diagram with seven expression faces corresponding to an 11-point Likert scale, with Face 1 (the saddest expression) on the left of the scale paired with a score "0" and a verbal description of "unable to perform", Face 4 (neutral expression) to agree with score "5" with a description of "moderately difficult" and Face 7 (the happiest expression) on the right of the scale matching with score "10" and a descriptor of "able to perform like before". The red "cross" on the left and green "tick" on the right accentuate the effects of the facial expression and association with the verbal descriptors.

(Insert figure 1 here)

Sample size

The size of the retest sample was estimated based on a method developed to calculate the required number of participants in a reliability study [25]. The probability of type I and type II error were $\alpha = 0.05$ and $\beta = 0.20$, respectively. An interclass correlation coefficient (ICC) value of less than 0.50 indicated poor reliability, whereas values between 0.50 and 0.75 indicated fair to good reliability; an ICC value greater than 0.9 showed excellent reliability [26]. We hypothesised that our findings would be consistent with a minimum coefficient of 0.75. This level of reliability is at least appropriate for person-level comparisons. Following these assumptions, a minimum of 50 participants will be necessary for the test-retest analysis for this study. According to COSMIN guidelines, validity calculations are considered good-excellent if the sample size exceeds 100 (n=100) [27]. To allow for a possible attrition rate of 20%, a minimum sample size of 120 will be needed.

Participant recruitment and selection criteria

Participant recruitment will occur at four public polyclinics that offer physiotherapy outpatient services in different districts of Singapore. Onsite physiotherapists familiar with the study protocol will identify eligible participants during the routine initial physiotherapy assessment. Inclusion criteria based on National Institute for Health and Care Excellence (NICE) criteria [28] will be: individuals diagnosed with knee osteoarthritis and referred for physiotherapy care at the polyclinics; age 45 years and above;

and proficient in colloquial/conversational English. Potential participants will be excluded if there are additional underlying medical or trauma condition(s) of the knees (e.g., trauma, fracture, infection, inflammatory disease, tumour), history of knee surgery within the last three months, or clinically recognisable cognitive impairment that inhibits the comprehension and completion of the questionnaires. Participation in the study is strictly voluntary and will not impact the type or quality of the individual's physiotherapy treatments based on prevailing evidence.

Instruments

The self-administered KOOS is a knee-specific instrument developed to assess the patients' opinions about their knees and associated short and long-term problems [2]. It is a validated tool in Singapore for knee osteoarthritis patients [29]. It consists of 42 items in 5 subscales, i.e., pain (9 questions), symptoms (7 questions), activities in daily living (17 questions), sport and recreation function (4 questions), and knee-related quality of life (4 questions). The 5-point Likert scale scoring system ranges from “0” (no problems) to “2” (moderate problem) to “4” (Extreme problem), and the score for each domain is calculated by summing the questions. Scores will be converted to a 0 to 100 scale, with zero representing extreme knee problems and 100 representing no knee problems. The use of the 0 to 100 score is practical as it projects a direct reference to the percentage concept [2].

The PSFS is a self-reported, patient-specific measure that assesses patients' functional status [8]. Patients are asked to identify three activities most affected by their conditions and then rate their ability on an 11-point Likert 0 to 10 scale for each activity, where “0” is unable to perform the activity, and “10” being able to perform the activity at the same level as before the onset of symptoms. The total score is computed by dividing the sum of the activity scores by the number of activities listed.

The global rating of change scale (GROC) is an outcome measure that assesses patients' self-perception of change in their condition between sessions [30]. The GROC is quantified on a 15-point Likert scale from "-7” (a very great deal worse) to “0” (about the same) to “7” (a very great deal better).

The scale is easy to administer as it requires minimal skills or training, has good reproducibility and is sensitive to changes [31].

Procedure

Eligible individuals will be informed of the study's purpose and data collection procedures. Written informed consent will be obtained from every participant before data collection commences. Participants' confidentiality and anonymity will be maintained throughout the study process with a unique identifier, and only the study researchers will have access to the data. Participants will receive standardised care, and their participation status will not be shared with the attending physiotherapists apart from the initial identifications for eligibility. All data collection forms will be coded with the same unique identifier, and the study team will not retain any identifiable information. Only anonymised data will be used for data analysis. The project investigators will perform all data collection. Demographic data, clinical characteristics and primary outcome measurements with the FAST, PSFS and KOOS will be collected during baseline assessment (week 0). Follow-up assessment with FAST, PSFS, and KOOS will be scheduled two to three weeks post initial assessment together with the administration of the GROC to evaluate the efficacy of the standard physiotherapy treatment that the participants will be receiving regardless of the participation status in this study. With reference to a prior study [10], the two to three weeks period is chosen as it is also the typical duration between the initial and follow-up physiotherapy session in the local setting. Figure 2 depicts the workflow of the data collection procedures. This study will not require any alteration or deviation from the standard protocol for knee osteoarthritis physiotherapy management.

(Insert figure 2 here)

Statistical analyses

All statistical analysis will be conducted using IBM SPSS 29.0 with the statistical significance set as $p < 0.05$. Descriptive statistics will be used to describe the demographic variables using mean and standard

deviation (SD) or median and interquartile range (IQR) for continuous variables and frequencies and percentages for categorical variables. To determine the profile of the subjects with the FAST scoring, Mann-Whitney U test or Kruskal Wallis test can be used for the continuous FAST score and the categorical demographics (i.e. gender, ethnicity, marital status, education level), while the Spearman's correlation can be used to compare the continuous FAST and demographics (i.e. age).

Validation

Face validity

The qualitative methods used to determine the face validity of FAST involved face-to-face meetings with an expert panel of academics, researchers and clinicians (n=7), and a series of cognitive interviews with patients (n=12). Three essential criteria were determined in establishing face validity: Clarity (the extent to which an item is open to more than one possible interpretation), Relevancy (the extent to which an item will be relevant to its component), and Acceptability (the extent to which readers would easily understand an item).

Content validity

The content validity index (CVI) and context validity ratio (CVR) will determine the content validity [32]. CVI is the most widely reported method for determining content validity in instrument development and assessing its relevance and clarity. There are two methods of calculation, namely item-CVI (I-CVI) and scale level-CVI (S-CVI) [33]. This study will use a 4-point Likert scale "1" = unacceptable, "2" = needs some revision, "3" = needs minor revision and "4" = acceptable, for the calculation of the I-CVI from the total rating scores from all panel members. Where I-CVI is greater than 0.79, the item is acceptable; between 0.70 to 0.79, the item will require revision; and when it is less than 0.70, the item will be eliminated [34]. Similarly, the S-CVI will be determined by the number of items in an instrument that receives a "highly acceptable" grade. The Universal Agreement (UA) among the panel members (S-CVI/UA) and the Average CVI (S-CVI/Ave) are two ways of determining S-CVI [33]. S-

CVI/UA will be calculated by the sum of all items with I-CVI equal to 1 divided by the total number of items, and S-CVI/Ave is equal to the sum of all the I-CVI divided by the number of items. Content validity is excellent when the S-CVI/UA is more than 0.8 and the S-CVI/Ave is more than 0.90 [34].

CVR quantify the essentiality of an item [35]. CVR ranges from -1 to 1; a higher score represents a greater agreement between panel members. CVR = (Ne – N/2)/(N/2), where Ne is the number of panel members who rated an item as “essential", and N is the total number of panel members [33]. Each element of the FAST scale will be evaluated on a 3-point Likert scale (1 = not essential, 2 = useful but not essential, 3 = essential).

Criterion Validity

The KOOS Singapore English version and PSFS will serve as the criterion for disability in the knee osteoarthritis population. The two validated self-administered questionnaires are specific and sensitive to change over time. The correlations between the FAST, KOOS and PSFS will assess the criterion validity of the FAST scale. Spearman correlation will investigate the criterion validity against PSFS, KOOS and GROC and the measurement of agreement according to the following criteria: high (rho ≥ 0.60); moderate (rho < 0.60 - ≥ 0.30); or low (rho < 0.30).[36] The higher the rho, the higher the agreement between the two instruments.

Responsiveness

Responsiveness is defined as the ability to measure and recognize change when a change has occurred. Similarly, Spearman’s correlation coefficients (rho) can be used to determine strong (rho ≥ 0.60), moderate (rho < 0.60 - ≥ 0.30), or weak (rho < 0.30) correlations [36].

Reliability

The test-retest reliability of FAST will be calculated via Intraclass Correlation Coefficient (ICC) for absolute agreement using a two-way mixed-effect analysis of the variance model between the scores

of two stable assessment periods (i.e. global rating of change less than 3). ICC values > 0.75 are indicative of good-excellent reliability [26]. Participants who scored between -3 and +3 on the GROC were included in the test-retest analysis and are assumed that they did not demonstrate any clinically relevant changes during this interval period [37].

Cronbach alpha measures the internal consistency of the instrument, a value of more than 0.7 is considered to be acceptable. For good internal consistency, the value should be >0.8 and for excellent internal consistency, the value should be >0.9 [38].

Measurement errors were determined by calculating the standard error of measurement (SEM) and the minimal detectable change (MDC). MDC is calculated using the formula $MDC = z \approx \sqrt{2} \times \sqrt{MSE}$, where z = 1.64 and is the score associated with a 90% confidence interval, $\sqrt{2}$ reflects the uncertainty introduced by using scores at 2 different points in time, and the square root of the mean square error (MSE) term represents the standard error of measurement (SEM) [39]. The MSE was found by constructing a 1-way analysis of variance (ANOVA) table of the baseline and follow-up scores of the stable group [39].

Ethics and dissemination

The SingHealth Centralised Institutional Review Board (CIRB) approved this research protocol: (CIRB reference number: 2022/2602). There are no potential risks for participants taking part in this study. All participants will provide written consent to participate and have the right to withdraw from participation in the project at any time without any compromise or disadvantage to them in any form. All participants will be assigned a unique de-identified code to protect the confidentiality of the participants. Access to the data is restricted to the project investigators, and only anonymised data will be used during data analysis. All investigators declare no financial or other competing interests at all study sites. This study will validate the new pictorial functional scale (FAST) in patients with knee osteoarthritis and hope to investigate if the new scale correlates with similar existing PROM with good validity and reliability.

The final results and establishment of the new PROM will be published via scientific publication. This will be advantageous to healthcare professionals in evaluating functional status changes in individuals with osteoarthritis regardless of English proficiency or language barriers.

Trial status

The study is at its pilot trial stage at the time of submission of this study protocol.

(3213 words)

References

1. Roland M, Morris R. A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. *Spine (Phila Pa 1976)* 1983;8(2):141-4. doi: 10.1097/00007632-198303000-00004

2. Roos EM, Roos HP, Lohmander LS, et al. Knee Injury and Osteoarthritis Outcome Score (KOOS)—Development of a Self-Administered Outcome Measure. *Journal of Orthopaedic & Sports Physical Therapy* 1998;28(2):88-96. doi: 10.2519/jospt.1998.28.2.88

3. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30(6):473-83.

4. EuroQol-Group. EuroQol--a new facility for the measurement of health-related quality of life. *Health Policy* 1990;16(3):199-208. doi: 10.1016/0168-8510(90)90421-9

5. Kyte DG, Calvert M, van der Wees PJ, et al. An introduction to patient-reported outcome measures (PROMs) in physiotherapy. *Physiotherapy* 2015;101(2):119-25. doi: 10.1016/j.physio.2014.11.003 [published Online First: 20141128]

6. Black N. Patient reported outcome measures could help transform healthcare. *Bmj* 2013;346:f167. doi: 10.1136/bmj.f167 [published Online First: 20130128]

7. Lundgren-Nilsson Å, Dencker A, Palstam A, et al. Patient-reported outcome measures in osteoarthritis: a systematic search and review of their use and psychometric properties. *RMD Open* 2018;4(2):e000715. doi: 10.1136/rmdopen-2018-000715

8. Stratford P, Gill C, Westaway M, et al. Assessing Disability and Change on Individual Patients: A Report of a Patient Specific Measure. *Physiotherapy Canada* 1995;47(4):258-63. doi: 10.3138/ptc.47.4.258

9. Pathak A, Wilson R, Sharma S, et al. Measurement Properties of the Patient-Specific Functional Scale and Its Current Uses: An Updated Systematic Review of 57 Studies Using COSMIN Guidelines. *Journal of Orthopaedic & Sports Physical Therapy* 2022;52(5):262-75. doi: 10.2519/jospt.2022.10727

10. Chatman AB, Hyams SP, Neel JM, et al. The Patient-Specific Functional Scale: Measurement Properties in Patients With Knee Dysfunction. *Physical Therapy* 1997;77(8):820-29. doi: 10.1093/ptj/77.8.820

11. Hefford C, Abbott JH, Arnold R, et al. The Patient-Specific Functional Scale: Validity, Reliability, and Responsiveness in Patients With Upper Extremity Musculoskeletal Problems. *Journal of Orthopaedic & Sports Physical Therapy* 2012;42(2):56-65. doi: 10.2519/jospt.2012.3953

12. Westaway MD, Stratford PW, Binkley JM. The Patient-Specific Functional Scale: Validation of Its Use in Persons With Neck Dysfunction. *Journal of Orthopaedic & Sports Physical Therapy* 1998;27(5):331-38. doi: 10.2519/jospt.1998.27.5.331

13. Espinoza J, Derrington S. How Should Clinicians Respond to Language Barriers That Exacerbate Health Inequity? *AMA Journal of Ethics* 2021;23(2):E109-16. doi: 10.1001/amajethics.2021.109

14. Williams MV, Parker RM, Baker DW, et al. Inadequate Functional Health Literacy Among Patients at Two Public Hospitals. *JAMA* 1995;274(21):1677-82. doi: 10.1001/jama.1995.03530210031026

15. Baker DW, Gazmararian JA, Sudano J, et al. The Association Between Age and Health Literacy Among Elderly Persons. *The Journals of Gerontology: Series B* 2000;55(6):S368-S74. doi: 10.1093/geronb/55.6.S368

16. Williams MV, Davis T, Parker RM, et al. The role of health literacy in patient-physician communication. *Fam Med* 2002;34(5):383-9.

17. Whitaker KL, Krystallidou D, Williams ED, et al. Addressing language as a barrier to healthcare access and quality. *British Journal of General Practice* 2022;72(714):4-5. doi: 10.3399/bjgp22x718013

18. Ghiassi R, Murphy K, Cummin AR, et al. Developing a pictorial Epworth Sleepiness Scale. *Thorax* 2011;66(2):97-100. doi: 10.1136/thx.2010.136879

19. Garra G, Singer AJ, Taira BR, et al. Validation of the Wong-Baker FACES Pain Rating Scale in Pediatric Emergency Department Patients. *Academic Emergency Medicine* 2010;17(1):50-54. doi: 10.1111/j.1553-2712.2009.00620.x

20. Houts PS, Doak CC, Doak LG, et al. The role of pictures in improving health communication: A review of research on attention, comprehension, recall, and adherence. *Patient Education and Counseling* 2006;61(2):173-90. doi: 10.1016/j.pec.2005.05.004

21. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of Life Research* 2010;19(4):539-49. doi: 10.1007/s11136-010-9606-8

22. Mokkink LB, De Vet HCW, Prinsen CAC, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Quality of Life Research* 2018;27(5):1171-79. doi: 10.1007/s11136-017-1765-4

23. Basch E, Spertus J, Adams Dudley R, et al. Methods for Developing Patient-Reported Outcome-Based Performance Measures (PRO-PMs). *Value in Health* 2015;18(4):493-504. doi: 10.1016/j.jval.2015.02.018

24. de Vet HCW, Terwee CB, Mokkink LB, et al. Measurement in Medicine: A Practical Guide. Cambridge: Cambridge University Press 2011.

25. Bonett DG. Sample size requirements for estimating intraclass correlations with desired precision. *Stat Med* 2002;21(9):1331-5. doi: 10.1002/sim.1108

26. Portney LG, Watkins MP. Foundations of Clinical Research: Applications to Practice. Upper Saddle River (NJ): Pearson/Prentice Hall; 2009.

27. Gagnier JJ, Lai J, Mokkink LB, et al. COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. *Quality of Life Research* 2021;30(8):2197-218. doi: 10.1007/s11136-021-02822-4

28. Skou ST, Koes BW, Grønne DT, Young J, Roos EM. Comparison of three sets of clinical classification criteria for knee osteoarthritis: a cross-sectional study of 13,459 patients treated in primary care. *Osteoarthritis and Cartilage* 2020;28(2):167-72.

29. Xie F, Li S-C, Roos EM, et al. Cross-cultural adaptation and validation of Singapore English and Chinese versions of the Knee injury and Osteoarthritis Outcome Score (KOOS) in Asians with knee osteoarthritis in Singapore. *Osteoarthritis and Cartilage* 2006;14(11):1098-103. doi: 10.1016/j.joca.2006.05.005

30. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: Ascertaining the minimal clinically important difference. *Controlled Clinical Trials* 1989;10(4):407-15. doi: 10.1016/0197-2456(89)90005-6

31. Kamper SJ, Maher CG, Mackay G. Global Rating of Change Scales: A Review of Strengths and Weaknesses and Considerations for Design. *Journal of Manual & Manipulative Therapy* 2009;17(3):163-70. doi: 10.1179/jmt.2009.17.3.163

32. Waltz CF, Bausell BR. Nursing research: design statistics and computer analysis: Davis Fa 1981.

33. Zamanzadeh V, Ghahramanian A, Rassouli M, et al. Design and Implementation Content Validity Study: Development of an instrument for measuring Patient-Centered Communication. *Journal of Caring Sciences* 2015;4(2):165-78. doi: 10.15171/jcs.2015.017

34. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Res Nurs Health* 2007;30(4):459-67. doi: 10.1002/nur.20199

35. Yamada J, Stevens B, Sidani S, et al. Content Validity of a Process Evaluation Checklist to Measure Intervention Implementation Fidelity of the EPIC Intervention. *Worldviews on Evidence-Based Nursing* 2010;7(3):158-64. doi: 10.1111/j.1741-6787.2010.00182.x

36. Hinkle DE, Wiersma W, Jurs SG. Applied statistics for the behavioral sciences. Boston (MA): Houghton Mifflin; 2003.

37. Lim HH, Tang ZY, Hashim MA, Yang M, Koh EY, Koh KH. Cross-cultural adaptation, reliability, validity, and responsiveness of the simplified-Chinese version of neck disability index. *Spine* 2020;45(8):541. doi: 10.1097/BRS.0000000000003325

38. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951;16(3):297-334. doi: 10.1007/BF02310555

39. Schmitt JS, Di Fabio RP. Reliable change and minimum important difference (MID) proportions facilitated group responsiveness comparisons using individual threshold criteria. *J Clin Epidemiol* 2004;57:1008-1018. doi: 10.1016/j.jclinepi.2004.02.007

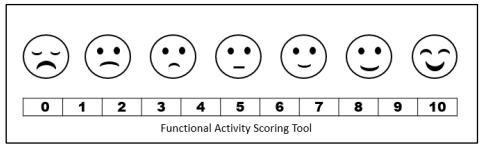
Figure legends

Figure 1. Functional Activity Scoring Tool (FAST) Conceptualisation process: Versions & Revisions

Figure 2. Workflow of the data collection procedures

Original: Version 1

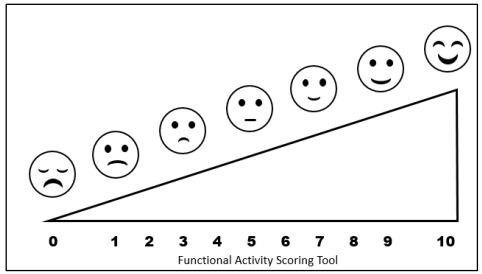
- Inspired by pictographs and Wong-Baker scale



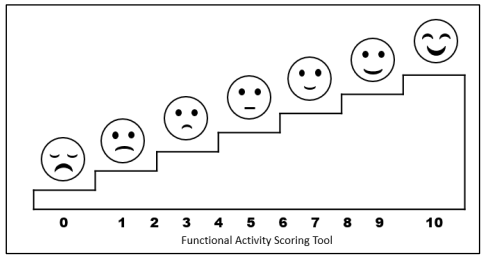
First Revision: Version 2a & 2b

- The expert review panel recommended delineation of 0 and 10 as stand-alone categories to imply 'inability to perform function' and 'no difficulty to perform function', respectively.
- Corresponding faces to score:
  - Face 1: Score 0
  - Face 2: Score 1
  - Face 3: Score 3
  - Face 4: Score 5
  - Face 5: Score 7
  - Face 6: Score 9
  - Face 7: Score 10
- Pictorial enhancement using a gradient to show decreasing difficulty. Options include slope and step ladder.

Version 2a



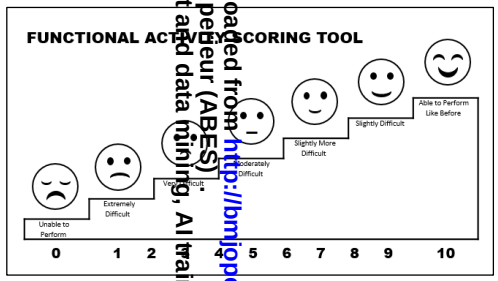
Version 2b



*Version 2b (Step ladder) was preferred by the majority (6 out of 7; 85%)

Second Revision: Version 3

- Cognitive interviews with elderly patients using version 2b.
- Common feedback was received to add some wording to improve the clarity of pictorial aids.
- Developers added short descriptors, i.e. "unable to perform, extremely difficult, etc." and redesigned the font display "Functional Activity Scoring Tool" as the questionnaire title after discussion with the expert panel.
- Corresponding faces to score is determined as below:
  - Face 1: Score 0
  - Face 2: Score 1
  - Face 3: Score 3
  - Face 4: Score 5
  - Face 5: Score 7
  - Face 6: Score 9
  - Face 7: Score 10
  - Scores 2, 4, 6, and 8 are for patients who indicate the level of difficulty between two faces.



Third Revision: Finalised Version

- Cognitive interviews with elderly patients using version 3.
- Common feedback received to add short instructions, statements or questions to improve the relevance of use.
- One patient suggested adding a tick and X at both ends to improve the efficiency of use.
- After discussion with the expert panel, visual enhancement with eye-catching symbols, green tick and red X at both ends were included, and a simple question, "How difficult is it to perform your activity?" was added.