PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (**http://bmjopen.bmj.com/site/about/resources/checklist.pdf**) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Study protocol for the validation of a new pictorial Functional Scale in patients with knee osteoarthritis: the Functional Activity Scoring Tool (FAST)
AUTHORS	Tang, Zhi Yin; Ng, Khim Siong; Koh, Yi Ling Eileen; Yeung, Meredith T.

VERSION 1 – REVIEW

REVIEWER	Pirayeh, Nahid Ahvaz Jundishapur University of Medical Sciences
REVIEW RETURNED	14-Jul-2023

GENERAL COMMENTS	Thank you for the opportunity to review this manuscript. The authors have presented a paper describing their Study protocol on the validation of a new pictorial Functional Scale (the Functional Activity Scoring Tool) in patients with knee osteoarthritis. The Introduction is well structured and the purpose of the study is clear. Also, the method is well-designed and consistent with COSMIN guidelines. Some specific comments and suggestions are included below.
	Although the researchers intend to assess the responsiveness of the FAST, the descriptions of this psychometric property are poorly written. For example: Examining the responsiveness is not mentioned in the objectives of the abstract and, also pg. 8, line 27. This sentence in the abstract is incorrect: The Global Rating of Change (GROC) scale will determine the perceived change in knee osteoarthritis.
	It is correct: The global rating of change scale will determine 'How the patient's knee status was changed compared to the beginning of the physiotherapy intervention?'
	Pg. 8, Lines 19-20: The authors write that "The new FAST scale can be used to measure function and difficulty in performing ADL in patients with osteoarthritis". The word of the knee is added to osteoarthritis.
	The inclusion criteria should be based on criteria set by the American College of Rheumatology.
	Pg.12 in procedure section: subheading "Responsiveness" is added.
	Pg. 12: Why is a period of two to three weeks considered to assess the responsiveness of the FAST? This treatment period seems to be short.
	Pg. 13 line 53: Assessing the correlations between the FAST, KOOS, and PSFS refer to construct validity, not criterion validity. How do you administer the test-retest reliability of the FAST? Given that patients receive treatment after the initial evaluation, when was the retest taken? As patients receive treatment, their conditions change, how do you administer this?

| | Pg.14 lines 33-36: why do you use the Mann-Whitney U test or Spearman's correlation for descriptive statistics?
Pg. 14 In the section statistical analysis: do you calculate Standard Error of Measurement (SEM) and Minimum Detectable Change (MDC) to assess absolute reliability?
Pg. 14 in the section Statistical analysis: Statistical analysis of responsiveness is not mentioned at all. |

REVIEWER	Bacardit, Jaume Newcastle University, School of Computing
REVIEW RETURNED	22-Sep-2023

GENERAL COMMENTS	This is to me a very clearly written paper describing a well-designed study protocol with a very good motivation. As my expertise is on data science, my only concern is on the described statistical analysis which follows a very classic and simple univariate statistical association approach. The use of multi-variate approaches, especially machine learning ones, is nowadays very widespread for the analysis of all kinds of biomedical data including Osteoarthritis. Hence it would be worth that the analysis plans include an assessment of the application of machine learning to analyse the FAST scores.

REVIEWER	Yue, Shouwei Qilu Hospital of Shandong University
REVIEW RETURNED	30-Sep-2023

GENERAL COMMENTS	The study protocol is well designed. However, considering the age of onset of the knee osteoarthritis, it is recommended that the age range of the study subjects should be expanded.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1:
Comment 1:
Thank you for the opportunity to review this manuscript. The authors have presented a paper describing their Study protocol on the validation of a new pictorial Functional Scale (the Functional Activity Scoring Tool) in patients with knee osteoarthritis. The Introduction is well structured and the purpose of the study is clear. Also, the method is well-designed and consistent with COSMIN guidelines. Some specific comments and suggestions are included below.

Response: Thank you very much for the encouragement.

Comment 2:
Although the researchers intend to assess the responsiveness of the FAST, the descriptions of this psychometric property are poorly written. For example: Examining the responsiveness is not mentioned in the objectives of the abstract and, also pg. 8, line 27.
This sentence in the abstract is incorrect: The Global Rating of Change (GROC) scale will determine the perceived change in knee osteoarthritis.
It is correct: The global rating of change scale will determine 'How the patient's knee status was changed compared to the beginning of the physiotherapy intervention?'

Response:
Thanks for highlighting. Revision was made according to your suggestion.

Examining the responsiveness of FAST has been added to the objectives of the abstract in pg. 3, line 28 and pg. 8, line 32.

The following sentence has been edited in pg. 3, lines 36-40: "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."

Comment 3:
Pg. 8, Lines 19-20: The authors write that "The new FAST scale can be used to measure function and difficulty in performing ADL in patients with osteoarthritis". The word of the knee is added to osteoarthritis. The inclusion criteria should be based on criteria set by the American College of Rheumatology.

Response:
Thanks for the highlight. We have revised the word insertion in pg. 8, lines 19-20.
Our inclusion criteria are based on the National Institute for Health and Care Excellence (NICE) criteria. According to Skou et al.1, the European League Against Rheumatism (EULAR) and American College of Rheumatology (ACR) criteria seem less appropriate to identify knee OA in primary care, as they only identified approximately half of treated patients, with no difference in those with or without self-reported radiographic knee OA. Conversely, the National Institute for Health and Care Excellence (NICE) criteria identified nine out of 10 treated patients and seem to be appropriate classification criteria for use in primary care.

Comment 4:
Pg.12 in procedure section: subheading "Responsiveness" is added.

Response:
We thank the reviewer for the recommendation. Necessary amendments have been made in pg. 14, lines 39-47.

Comment 5:
Pg. 12: Why is a period of two to three weeks considered to assess the responsiveness of the FAST? This treatment period seems to be short.

Response:
In our clinical setting, two to three weeks is the usual period between the initial and review physiotherapy sessions. Furthermore, our chosen interval duration is consistent with those of previous studies.2,3

Comment 6:
Pg. 13 line 53: Assessing the correlations between the FAST, KOOS, and PSFS refer to construct validity, not criterion validity.

Response:
We are measuring criterion validity, as our intention is to compare results of the FAST versus the gold standard measurement (KOOS or PSFS), and if the correlation is high, it shows that FAST is measuring what it intends to measure. Criterion validity is also part of construct validity, but the latter is usually evaluated using factor analysis whereby we measure how well the multiple items of the test instrument measures the concept that is not directly measurable, in this case, because FAST is only one rating question, there is no need to assess using factor analysis.

3

Comment 7:
How do you administer the test-retest reliability of the FAST? Given that patients receive treatment after the initial evaluation, when was the retest taken? As patients receive treatment, their conditions change, how do you administer this?

Response:
The retest was taken two to three weeks after the initial evaluation. Even though patients received treatment after the first test, only those that score a Global Rating of Change (GROC) of -3 to 3 are included in the test-retest analysis as they are considered to be stable and should not present any clinically large changes in the measurement. This approach is consistent with other studies.[4]

Comment 8:
Pg.14 lines 33-36: why do you use the Mann-Whitney U test or Spearman's correlation for descriptive statistics?

Response:
We have added in detailed description (pg. 13 lines 5-13) under statistical analyses for the use of these two tests: "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)."

Comment 9:
Pg. 14 In the section statistical analysis: do you calculate Standard Error of Measurement (SEM) and Minimum Detectable Change (MDC) to assess absolute reliability?

Response:
Thanks for the highlight. Yes, these two are added in to assess reliability (pg. 15, lines 19-31).

Comment 10:
Pg. 14 in the section Statistical analysis: Statistical analysis of responsiveness is not mentioned at all.

Response:
Thanks for highlighting. We have added in the assessment of responsiveness (pg. 14, lines 39-47).

Reviewer 2:

Comment 1:
This is to me a very clearly written paper describing a well-designed study protocol with a very good motivation. As my expertise is on data science, my only concern is on the described statistical analysis which follows a very classic and simple univariate statistical association approach. The use of multi-variate approaches, especially machine learning ones, is nowadays very widespread for the analysis of all kinds of biomedical data including Osteoarthritis. Hence it would be worth that the analysis plans include an assessment of the application of machine learning to analyse the FAST scores.

Response:
Thank you very much for the encouragement. This study is meant to be an instrument validation study, not a study on prediction model with multiple variables. We have not come across any published validation study using machine learning methods so far, if possible, we would appreciate if the reviewer can provide some examples for our learning as well.

Reviewer 3:

The study protocol is well designed. However, considering the age of onset of the knee osteoarthritis, it is recommended that the age range of the study subjects should be expanded.

Response:

Thanks for the recommendation. However, our inclusion criteria are based on the National Institute for Health and Care Excellence (NICE) guidelines. According to the criteria from NICE, patients can be diagnosed with knee OA if they are 45 years or older, have movement-related joint pain and either no morning knee stiffness or stiffness of 30 min or less.1

We hope that our responses adequately address the reviewers' concerns. We believe our manuscript has been improved and is suitable for publication in BMJ Open. We look forward to hearing from you regarding our submission.

Sincerely,
Zhi Yin, Khim Siong, Eileen Koh and Meredith Yeung

References
1. 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 Feb 1;28(2):167-72.
2. Yilmaz OO, Senocak O, Sahin E, Baydar M, Gulbahar S, Bircan C, Alper S. Efficacy of EMG-biofeedback in knee osteoarthritis. Rheumatology international. 2010 May;30:887-92.
3. Chatman AB, Hyams SP, Neel JM, Binkley JM, Stratford PW, Schomberg A, Stabler M. The Patient-Specific Functional Scale: measurement properties in patients with knee dysfunction. Physical therapy. 1997 Aug 1;77(8):820-9.
4. 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

VERSION 2 – REVIEW

REVIEWER	Pirayeh, Nahid Ahvaz Jundishapur University of Medical Sciences
REVIEW RETURNED	29-Nov-2023

GENERAL COMMENTS	I am glad to read this article again. I appreciate the author's efforts in this article. It was revised as well and the defect of the article were resolved. In my opinion, the article is accepted.