Author's response to reviews

Title: Prioritisation of patients on waiting lists for hip and knee arthroplasties and cataract surgery: instruments validation

Authors:

Alejandro Allepuz (aallepuz@aatrm.catsalut.net)
Mireia Espallargues (mespallargues@aatrm.catsalut.net)
Montse Moharra (mmoharra@aatrm.catsalut.net)
MercA Comas (mcomas@imas.imim.es)
Joan MV Pons (jpons@aatrm.catsalut.net)
Network Research Group on Support Instruments IRYSS (mespallargues@aatrm.catsalut.net)

Version: 2 Date: 20 February 2008

Author's response to reviews: see over
Barcelona, 20 February, 2008

Dear Editor,

We send you back our response to each of the issues raised by the reviewers. To facilitate the reading, the reviewers’ comments appear in italic and our responses in bold. When we’ve introduced any change to the previous version of the manuscript we’ve referenced it according to the paragraph where the change has been made.

We’ve added a new file as an appendix containing two examples of hypothetical scenarios that were used in the study and that were requested by two of the reviewers. We’ve also added a new ascription. The Catalan Agency for Health Technology Assessment (CAHTA) has become member of the Biomedical Research Net (CIBER) on epidemiology and public health. We’ve included also in the acknowledgement section that the Catalan Government has recognised CAHTA as a Health Services and Outcomes Research Group.

We hope that this new version of the manuscript could be considered adequate to be published in the BMC Health Services Research journal.

Sincerely,

Alejandro Allepuz
MINNOR ESSENTIAL REVISIONS (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

There are some minor grammatical errors. Following are suggested revisions:

Methods: This was a multicentre validation study which included patients scheduled for hip and knee arthroplasties and cataract surgery between June 2001 and June 2002 and May 2004 and March 2006 in 10 hospitals of different Spanish Autonomous Communities: 2 in Andalucia, 2 in Aragon, 2 in the Canary Islands and 4 in Catalonia. (p. 5)

We accept the suggested correction and we include it in the manuscript. First paragraph in methods.

This technique has been used in health care to involve patients and the community in planning and developing healthcare services and to investigate priority of patients on waiting lists and differences on judgements among different stakeholders [18,19]. (p. 5)

We accept the suggested correction and we include it in the manuscript. First paragraph in CAHTA’s prioritisation instruments in methods.

Construct validity was evaluated only through correlation analysis due to the fact that our theoretical framework for the development of the prioritisation instruments did not contemplate a previous hypothetical structure.

We’ve changed this paragraph after another reviewer’s comments. First paragraph in Convergent and discriminant validity analysis in methods.

Although simplicity could threaten validity, a prioritisation instrument with few criteria reduces the possibility of missing information on some of them. (p. 15)

We accept the suggested correction and we include it in the manuscript. Paragraph nine in discussion.

All health professionals and managers must be involved in their implementation. It could be also necessary to supervise its application to prevent distortions in the priority assessment, and it has to be set up as an information system able to both incorporate and apply this new information.

We accept the suggested correction and we include it in the manuscript. Paragraph nine in discussion. We’ve also added some new comments after another reviewer’s suggestions.

ME designed the study and also drafted the manuscript.

We accept the suggested correction and we include it in the manuscript. We’ve also changed slightly this sentence. In authors’ contributions.
Table 2 The speech dimension was not included in the analysis as patients had difficulty with it.

We accept the suggested correction but with a different meaning: The speech dimension was not included in the analysis as none patient had difficulty with it.

DISCRETIONARY REVISIONS (which the author can choose to ignore)

Recommendations are as follows:

1. For the convergent and discriminant validity analysis, rather than comparing the overall CI and AI with each dimension or item of the EQ-5D and HUI3, it would be more meaningful to compare the specific dimensions of the CI and AI with conceptually similar items of the EQ-5D and HUI3 (e.g., pain). (page 8 and Table 2) If this is not revised, the issue should be addressed either in the methods or discussion. It is acceptable to compare the overall CI and AI with the summary scores (i.e., the EQ-5D index and HUI3.)

When the strategy of analysis was planned we considered that our main objective was to assess convergent and discriminant validity analysis only for the overall priority score and not for its criteria, as any dimension structure was defined for its development. The priority construct is defined by the presence of a number of criteria, but it’s the result of the combination of these criteria –the priority score- that we were interested in analysing.

We include a comment on this issue in “Convergent and discriminant validity analysis” in the Methods section.

2. I don’t understand the rationale for the hypothesized convergent correlation coefficients between the CI and the dimensions of mobility (EQ-5D) and ambulation (HUI3). They assess problems or difficulty with walking. (Table 2).

The rationale of this hypothesis was the difficulty that the patient might have for moving around because of his or her problems with vision. The hypotheses were formulated before analysing the data and we decided not to change them after reviewing the results.

3. It wasn’t clear how the scenarios were constructed. Did each one include all of the criteria? Maybe an example of a scenario would be helpful. (page 9)

The scenarios were based on real situations obtained through clinical chart reviews. Each scenario consisted on a summary of the main patients characteristics related to the prioritisation criteria. Two examples, one for each instrument, have been included in this new version of the manuscript.
4. Results: I assume that consecutive patients were asked to participate in the study. The response rate and the characteristics of the responders vs. non-responders should be reported. If the data are not available, then this should be included under study limitations. (p. 10). In addition, how many physicians were asked to participate and how many did participate in the study?

We completely agree with the reviewer that this information should have been included in the manuscript. The patients recruited for this study were probably not consecutive. We encouraged participating doctors to include all their patients until they had recruited the established sample size. Their workload prevented them from doing it. However, when we compared patients’ priority profiles with that from the Catalan Health Service they were similar.

Regarding doctors, we invited all the Ophthalmologists and Orthopaedic Surgeons of the selected centres to participate in the study. From those who agreed to participate not all of them finally collaborated. The participation percentage of Ophthalmologists and Orthopaedic Surgeons was 53% -19 out of 36- and 73% -40 out of 55- respectively.

We include a comment on these limitations in the sixth paragraph of the discussion section.

5. Under study limitations, another factor that could influence the correlation between the priority scores and the patient-reported HRQOL questionnaires is a difference in assessment time between the surgeon and patient assessments (upto 4 months). (p. 14)

We agree with the reviewer that during these 4 months the correlation between patients’ HRQL and the priority score might have been affected.

We include a comment on this limitation in the sixth paragraph of the discussion section.

6. Table 3: Does hip or femur fracture for the hip/knee arthroplasties mean a previous fracture? Under occupational status, please clarify how you calculated the percentages as I would assume that the ‘housewife’ category only refers to the females while the other categories refer to both.

Hip or femur fracture means a previous fracture not related to the current indication for surgery. These categories for describing patients’ comorbidities were taken from Barcelona’s Health Survey.

Occupational status percentages were calculated over the total of included patients. They sum up 100%. Not all women were housewives, part of them were retired.
REVIEWER: Lorne Bellan

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

*I do not think that the last sentence on page 7 is a proper sentence and after reading it multiple times I still can't figure out what it means. It needs to be rewritten so that its meaning is clear to the average reader.*

We agree with the reviewer and we’ve rewritten this last sentence: “We did not use factor analysis to analyse construct validity due to the fact that our theoretical framework for the development of the prioritisation instruments did not contemplate a previous hypothetical structure.”

*I have real concerns about your statements on page 15 about how you were perturbed about patients who had no symptoms or disability being placed on waiting lists and then later on about how your prioritization tool could be used to flag inadequate (I think you meant "inappropriate" the way the rest of the sentence reads) resources allocation. I have experienced the dangers of these arguments first hand in Canada where the VF-14 was used for prioritization in some provinces. The problem is that a tool developed for prioritization isn’t automatically a valid tool for measuring appropriateness of surgery. Cataract surgery is sometimes undertaken when people don’t feel they have any symptoms but have been shown to have vision that falls below the minimum standard for driving and so they will lose their driver’s license without the surgery. Many elderly patients tend to minimize their complaints and will present saying they have no problems and yet, on examination, are found to have profound visual impairment. Cataract surgery is also undertaken in some cases where patients are asymptomatic but their lens cloudiness impairs management of retinal diseases. These are only three easy examples of when using a priority tool breaks down in addressing appropriateness and there are more. The one study (Can J Ophthalmol. 2005 Aug;40(4):433-8)that I am familiar with that tried to look at the problem of using a prioritization tool with a ceiling effect for assessing appropriateness found that 75% of those study who had no potential for improvement based on their pre-operative prioritization assessment reported after their surgery that they had markedly improved. It is harmful to the profession to claim that a tool that has been shown to have construct validity for prioritization of patients once they have been booked for surgery automatically has construct validity for assessing appropriateness for cataract surgery.*

The prioritisation instruments, as the reviewer highlights, were developed to order the waiting list not to analyse appropriateness of indication. In fact, when you review the levels of each criterion, a score of 0 doesn’t mean that there are no symptoms. Anyhow, the Scientific Advisory Committee of the project where the prioritisation instruments were developed, agreed with the possibility of using these instruments to decide maximum waiting times and if the patient could be placed in the waiting list or if he or she should wait a number of months for a new priority assessment.
The Scientific Advisory Committee reviewed the different patients’ priority profiles associated with each priority score, and decided that for cataract surgery patients who scored 0 could wait 6 months, 3 months if they were between 0 and 50 and 1 month if they were between 50 and 80. In the case of hip and knee arthroplasties the cut-off point was established in a score of 60. Patients who scored below this score needed to be reassessed 6 months later and the ones over 60 could be placed in the waiting list.

We agree with the reviewer that “It is harmful to the profession to claim that a tool that has been shown to have construct validity for prioritization of patients once they have been booked for surgery automatically has construct validity for assessing appropriateness for cataract surgery.” In fact, when these instruments were presented at each of the Health Care Regions of Catalonia, this was one of the main concerns of some of the doctors.

We’ve decided to exclude this last part of the paragraph. Its interpretation in different contexts and the fact that to extend its application is necessary to reach a previous consensus and further research justify its exclusion to avoid misinterpretations.

---

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

*On page 6 line 8 shouldn't it be "logic" instead of "logit"?*

The rank-ordered logit model fits a logistic regression by maximum likelihood on the ranked alternatives. In this case the ranked alternatives were the scenarios ordered by the participants when the prioritisation instruments were developed. The coefficients of each criterion levels in the model were their final score after applying a transformation to a 0-100 point-count linear scale.

*In the last sentence on page 6 the word "from" should be deleted and the word "a" should be inserted before the word "doctor's".*

We accept the suggested corrections and we include them in the manuscript. Second paragraph in convergent and discriminant validity study methods section.

*On page 14 line 10 the word "suppose" is incorrect. I think you meant "impose".*

We’ve changed “suppose” by “represent”. We think this word explains better the idea we wanted to transmit regarding the reliability analysis limitations. Paragraph seven in the discussion section.
On Table 3 I didn’t understand whether "Previous surgery" referred to any form of surgery in the patient’s lifetime or organ specific surgery (i.e. eye surgery for the cataract patients)

We were referring to a previous arthroplasty or cataract surgery. We’ve modified the text in table 3 in order to make the meaning clearer.

Discretionary Revisions (which the author can choose to ignore)

In the methods section I wasn’t clear on the two time windows that were mentioned: June 2001-2 and May 2004-March 2006. Was one for arthroplasties and one for cataract surgery? If so, it would be clearer if you added the word "respectively" after March 2006.

The study for both hip and knee arthroplasties and cataract surgery was conducted during the 2 periods. During the first period the study was carried on in Barcelona and when an extra budget was available hospitals from other regions were invited to participate in the study.
MINOR ESSENTIAL REVISIONS (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

1. sp. p7 line 14 'New Zealand'

We’ve corrected this error. First paragraph in convergent and discriminant validity analysis methods section.

DISCRETIONARY REVISIONS (which the author can choose to ignore)

1. The correlations between VAS and the IA and IC tools were good. Could the VAS be correlated to the 7 and 6 criteria of IA and IC respectively to gain an understanding of the weighting that surgeons were placing on them?

In a previous work presented in the annual congress of the Spanish Society of Epidemiology (SEE) we analysed through a linear regression model where the weights assigned by the doctors to each criterion, but only for hip and knee arthroplasties. The dependent variable of the model was the VAS and the explanatory variables each of the priority criteria.

The criteria order according to their weights in the prioritisation instrument was the same when the criteria weights were calculated through the linear regression on VAS. However, the weights for severity of the disease, pain and difficulty in doing ADL increased. On the contrary, the weights of probability of recovery, limitation on ability to work, has someone to look after the patient and to be a care-giver were reduced and weren’t statistically significant.

Our interpretation of these results was that the method used to the prioritisation instruments development –conjoint analysis- allowed to incorporate criteria that are considered important but that are not taken into account during the daily medical practice. The different points of view that the prioritisation instrument includes –medical specialists, general population, patients and close relatives and allied professionals- it’s an added value that could explain these results.

We’ve not included any commentary in this issue as we considered that it was not related to validation, but with the interpretability of the priority score.

2. The inter-surgeon reliability was surprisingly high for this sort of clinical judgment exercise. Were the scenarios too bland in their range of variation leading to a 'damping' effect on surgeon judgment? More information on scenario construction required.
We’ve included two examples for both hip and knee arthroplasties and cataract surgery that can help to clarify how the scenarios were presented to the doctors. Some comments on this issue were included in the discussion section were the limitations are explained (seventh paragraph).

3. P13 para 2: a little dismissive of equity of access for the individual patient. This is a real ethical issue which also has legal and political implications

We agree with the reviewer that an incorrect allocation of a patient in the waiting list because of the variability in his or her priority assessment could be an ethical issue with possible legal consequences. Medical judgement is difficult to standardise. The prioritisation instruments have tried to make explicit criteria that in most cases were implicitly used. Although some sort of variability might exist in the priority assessment, once some criteria are established this variability could be reduced but not eliminated. This irreducible minimum is related to subjective judgement of each criterion. However, there is an important advantage. If a patient is incorrectly allocated in the waiting list, there are criteria that could be reassessed to evaluate if patient’s priority score was correctly assessed. Otherwise, patients’ complains could be difficult to respond.

We’ve modified paragraph four of the discussion section.

4. p15 para 2: consider perhaps the role of a ‘feedback’ audit of surgeon scoring to reduce temporal score inflation and maintain confidence by funders and users of elective surgery?

This point was discussed with the Catalan Health Service (CHS) when health care centres started to send information on priority scores. At the moment we don’t have information on this aspect, but when our study was started we compared our data with CHS data and priority scores distributions were similar.

We’ve modified a sentence of the paragraph nine of the discussion section.
This study is important as it validates an instrument for prioritization of three common diseases. The title is relevant and the study concentrates on a few psychometric properties of the instrument according to the title. The statistical analyses are relevant.

However, I have some concerns about the construct of the instrument. The prioritization instrument basically contains two sets of items related to the patients’ self-assessed situation; difficulty in doing ADL for both diseases and pain for arthroplasty patients. The severity of the disease is ranked very high, especially in cataract with 45 as maximal ranking. This may explain the moderate correlation between patients’ perception of the difficulty caused by the condition and the instrument. The dominance of objective scoring may also explain the better correlation with the doctor’s opinion. In order to make this clearer I suggest that the authors include the relative weight of each part of the instrument in the two patient groups. My guess is that the ‘objective’ parts of the instrument contribute to a high percentage of the total scoring. This will explain some of the correlations that may seem surprising for the authors. The authors are concerned about the relatively weak correlation with the doctors’ opinion though this was the strongest correlation in the study. Furthermore the authors think that this ‘weak’ correlation depends on influence by the patients’ opinion. I make the opposite conclusion. This instrument has too little scoring from the patients’ self-assessed situation.

We agree with the reviewer that the correlations between the prioritisation instruments and doctors’ opinions were strong. In fact, doctor’s opinion was our principal variable to analyse construct validity. They were the only ones who were directly asked on their opinion on priority. On the contrary, patients weren’t directly asked and they were only questioned on their perception on different aspects of their HRQOL. This limitation was explained in the discussion section. The relatives weights of each criterion of the instruments correspond to the scores in table 1.

The reviewer raises some questions on the instruments development process. We agree that the weight in the cataract surgery instrument for severity of the disease is really high. This isn’t the case in the arthroplasty instrument were severity of the disease has the third weight in importance after pain and difficulty in ADL. The weights for each criterion were calculated through a rank-ordered logit model using as the response variable the order of the priority scenarios established by the study participants - general practitioners, nurses, social workers, optometrists, physiotherapists, orthopaedic surgeons, rheumatologists, rehabilitators, ophthalmologists and general practitioners.

Prioritisation instruments include objective criteria and criteria related with patients’ self-perception. This could have affected the correlations between VAS and overall priority score. A higher variability in patients’ self-perceived criteria might be expected, although this influence would be more important for the cataract instrument – severity of the disease
maximum score is 45- than for the arthroplasties instrument – severity of the disease maximum score is 18.

We’ve introduced some comments on this issue in the first paragraph of the discussion.

*It is well-known that generic instruments show little influence by cataract; this should not be as surprising as the authors state.*

We agree with the reviewer and we’ve added to the second paragraph of the discussion section this sentence.

*The inter-observer test adds some information about the instrument, but I don’t think this test can fully replace a test in realistic clinical daily work. This is also commented by the authors. In reality patients’ influence over the doctor and the situation may change the result and increase the variation between different examiners. I suggest that the authors point out this weakness of the study more clearly.*

We agree with the reviewer and we’ve introduced some modifications in paragraph seven of the discussion section.

*I have a few questions:

1. How was the unilaterally situation handled in the item “severity of disease”. Dense cataract in one eye and normal vision in the other may not interfere too much in daily living but how was it scored? Describe this.*

In table 1, because of simplification, not all the information related to how to score cataract surgery severity of the disease criterion was shown. The score is assigned according to the visual acuity in both eyes. Here are the different combinations:

Best eye >0.4 and worst eye >=0.2 severity of the disease is mild
Best eye >0.4 and worst eye <0.2 severity of the disease is moderate
Best eye 0.2-0.4 and worst eye >=0.2 or <0.2 severity of the disease is severe
Best eye <0.2 and worst eye >=0.2 or <0.2 severity of the disease is extremely severe

We’ve added this information in table 1.

2. When ICC was used, do the values refer to average measurements or single measurements?

The values used to calculate the ICC were the individual priority scores assigned to each scenario by the participating doctors. A commentary on this issue was included in paragraph three of reliability study and analysis methods section.
3. Why did the authors not use a Rasch analyzed questionnaire? VF-14 is not very good according to the state of the art technology (Rasch analysis).

We decided to use the VF-14 because the scores provided by the Rasch analysis has shown a high level of agreement with traditional scores (ICC = 0.80). Valderas JM, Alonso J, Prieto L, Espallargues M, Castells X. Content-based interpretation aids for health-related quality of life measures in clinical practice. An example for the visual function index (VF-14). Qual Life Res. 2004 Feb;13(1):35-44.

4. Did the same doctor group severity of the disease in the instrument and assess the priority through a visual analogue scale for a certain patient in this validation test? According to the manuscript the doctor could not see the instrument during the VAS assessment, but was it the same doctor who did both evaluations?

The same doctors answered the VAS and administered the prioritisation instrument. We tried to avoid the influence of the priority criteria on doctor’s opinion answer to the VAS by placing first the VAS question and then the priority criteria. Of course, after having had few administrations the doctors could have learn which the priority criteria were. This fact was also included as a limitation in the discussion section.