Reviewer's report

Title: Determining Utility Values in Patients with Anterior Cruciate Ligament Tears using Clinical Scoring Systems

Version: 2 Date: 8 April 2011

Reviewer: Stefano Lupparelli

Reviewer's report:

Please find herewith enclosed my second peer-reviewing comments that have been pasted into the reply text the authors have kindly provided following my first peer-review. All my current comments should be regarded as major compulsory revisions when this applies to my point-to-point responses.

"Determining Utility Values in Patients with Anterior Cruciate Ligament Tears using Clinical Scoring Systems," authored by Mazda Farshad, Christian Gerber, Thomas Szucs and Dominik C Meyer

Note from the authors:

The authors thank the editor and reviewer for their thorough review of our manuscript and the opportunity to revise the text. We have directly addressed the reviewer’s comments on a point-by-point basis (enumerated below) and hope to meet his expectations. Thank you again for investing your time in considering our work.

Author Response to Reviewer Stefano Lupparelli

General Comments: The authors have addressed an important topic, which is determining utility values in patients with an anterior cruciate ligament tear. Incorporating such information into decision-making would doubtless enhance our understanding of the impact an ACL tear may have on patients. The following observations would improve the quality of their submitted article if the authors could consider a major compulsory revision.

Author response: The authors appreciate the reviewer’s view finding the topic and concept to be of interest. We have done our best to address the questions and comments you raised below.

Reviewer response. Thank you for your appreciation.

BACKGROUND:

Comment 1. The authors correctly report that a reasonable outcome following an ACL tear can be expected either by conservative or surgical treatment. This is indeed the case so much so that a distinction between “copers” and “non copers” has been made in the literature. However, it should be also clarified to the
readers that the indications for surgery depends on a number of variables including patients' level of activity and their willingness to modify it, response to rehabilitation, instability episodes despite activity modifications and rehabilitation. In other words, assigning patients with an ACL tear to either a conservative or surgical intervention is not based on an equivalent spectrum of symptoms and physical findings. An important point is also represented by the role a surgeon may play in indicating surgery versus a conservative treatment.

Author response: We fully agree. We have enhanced the paragraph with the aspects the reviewer as raised and think that this has enforced the background section of the manuscript. It now reads:

“A decision towards one or the other treatment option is challenging and should include assessment of individual factors, such as time since injury, patients level of activity, amount of instability episodes despite activity modifications, patients compliance for postoperative rehabilitation programs or performance of physiotherapy as a cornerstone of the conservative approach. However, beside individual factors, the decision should also consider economic and public health aspects.”

Reviewer response. Thank you for agreeing, the amendment appears to be satisfactory

Comment 2. Information about the impact of an ACL tear on quality of life as assessed by the SF-36 is indeed reported in the literature. Admittedly, no information regarding utilities is available. However, assessing patients' utilities is a slightly different concept also using different assessment tools and the readers would benefit from a concise explanation about what utilities are.

Author response: Thank you for pointing out this aspect. We agree and have included a description of the concept of utilities with an according reference. It now reads: “Utilities, as understood in health economics, are values that reflect an individual’s preferences for different health outcomes measured on a scale with one reflecting a state of perfect health and zero being the state of death*. The scores are used to generate quality-adjusted life years (QALY’s) for use of cost-utility analysis.”


Reviewer response. Thank you for agreeing, the amendment appears to be satisfactory. I have noticed that the concept of utilities being instrumental to generating QALYs does not crop up anywhere in the manuscript. Do the authors think that it could be useful to foresee their study could have any relevant role in encouraging ACL-related cost-utility analysis the future ?.

Comment 3. The authors succinctly detail the IKDC, Lysholm and Tegner score in this section. My suggestion is that the information relevant to these scoring systems might be more suitably placed in the “Methods” section.
Author response: We agree. Done as suggested. Thank you.

Reviewer response. Thank you for agreeing, the amendment appears to be satisfactory

Comment 4. The authors should clearly state both the design of the study and the null and test hypothesis of the study they have conducted.

Author response: Since this survey based study aims to find utility values for the most commonly used knee scores, formulating a hypothesis was more challenging than for an traditional study. However, we added both the study design and a hypothesis at the end of the “background” section of the manuscript. It now reads:

“The aim of this survey study was to correlate the most commonly used scores for the evaluation of patients with ACL injury, namely the IKDC score, the Lysholm score and the Tegner score, not only between each other but also to the quality of life, to verify the ability of the scores to discriminate among different situations and to provide information for the transformation of reported outcomes to utility values. The hypothesis is that the most commonly used scores for the assessment of patients with a ruptured ACL and utility values as an indicator of quality of life are correlated.”

Reviewer response. The authors are right in that a survey is not usually associated with a null/test hypothesis formulation. Do they think their study could be regarded as a cross-sectional study conducted on a sample of orthopaedic surgeons as proxies?

METHODS:

Comment 1. The authors use the term “survey” in the “Methods” section even if they have employed the term study (admittedly without specifying the type of design) in the “Background” section. This should be corrected as already suggested. To my best knowledge, surveys are less likely to be associated with hypothesis testing. Conversely, it would seem that the underlying aim of the study might be to find out whether the IKDC, Lysholm and Tegner scoring systems correlate with patients’ utilities and whether these can discriminate the scenarios proposed by the authors.

Author response: See answer to comment 4 in the Background section. We appreciate the reviewers understanding that hypothesis testing can include surveys as methodological tools.

Reviewer response. I am not entirely sure the study cannot be regarded as a cross-sectional study instead of a survey, please see my previous response

Comment 2. My understanding of the study methodology is that four scenarios (or vignettes) characterizing different level of activities of patients with an ACL tear were created according to the method validated by Gottlob. Then a study
sample of orthopaedic surgeons (used as proxies in lieu of real patients) were asked to fill in the IKDC, Lysholm and Tegner scores as well as the utility scores. The authors finally performed their correlation analyses, which is the main focus of the study. If my understanding is right, it would then be useful to give further information in this section about the number of orthopaedic surgeons constituting the study sample, whether they work in the same institution, whether they share the same indications for surgery or conservative treatment. Also, the authors state that “the opinions of the orthopaedic surgeons who had treated more than 100 ACL patients...”. It would be important to report whether these 100 ACL patients are treated monthly or annually, for example. Finally, the authors should report what is the percentage of surgeons treating more than 100 ACL patients and 300 patients within the study sample. It would also be interesting to know what criterion they used to weigh as double and triple the opinion of surgeons respectively operating more than 100 and 300 ACL patients.

Author response: We agree that the sample that was asked to fill in the scores should be described more precisely. We have addressed the suggested specifications: the according sections now read:

“Four hypothetical vignettes (Table 1) of patients all with an isolated rupture of their anterior cruciate ligament on one knee and differently active were presented to 27 orthopedic surgeons of the same institution.”

Reviewer response. Could the authors please specify if the orthopaedic surgeons participating in the study all had the same level of seniority (e.g. consultant vs registrars ?)

“The opinions of the orthopedic surgeons who had treated more than 100 ACL patients in their career (n=9) were counted as twofold and those from experts who had taken care of more than 300 ACL patients (n=2) were counted threefold.”

We have chosen this approximation (<100 patients 1x, >100 patients 2x, >300 patients 3x) based on the assumption that the opinion of more experienced samples should not be set equal to those with a lower level of experience. We found however no evidence or guideline on how to weigh this phenomenon.

Reviewer response. I would suggest it would be useful if the authors could specify in the text that their approximation was based on an arbitrary assumption as there is no evidence or guideline to weigh an experience-related opinion

Comment 3. As previously stated, a succinct but informative description of the scoring systems, including the Health Utility Index, used in the study should be incorporated in this section.

Author response: Done as per suggestion. An additional reference with information about the Health Utility Index has been added accordingly:

Reviewer response. Thank you, this appears to be satisfactory.

Comment 4. One important issue is whether the orthopaedic surgeons used as proxies regarded patients as described in the scenarios as patients capable of living with their impairment or whether they completed the scoring systems having a potential surgical candidates in mind. Not differentiating between the two types of patients risks introducing a selection bias: surgical candidates and non-surgical ones could be characterized by different utilities. In other words, using, for example, physiotherapists as proxies could yield totally different results.

Author response: We do understand this concern and had explicitly not distinguished the hypothetical patients as candidates for surgical or conservative treatment. We believe that this is not a selection bias confounding the results, since the surgeons assessment of the performance in the scoring of the typical hypothetical patient was independent of the fact whether the patient would need or had surgery or not. Further, the survey was performed in setting of an academic institution with no relevant rewards for surgical over conservative treatment, as both are performed there daily.

Reviewer response. Please allow me to disagree as we do not know whether patients undergoing surgery for an ACL tear may score their utilities differently from those who benefit from physiotherapy alone. Thus, I would suggest the authors could include their response, which appears to rule out the risk of a selection bias, in the text. My observation was strictly methodological and did not relate to any potential “reward” or academic versus non-academic setting being acting as confounders.

Comment 5. It would seem that the authors used the same age (35 year-old) to set the four hypothetical patient scenarios. This would represent a major limitation to their analysis as it has been demonstrated that quality of life as measured by the SF-36 in ACL patients is affected by the patients’ age, gender and associated comorbidities. Could this apply to utilities as well? We actually do not know but, should this be the case, the findings of their studies could not be extrapolated to patients who are younger or older than 35 years of age.

Author response: This is clearly a limitation of the study. We had chosen this age (35 year old) based on the Swiss national statistics and our own statistics that showed that 35y is the mean age of patients presenting with an ACL injury. We agree that the question whether age, gender and other cofactors might affect the utility values of a patient with ACL injury is very interesting. The influence of age as a patient characteristic is however difficult and complex to grasp and reflected in many other, in part competing aspects: Younger age may simultaneously mean a tendency towards higher activity, higher ability to compensate for a ligamentous deficit, better healing, less weight, higher expectations, a more physical strenuous profession, etc. However, it was not the aim of this study to answer this question and as stated at the end of the manuscript, we suggest that
a measurement instrument for utility values should be included in future prospective clinical studies in addition to commonly used knee scoring systems to confirm and adjust the here presented data and to answer the interesting question that the reviewer has raised. More importantly, we also feel that the current standardization to a defined type of patient does not relevantly influence the comparison of the different scoring systems, which was a major aim of this work.

Reviewer response. I wonder whether the authors could concisely incorporate the above thoughtful observations into their “Discussion” section.

Comment 6. I am not entirely sure that scoring data could be regarded as continuous values regardless of their distribution. I would have personally used non-parametric testing but I would recommend a statistician consultation.

Author response: As mentioned in the statistical description part of the “methods”, a test was applied to find out whether data represented normal distribution or not before applying parametrical or non-parametrical statistical tests. A professional statistician was involved in the analysis.

Reviewer response. I oblige the expertise of a professional statistician although I would not have used parametric tests to perform an analysis on score data regardless of their distribution. Yet, this is only my preference.

RESULTS

Comment 1. The authors might wish to organize their “results” section in a way that avoids overlapping information. Specifically, the results of correlation of utility values are reported under the subparagraph “Utility values” and then again under the subparagraph “Correlations of knee scores and utility values”.

Author response: We have re-read the result section carefully and found no overlap. Under the heading “utility values” we present total values and under the heading “correlations of knee scores to utility values” we do not repeat them, but give the correlation values.

Reviewer response. In the “Results” section and under the “Utility values” paragraph it reads (page 8 and lines 8-10 of the paragraph): “The utility values over all vignettes correlated best with the IKDC subjective score (r=0.86, p<0.001) followed by the Lysholm score (r=0.77, p<0.001) and the Tegner score (r=0.77, p<0.001)”. In the “Results” section and under the “Correlation” paragraph analysis it reads (page 9 and lines 1-2): “The scoring system, which had the strongest correlation to the utility values was the IKDC subjective score (r=0.87, p<0.05) over all vignettes (Figure 2)” and again (page 9 and lines 5-7 of the paragraph): “The overall correlation of the Lysholm and the Tegner score to the utility values was similar (r=0.77, p<0.05 for both) (Figure 2)”. I thought this might be a repetition.

Comment 2. Two sets of graphs relevant to Figure 1 and Figure 2 were supplied with the manuscript, which do not seem to differ aside from the layout. The
authors might wish to supply only one set of graphs to avoid redundancy.

Author response: While figure 1 shows the IKDC subjective knee score, Lyshom and Tegner score for different activity levels, figure 2 documents the correlation of those scores with the utility values. We have re-uploaded the figure to assure that this comment was not caused by an error in the uploading process of manuscript submission.

Reviewer response. I can see that Figure 1 and Figure 2 do convey a different information, yet I still have TWO sets of figures which would seem to differ only in the colour (black versus lilac) of the caption text.

DISCUSSION

Comment 1-There would seem to be some redundancy at the beginning of this section as the authors have repeated some of the findings reported in the “Results” section. There is little discussion of the clinical meaningfulness regarding their observations, for example what they refer to as “difference in gain” of utility values among different vignettes.

Author response: Thank you for this advises. We have removed some of the redundant sentences in the beginning of the discussion part. The most important clinical meaning in our eyes that can be gained from this study is that the correlations found here can be used to transform reported score-values of IKDC subjective, Lysholm and Tegner score to utility values by the transformation key as provided in the manuscript. This has been stated several times in the discussion part.

Reviewer response. In the “Results” section and under the “Utility values” paragraph it reads (page 8 and lines 6-8 of the paragraph): “The difference in gain of utility values from vignettes 1 to 2 (0.25± 0.15) was significantly higher than from vignettes 3 to 4 (0.13± 0.12)”. I would be interested in knowing whether the authors think the statistical difference also conveys a clinical meaning. If so, should any importance be attributed to this finding or is it just a numerical observation?

Comment 2. The potential (selection) bias and confounders (age, gender, comorbidity) should be better highlighted in this section.

Author response: We agree. See also answer to the comment. We have further enhanced this already stated limitation that now reads:

“Second, the designed hypothetical patients might not perfectly represent the range of patients seen in daily practice with a variety of potential confounders such as age, gender, comorbidities and activity level.”

Reviewer response. Thank you, this appears to be satisfactory.

Comment 3-In conclusion, I have found two major limitations:
1. The lack of a study sample including actual patients risks introducing an uncontrollable selection bias that could flaw the study.

2. The four hypothetical scenarios are limited to a fixed age that is not truly representative of the wide spectrum of patients encountered in real practice. Hence, I would strongly encourage the authors to perform a compulsory major revision including a cohort of patients that could be contrasted with the findings obtained by the authors using orthopaedic surgeons.

Author response: Thank you for these comments, which represent the two main limitations already stated in the manuscript. We agree that a very large individual level prospective study with inclusion of a measurement instrument for utility values in addition to commonly used knee scoring systems would have been ideal to answer the here asked study question. However, we believe that the here used methodology is valid and was able to provide an accurate answer to the study question with an established method and reasonable amount of used resources. As stated above, the definition of a standardized patient is an accepted and well defined approach and does avoid hard to define biases and confounders, such as for example the competing effect of age on activity and ability to compensate for functional deficits. This is a complex physiological and functional set of topics which is subject to several clinical investigations and would disproportionally complicate the present manuscript.

We have addressed the limitations of the here used approach in the discussion as:

“...the results are based on opinions of physicians on how the hypothetical patient in average would perform and not on real patient data. Although the approach of assuming values based on expert’s opinion is frequently used in health economic research [36-38], it might bear potential disadvantages; two experts might not agree fully on their opinions. We tried to reduce the potential of discrepancy of opinions by including a large number of experts. Further, experts might not represent accurately their patients as surrogates. However, we believe that for production of a transformation key from knee scores to the utility values, such as performed here, using real patient data might bear too much uncontrollable confounders and that standardized hypothetic patients might better represent the cohort of all ACL patients.”

and

“Second, the designed hypothetical patients might not perfectly represent the range of patients seen in daily practice with a variety of potential confounders such as age, gender, comorbidities and activity level.”

A cohort of patients that could be set in contrast to the findings obtained here would need to be very large in order to be valid and there are practical limitations to perform such a survey: Patients would need to be sampled at the same institution for an impractically long period (we calculated over ten years KTÖRRELL) or at a very large number of institutions. Both approaches
would doubtlessly increase variability of data due to the large number of investigators with nonuniform care for detail and the inherent bias of where and when the patients are recruited. Although such an approach is currently planned in a prospective project, it is not possible to include such a study in this manuscript revision. Although enhancement of the current study with such data would increase the value of the manuscript, we feel that the current results are valuable, transparent and based on an established and well accepted method.

Reviewer response. I do acknowledge that the authors have discussed the limitations of their study owing to the lack of a cohort of actual patients. I also appreciate that their methodology is sound and derived from previous economic research.

However, it is my view that clinical research should improve our knowledge in order to achieve higher standards of care. The authors have above stated that “we believe that for production of a transformation key from knee scores to the utility values, such as performed here, using real patient data might bear too much uncontrollable confounders and that standardized hypothetic patients might better represent the cohort of all ACL patients.”

My main concern is therefore whether these very same challenges they seem to allow for would prevent orthopaedic surgeons from including a measurement instrument for utility values in addition to commonly used knee scoring systems as they advocate in their conclusions.

Could the authors comment on this and perhaps incorporating their views into their conclusions?

Finally, do they think that their study would bear implications to calculate QALYs at a time of limited healthcare resources when applied to

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

'I declare that I have no competing interests'