Author’s response to reviews

Title: A comparison of the sensitivity of EQ-5D, SF-6D and TTO utility values to changes in vision and perceived visual function in patients with primary open-angle glaucoma

Authors:

Fiammetta M Bozzani (fiammetta.bozzani@lshtm.ac.uk)
Yasmene Alavi (yasmene.alavi@lshtm.ac.uk)
Mireia Jofre-Bonet (mireia.jofre-bonet@city.ac.uk)
Hannah Kuper (hannah.kuper@lshtm.ac.uk)

Version: 2 Date: 29 June 2012

Author’s response to reviews: see over
Dear Sirs,

This letter accompanies the submission of a revised version of our manuscript to BMC Ophthalmology. We are very grateful to the editors and reviewers for the suggestions and comments that greatly improved our work.

Changes in the revised version include moving a statement concerning funding received from Pfizer to the Competing Interests section to comply with the journal’s editorial policy. Other points raised by the reviewers are addressed in detail below.

Kind regards,

Fiammetta Bozzani
Fiammetta.bozzani@lshtm.ac.uk

Reviewer #1: Louise Longworth

Reviewer’s report:
The paper describes a small but interesting study designed to estimate and compare utility values for people with primary open-angle glaucoma. My main comment relates to a lack of detail provided on the methods used to elicit the utilities. This is key to the study design and results. If the methods used really are comparable (depending on responses to concerns about the TTO methods below) then I propose that the clarifications are made within the paper along with additional drafting clarifications noted below. However if the methodology used in TTO exercise affects comparability of utilities with EQ-5D and SF-6D, then this is likely to substantially weaken the paper and significant revisions would have to be made to reflect this. I have elaborated further below on the specific concerns relating to the TTO. The manuscript also requires some amendments for accuracy and clarity.

Major essential revisions
Main comment: TTO design
My main comment is that there is a lack of detail describing the TTO elicitation. This is extremely important for the paper and it is difficult to judge the methods and data without that information.

This is a good point. More detailed methods are published in a previous publication by the group (Alavi et al, 2011), referenced in the first paragraph of the methods. However, we agree that more detail is needed in this paper too. We have therefore added more detail and references about the TTO methods used in the methods section. As explained in our responses to authors’ comments below, the 3 methods for UV elicitation (SF-6D, EQ-5D, TTO) are different on several levels and this is now acknowledged as a limitation in the Discussion section.
- What were the anchors used for the TTO? This is crucial. Was it ‘dead’ and ‘perfect health’ or ‘full health’ as used for QALY estimation and EQ-5D and SF-6D. This needs to be explicitly stated. If these anchors or equivalent one were not used the comparability with EQ-5D and SF-6D will be flawed as the TTO will be measuring a different concept. In addition it would be inappropriate to compare to most QALY calculations if used to estimate QALYs.

We agree. As now described in the methods, participants were asked how many years of their remaining life they would be willing to trade for perfect vision. Therefore the anchors were perfect vision and death. As discussed in our previous paper (Alavi et al 2011), this does limit the comparability of the TTO utility values with those of the EQ-5D and the SF-6D, and their application in QALY estimation. We have added these limitations to our discussion.

- The length of time for trading was set as the total that they expected to live (Y). As a method this is fine. However further clarification required: was this asked explicitly at the outset about the individual value for each person? Why was this approach chosen? It would also be helpful to present these data. This differs from the approach used in the EQ-5D and SF-6D which use a 10 period. Consideration as to how this will affect comparability is required in the discussion section.

The amount of years participants expected to live was indeed asked explicitly to each individual from the outset. The reason for this is that we followed the method used and published many times by Gary and Melissa Brown, and Sharma, all now referenced in the methods section. At the time this study was designed, their health economics papers dominated the scientific literature in the field of ophthalmology. Therefore we based our questionnaire on theirs (shown below).


I’m going to ask you some purely theoretical questions that require careful thought. Please take your time in answering.
1. How many years do you expect to live?
2. Suppose there was a technology that could return your eyesight to perfectly normal in both eyes. The technology always works but decreases your survival. Essentially, it theoretically increases your quality of life, but decreases the length of time you live. What is the maximum number of years—if any—you would be willing to give up if you could receive this technology and have perfect vision for your remaining years?

- What approach was used to elicit values for states considered to be worse than dead (SWD). I note that the authors consider zero to be a lower bound for the TTO, which is concerning. The authors need explain how they handled the possibility of SWD explicitly. Again they need to consider comparability to other utility estimation methods in their explanation and discussion.

The Brown method did not include a method for estimating states worse than death, and therefore nor did our method. This has now been acknowledged in the Discussion as a limitation.
Were props used? If so what? Props have been found to significantly affect the values obtained (see Dolan et al, JHE 15:2)

Following the published methods of Brown and colleagues, props were not used. Preferences were stated in the participants’ answers to the second part of the question.

What iteration procedure was used to obtain the point of indifference?

The point of indifference (preference) was not revealed in an iteration procedure. As described above, the maximum number of years a participant was willing to trade was asked directly as a question, and participants stated their preference in answers to this. This has now been explained in the methods section.

Other points of clarification
- SF-6D is obtained from 11 not 6 questions on the SF-36. Is this a typo or has it been incorrectly calculated?

It was indeed a typo and it has been corrected. The sentence meant to explain that 6 of the 8 dimensions, not questions, are used for generating the SF-6D index.

Minor essential revisions
- Some grand statements are referred to in the introduction without sufficient explanation or justification. For example, it states that ‘interventions should be compared to each other through cost-utility analyses’. Why should they? What about other methods such as cost-benefit analysis as generally recommended by the Treasury? I agree that CUA has a significant role in the evluation of health interventions but it is not the only approach.

Assigning a monetary value to the benefits afforded by different glaucoma interventions is complicated by the fact that the pace of disease progression varies across individuals. In the absence of a universally accepted system for staging glaucoma, as described in the manuscript, it is not possible to calculate productivity gains from different interventions to be used as an outcome measure in cost-benefit analysis. Moreover, the fact that glaucoma progression cannot be reversed to restore vision, makes it difficult to quantify effectiveness across treatment options for cost-effectiveness analysis. The lack of comparability of outcomes in economic evaluations of glaucoma interventions was pointed out in the most recent systematic review on the subject (Hernandez et al. ,J Glaucoma). As NICE advocates the use of utility elicitation methods for use as measures of effect in economic evaluations, we feel it is correct to state that CUA should be the standard analysis method for comparing glaucoma treatment options.

- Greater clarity is required in the second paragraph of the introduction as it confuses several concepts. It states that utility values are preference-based measure of quality of life. This is not correct. Utility values can be produced by preference-based measures of quality of life. They also don’t range from 0 to 1. As by definition, values of less than zero (dead) are possible to reflect health states considered to be worse than being dead. In the context of utility measurement the MID can
be considered to be the minimum detriment in health that a person would give up some length of life to avoid.

We agree. The description of utility values from different elicitation methods and their anchor points have been amended accordingly.

- TTO is a method rather than an instrument

The manuscript was checked to ensure it contains no reference to TTO as an instrument rather than a method.

- What types of conditions prevented reliable visual testing?

Patients in the advanced stages of glaucoma found it difficult to do VF tests (often their eyes would water). We have added the following detail to the methods from (Alavi et al 2011): IVF scores were based on the monocular VF for participants with (1) no perception of light (NPL) in one eye, or (2) severe visual loss (mean deviation ≤ −25dB) in one eye in their most recent test, and whose eyesight had deteriorated to the extent that their VF was unobtainable in that eye. An IVF score of 104 was assigned to those participants who had NPL/severe visual loss in both eyes, such that a reliable VF was unobtainable from either eye.

- What happened to those people who failed the depression screen?

A trained counsellor was available to talk to participant upset by any part of the interview, this has been added to the methods.

- Response rates (% people approached and completing interviews) and missing data (% item missing responses for each instrument/measure) should be reported

574 patients were contacted of which 132 (23%) were included in the study. We have not included this statistic as a response rate because it is not simply a reflection of how many chose to/not to participate. There were several other reasons people contacted did not participate. For example, some failed the inclusion/exclusion criteria based on their clinic appointment data on the day; many were in the very early stages of visual field loss, and those with more severe visual field loss were prioritised half way through recruitment, sometimes the logistics of their appointment and interview times did not work out.

As for missing data, Table 2 reports the total number of respondents with complete sets of answers, and for whom the corresponding utility values could be derived, for each elicitation method. The % of items missing responses for each instrument has now been added to the Results section.

- Discussion. The authors are correct that a 5 level version of the EQ-5D has been developed (see Herdman et al, 2011) although this wasn't developed to specifically address concerns around vision.
This was clarified in the discussion, explaining that the 5-level version of EQ-5D was introduced to address the original instrument’s poor sensitivity and ceiling effects for certain conditions. The suggested reference was also added.

Discretionary revisions

- 'UV' isn’t a commonly used abbreviation. I think it would improve readability if excessive abbreviations were avoided.

We agree that the reading would flow better without too many abbreviations but we had to use ‘UV’ for the sake of brevity.

Discussion:
- How do the results compare to other studies examining the performance of utility instruments for reflecting the utility associated with glaucoma and other vision conditions? E.g. Tosh et al, Value in Health 2012.

The discussion was expanded and the very good suggested reference was included.

- The discussion doesn’t include any consideration of why instruments such as the EQ-5D and SF-6D may have been designed in the way they have. That is, to focus on how health impacts on general quality of life and functioning (as opposed to specific symptoms), to enable multiple health impacts of conditions (and adverse effects of treatment) to be captured within a single utility estimate and to enable comparison between interventions/conditions. There will clearly be a trade off between these abilities and the to capture small/specific changes for specific symptoms. In this specific case where is that balance of trade off? Is the sensitivity more important - will that be the case for all interventions in glaucoma?

This valid point was expanded in the final paragraph of the conclusion. Here, we specified that, in the case of a degenerative condition such as POAG, what matters is the timing of the intervention. For this reason, utility elicitation methods most sensitive to small progressive QOL declines are to be preferred over those displaying the properties mentioned above.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

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

**Declaration of competing interests:**
I have also published in this field but do not consider myself to have any conflicting interests and I have no related financial interests to declare

**Reviewer # 2:** Augusto Azuara-blanco

**Reviewer’s report:**
The authors should be congratulated for this important contribution to the literature. This study will be very important for evaluating effectiveness of glaucoma interventions. Surprisingly there is very little published data about health states and utilities in glaucoma, and this study fills some major gaps.

We would like to thank the reviewer for the very encouraging comment on our work.

**Level of interest:** An exceptional article
**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

**Reviewer # 3:** Georgios Labiris

**Reviewer's report:**
In this paper, authors deal with the very contemporary issue of Quality of Life (QoL) among individuals with glaucoma and the sensitivity of QoL Questionnaires and utility values (UVs) in detecting minor deterioration of visual function. Despite, the clinical and research significance of such studies, this paper does not significantly add to the body of knowledge of this important issue.

**Major Compulsory Revisions**

Regarding primary objective as stated by the authors “…study objective is to identify among 3 widely used methods (EQ-5D, SF-6D, Time-Trade-Off) the one most sensitive changes in both binocular VFL and visual functioning (measured using the VFQ-25).” However, former studies have already indicated that generic health-related QoL instruments (i.e EQ-5D, SF-6D) inefficiently reflect vision specific QoL changes and that vision-specific (i.e. NEI-VFQ 25) and glaucoma-specific (i.e. GQL-15) tools correlate better with glaucoma-related clinical parameters than generic ones. Similar conclusions could be drawn by this research work, as well, although they are indirectly presented in the manuscript.

As correctly pointed out by the reviewer, available evidence does suggest that vision-specific tools more adequately reflect glaucoma-induced QOL changes. However, since utility values are necessary for economic evaluations of alternative treatment options, the objective we stated in the above sentence was to identify among the 3 more widely used generic elicitation methods the ones that are most sensitive to such QOL changes and to self-reported changes in visual functioning as measured by VFQ. We did not aim to compare the performance of generic health-related and vision-specific instruments directly.

We have expanded both the background and discussion sections to include further references and explanations with regards to the known shortcomings of generic health-related elicitation methods to clarify our objectives.
For example in:

- Results / Stage-dependent changes in visual functioning and QOL / line 21
  “...There was strong evidence of a relationship between VFL and every VFQ-25 sub-scale (all P<0.001), except for general health (P=0.10) and ocular pain (P=0.16). The same associations were found for VAB (P<0.001 for all tests but general health and ocular pain, both P=0.35)...” however, no table with the specific scores is included in the paper.

The median and inter-quartile ranges for changes in each VFQ-25 sub-scale by level of VFL severity are shown in Figure 1. We did not show them in a Table for the sake of brevity, as the description of these observed associations was outside the scope of our main objective. We opted for a graphical representation accompanied by the p-values in the above paragraph instead.

- Again in Results / Stage-dependent changes in visual functioning and QOL / line 29 “...Only the SF-36 sub-scales related to its physical component were associated with VFL and VAB, (P=0.03 and P=0.01, 30 respectively; P=0.4 and P=0.92 for the psychological component)...”. No specific data are displayed in a table.

As above, we opted for not showing these results in a table as investigating the associations between generic QOL dimensions and vision loss was beyond the central objectives of the study.

- Regarding the “Results/ Sensitivity of UVs to changes in visual functioning / line 11 " Spearman’s rank correlation coefficients indicate statistical dependence between the VFQ-25 composite score and the UVs from EQ-5D (r=0.38), SF-6D (r=0.43) and, in particular, TTO....", it would be to the reader’s interest if all the results mentioned in this section would be analytically displayed in an additional table.

As suggested, the correlation coefficients and P-values for the association between utilities and VFQ-25 dimensions are now displayed in Table 4. The above paragraph was modified accordingly.

Minor Essential Revisions

- Please, spell out the acronyms (i.e. EQ-5D, SF-6D etc.) in the abstract and the first time they are mentioned in text, as well.

All acronyms have now been defined both in the abstract and in the text.

- In Methods / Design and sample / Line 13 "Exclusion criteria were eye surgery in the preceding....", it is mentioned that exclusion criteria included eye surgery 6 weeks prior to the response to the QoL instruments and other significant ocular co-morbidities. Please define which ocular diseases were included.

Any ocular co-morbidities contributing to the patients’ loss of vision were excluded, as these could confound the effect of glaucoma-induced vision loss on quality of life. This has now been specified in the Methods section.
In Results / Line 5 "The VAB of most patients was within the normal range....", it is stated that the range of the binocular visual acuity (VAB) was: -0.16 - 1.85, while in Table 1 the same range is: -0.18 - 1.85. Please define the discrepancy.

The typo in the text has been corrected. The value reported in Table 1 is accurate.

Discretionary Revisions
- The second paragraph of the “CONCLUSIONS” section belongs better to the “DISCUSSION” section.

The suggested move was carefully considered, but decided against to emphasise a point-raised by reviewer n. 1 that we feel should be an important take-home message from our results i.e. TTO’s sensitivity, albeit still not perfectly adequate for detecting small QOL changes at early stages, is the property that makes it preferable to EQ-5D and SF-6D as an elicitation method for a degenerative condition such as POAG

Level of interest: An article of limited interest
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