Comment on “Psychometrical properties of the Turkish translation of the New Knee Society Scoring System”

Mahmut Enes Kayaalp

Department of Orthopaedics and Traumatology, Cerrahpaşa University-Cerrahpaşa School of Medicine, Istanbul, Turkey

ARTICLE INFO
Article history:
Received 29 April 2019
Accepted 16 May 2020
ORCID ID of the author:
M.E.K. 0000-0002-9545-7454.

Ozden et al. translated and adapted the new Knee Society Score into Turkish and determined some psychometrical properties. They have conducted a valuable study but with some major shortcomings (1).

Firstly, the authors state that they used recommendations by Beaton et al., although they don’t mention a pretest version in their study design; but only a pilot version exclusively in the discussion section. Beaton’s recommendations involve evaluating the prefinal version on at least 30-40 patients, which is of significant importance for the final shaping of the scoring tool (2).

Secondly, although sample size recommendations in the literature are inconsistent, and there is a lack of guidance regarding validation of scoring tools; the number of items in a questionnaire should be considered or at least 100 patients are needed for internal consistency, whereas 50 patients are required to effectively evaluate floor and ceiling effects (3, 4). On the other hand, authors included only 69 patients in total, i.e. from preoperative and postoperative patients groups in total. In detail, it is further explained that 11 of these patients, who were operated bilaterally, were included in preoperative as well as postoperative groups. This gives an actual total of 58 individuals, which is quite low for validating a 34-item questionnaire (3). Authors use of “number of patients” expression in tables indicates; however, number of knees included in total, which is very deceiving. Regarding validation studies, the number of knees is of a lesser interest but the actual number of patients because this is a study of psychometrical properties as stated in the title.

Thirdly, the new KSS has two forms to evaluate patients either in the preoperative or postoperative period with a different expectations section. Therefore, it is necessary to evaluate both forms in its respective aim, i.e. preoperative patients should be compared to corresponding preoperative scoring tool results and postoperative to its corresponding results. However, authors mixed both patient groups and stated only mean values from both groups and only than they have compared it with other scoring tools, i.e. they gave mean values of mixed preoperative and postoperative scores. This method is unique, but wrong in its methodology due to the reasons explained above. Furthermore, this approach makes it impossible to compare data from this study with any of the other studies out in the literature; because there are either exclusively preoperative or postoperative results in every other study, as it should normally be. The methodology in this study is therefore questionable and less creditable compared to other studies. Besides the methodological errors, there are also some issues with the obtained results. Firstly, the physical function domain of SF- 36 should have ended in a positive correlation rather than a negative one when compared to the corresponding symptom domain of the new KSS. Furthermore, the functional activities domain of the new KSS shows almost no meaningful correlation with any of the corresponding functional domains of any other scoring tools, neither WOMAC nor KOOS or SF-36, which makes the cultural adaptation and successful translation process of this study questionable. Furthermore, although the authors claim that their results are “in conformity” with the literature, referenced French valida-
tion study includes major methodological mistakes that were addressed in a letter to the editor (5). Besides, the Japanese version does not have a reasonable correlation regarding the symptoms domain of the new KSS, because the authors failed to address issues mentioned in a recent study (4). The authors included only one center in their study, which limits the representativeness of the cohort used in this study. Even though the center would be a regional reference institution regarding the uniquely rich cultural and ethnic variabilities in Turkey, a multicenter study should have been considered.

Additionally, authors claim that their version is also aimed at Turkish speaking people living abroad, namely the European Union. This claim, however, requires further investigation because of the cultural settings. It is called external validity. Without investigating the external validity, this claim is not justified.

Last but not least, the psychometrical property is a term used to cover the validation of a scoring tool. Responsiveness is an important and relevant factor in this regard. However, it is not evaluated in this study. Therefore, it must have been inevitably mentioned in the discussion section since scoring tools are meant to reflect the effect of the performed intervention, and responsiveness is one of the major components of the psychometrical properties of a scoring tool.

Conclusion section, unfortunately, includes statements not investigated in this study. Because of the reasons explained above, a correction must be made to this study, whereas also a well-designed and executed new study is needed to validate the Turkish version of the new KSS for preoperative as well as postoperative versions.

Conflict of Interest: The author has no conflicts of interest to declare.

Financial Disclosure: The author declared that this study has received no financial support.

References

3. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007; 60: 34-42. [Crossref]

Author's response:

Dear Editor,

First of all, we thank the authors for their interest in our paper.

In our study, we used the procedures by Beaton et. al. and Guillemin et al. The recommendation mentioned in both studies already stated the necessity of making a pretest. It contains enough information to demonstrate that we are doing a pilot study. As a matter of fact, we also mentioned our pilot study in the discussion section. Beaton et al. recommended at least 30–40 patients. Guillemin did not show a minimum requirement for the number of cases taken for the pretest and developed a score to indicate that this was related to the level of evidence of the study. It would be true if the author’s critique was that we only considered the advice of Beaton et al.

We stated in our paper that the sample size was based on the general recommendations of Altman; at least 50 subjects in a methods comparison study. There are numerous qualified reliability and validity studies conducted in accordance with this recommendation. According to another recommendation study, validation studies should be conducted with at least 5-10 times the number of survey items. However, considering the long structure of the New Knee Society Scoring System, it is not easy to reach these numbers in limited time periods of cross-sectional studies conducted in one center. It is seen that these numbers couldn’t be reached in German and French version studies too. It will be more accurate to state that the lower number of cases of these studies is a minor limitation of the study.

The user manual of the questionnaire emphasizes the necessity of using a different questionnaire for each knee in bilateral cases. Therefore, in a study involving individuals assessed for both knees, giving statistical results depending on the number of individuals will not be mathematically reasonable in some situations. We explained the reason and examples of the criticism you made about the number of cases just above.
Only the expectations section was analyzed separately because of differences in the preoperative and postoperative questionnaires. There were no difference in the other sub-sections and therefore this issue did not cause problem for statistical analysis.

The version studies of the New Knee Society Scoring System have been carried out with different methodologies and it is not possible to claim that they are wrong. In the other version studies, there is no consensus in terms of methodology and the comparison of the data (especially due to the use of different surveys for validity) was not compatible.

We agree, physical function domain of SF-36 should have ended in a positive correlation rather than a negative one when compared to the corresponding symptom domain of the new KSS. However, the results obtained from our study were in this direction. So, there was a negative correlation. This may suggest that the questionnaire differs from SF-36 because of its specificity to pathology. In the User’s guide of the questionnaire the expression "Part of the validation process involved confirmation that it was generally consistent with other” knee-specific scores, “ support our results. Furthermore, the fact that the questionnaire has a broad spectrum of functional activity sub-scales with four different levels indicates that it may be natural to be incompatible with KOOS and WOMAC subscales. The comparison of a questionnaire developed for the evaluation of preoperative and postoperative periods of total knee arthroplasty with a questionnaire developed for Knee Osteoarthritis should not lead to questioning the validity and reliability of the study.

In discussion section of our study we clearly stated some relationships between other version studies. Addressed paper (Letter to the editor) and reply to this comment showed that subjective interpretations are a product of different perspectives. Also, our expression “in conformity” with the literature does not have to show the conformity of all parameters. Although the idea that multicenter study may be more effective in terms of increasing the level of evidence, it will be more accurate to consider with the advantages and disadvantages of its applicability. Advantages of multicenter study is already expressed in the literature. But, cost and time issues limits lots of studies to do so.

We have claimed that the questionnaire can be useful for Turkish-speaking individuals living abroad. Investigating the external validity in different countries could be better. Considering that the culture of each country is different, we think that it will be very hard to overcome this manner. It is just a concern that only one high quality version will be produced in one country for one given language will be more suitable, in our opinion.

Responsiveness analysis is a methodological choice in these studies. There is no rule that all studies are required to perform all sort of validation procedures. For a 34-item questionnaire specifically developed for Total Knee Arthroplasty, reliability analysis with Cronbach’s alpha value is much more important to identify the internal consistency. It would be appropriate to observe validity by comparing it with more than one questionnaire that is neglected in other studies. In conclusion, we reject neither revision nor any corrections on account of the replies above, but some issues like not conducting responsiveness analyses could be added as a suggestion for further studies.

Fatih Özdén*
On behalf of authors

* Department of Physiotherapy and Rehabilitation, Muğla Sıtkı Koçman University, School of Health Sciences, Muğla, Turkey

E-mail address: fatihozden90@gmail.com

DOI: 10.5152/j.aott.2020.04.01