**Author's response to reviews**

**Title:** Pre-validation methods for developing a patient reported outcome instrument.

**Authors:**

Maria E Prior (m.e.prior@abdn.ac.uk)
Jemaima Che Hamzah (j.hamzah@abdn.ac.uk)
Jill J Francis (j.francis@abdn.ac.uk)
Craig R Ramsay (c.r.ramsay@abdn.ac.uk)
May Castillo (m.castillo@abdn.ac.uk)
Susan E Campbell (Susan.Campbell@uae.ac.uk)
Augusto Azuara-Blanco (aazblanco@abdn.ac.uk)
Jennifer M Burr (j.m.burr@abdn.ac.uk)

**Version:** 2  **Date:** 26 May 2011

**Author's response to reviews:** see over
Dear Editor,

Re: Resubmission of: Pre-validation methods for developing a patient reported outcome instrument

We are pleased to re-submit the above article for consideration in BMC Medical Research Methodology. We thank the two reviewers for their thorough and scholarly critique of this paper. We have revised the title and the paper in accordance with their advice and we believe this process has resulted in a higher quality manuscript. Revisions are underlined within the main body of the text. Below, we present the comments of each reviewer, our responses to them and the page reference of the relevant text in the revised manuscript. We trust that these revisions are acceptable and look forward to your decision in due course.

Maria Prior for the Author Group
Reviewer 1

I have serious reservations about the use of the term ‘item bank’ throughout this paper. It is my belief that the authors here refer to a pool of items that they have amassed throughout literature and other means. An item bank is the end product of several important developmental and validation phases where resulting items have appropriate fit; have reasonable targeting; are well calibrated; and have undergone some form of item response theory. What the authors claim to be an item bank in the paper is simply an initial pool of items. It shows a serious misunderstanding of the appropriate terminology.

*We thank the reviewer for their opinion and acknowledge the differences between their interpretation of an item bank and ours. We have revised the manuscript accordingly.*

I also surprised that the focus of the AGQ is very narrow as the authors are only interested in functioning and disability which are similar. To generate an item bank with such a narrow focus does not justify the invested resources and time. What are the authors' plan following this?

*We propose a systematic, 5-step process for developing a new instrument for assessing patient reported outcomes where there is an existing body of relevant research and existing instruments derived from qualitative enquiry. This is the generalisable component of our article. The specific component is that we used this 5-step process to develop the Aberdeen Glaucoma Questionnaire, which is the primary patient reported outcome measure for a future randomised controlled trial (RCT) evaluating the effectiveness of glaucoma screening compared with no formal screening (opportunistic case detection). The aim of the AGQ is to compare patient reported vision related disability between the intervention (glaucoma screening) and comparator (opportunistic case detection) arms at the end of the proposed RCT. Existing instruments used in glaucoma populations that whilst validated and mostly arising from qualitative studies are not suitable in their entirety for this purpose. We wanted to build on previous research by making use of the existing body of knowledge in terms of items that are known to be relevant to people with glaucoma and to develop a new instrument from these. We have clarified this point on page 4 of the manuscript.*

Introduction

I fail to make the connection between glaucoma screening and a valid glaucoma instrument. It is one of the reasons but there are dozens more. One line is enough and other reasons also need to be mentioned. In addition the authors claim further that they want the tool to be used in clinical trials. Are the items geared towards screening or treatment and clinical trials?
The AGQ was developed to be used as the primary patient reported outcome measure for a future randomised controlled trial (RCT) evaluating the effectiveness of glaucoma screening compared with no formal screening (opportunistic case detection). The aim of the AGQ is to compare patient reported vision related disability between the intervention (glaucoma screening) and comparator (opportunistic case detection) arms at the end of the proposed RCT. We have revised the manuscript on page 4.

Methods

The authors report ‘From the resultant list of instruments, we selected those instruments that met the following criteria: suitable for self report; validated in a glaucoma population; in the public domain; items and response options fully described in the text article reporting the instrument’. Rather than using the term ‘instruments’, it should be items. The authors are seeking to generate a pool of items, as such they should focus on items not instruments.

We appreciate the distinction that the reviewer makes. Our initial focus was on instruments that met our inclusion criteria. Our pool of items consisted of all items contained in the selected instruments (step 1 of our reported method). We then applied Steps 2-5 of our method to this pool of items. We have revised the manuscript to be explicit regarding the shift in focus from instruments to items (Table 1: Step 1).

The authors argue the following: C) It is recommended that both generic and condition-specific PRO instruments be used in clinical trials [26]. The glaucoma-specific AGQ was therefore designed to be administered alongside the widely validated generic measure, SF-36, in the definitive trial [26]. The third phase of de-duplication resulted from this decision. All SF-36 items and any extra items considered, by the multi-disciplinary team, to be directly covered by SF-36 items were removed from the item bank.

What is the rationale for including generic instrument? Just because they have been recommended? How does including a generic instrument contribute to the aim of the study which is ‘(to develop a self-report measure of functioning and disability associated with glaucoma, and its treatment, for use in a clinical trial)’.

Capturing vision related disability is important, but within a RCT context it is important to compare other health effects between the screened and unscreened population to capture any wider benefits or harms of glaucoma screening. Existing generic instruments, such as the SF 36 are adequate for this
purpose and can be used alongside glaucoma specific measures. We have clarified this point in the manuscript at page 4.

I have serious reservations concerning the two strategies used by the authors to item reduce the pool. It is too ‘expert’ driven as opposed to patient driven. How about pilot testing these items with a group with glaucoma, look for inter-item correlation, poorly understood items, those with little variance, those with high floor and ceiling effects, etc. The ‘think aloud section’ is encouraging as it is the only patient involvement in this preliminary phase but it was limited to only a group of eight subjects. Clearly more is required.

The aim of this paper is to present a pre-pilot, pre-validation methodology for developing new PRO instruments from an existing body of knowledge (i.e. existing validated instruments). It articulates the phase that is usually not presented in the literature, to address the question, which items should be pilot tested with a relevant user group? This pre-validation phase includes a qualitative ‘think aloud’ study that explores acceptability of the AGQ to people with no visual impairment and with differing severity of glaucoma. We have clarified this point in the manuscript on page 3.

We agree with the reviewer that the next stage of this work is to examine the statistical attributes of the items and we have done this with a sample of 656 people with no visual impairment or with varying glaucoma severity. However, the testing of the AGQ (e.g. investigation of inter-item correlation, lack of variance, floor/ceiling effects, factor structure, item discriminability) and any subsequence item reduction is not the focus of this pre-validation (pre-quantitative) paper.

Result

I am staggered by the final figure of 68 items from an initial pool of 725.

We agree, and are pleased to have been able to reduce the items to this number although clearly this would still be too large. We are working on reducing this further to minimise participant burden.

Discussion

One of the limitations of the AGQ, even at this early stage, is its relatively low overall number of items in the pool. This will invariably shrink in the future developmental and validation phases. Yet, this is supposed to be an item bank with a range of items able to address functioning and disability across the spectrum of the condition. I think the authors have missed a tremendous opportunity here. I suspect not having done focus groups from the start, was always going to be an issue in the end.
How do the authors plan the next stages?

*Our aim is to develop an instrument that is acceptable to members of the public in terms of length, comprehensibility and appropriateness to the condition we are investigating. We refer the reviewer to our earlier response regarding pilot testing of the AGQ with 656 people with no visual impairment or with differing severity of glaucoma.*

Reviewer 2

This is a good manuscript, quite long but easy to read, matching the current standards, worth the publication, contributing to addressing the current challenge of creating new measurement instruments in the field of ophthalmology.

The methods are well described and consistent, the work is based on a reasonable set of data, clearly documented.

However, there are a couple of important points that, in the reviewer’s opinion, call for essential revisions:

The first one is about the title, as well as about the abstract introduction and conclusions: the title suggests this is a paper discussing the methods, illustrated by a specific measure in glaucoma. I think the authors should give a clear priority to their paper: if they advocate for a specific method in developing new PRO measurement instruments, the discussion on the theories (those that have been used in the past, those currently used for new instruments, and the specific approach they suggest) should be more developed and supported by more references and an accurate discussion, which is not the case here; if they essentially promote a new instrument – which seems to be the case – then the title is inadequate, as are the statements in the abstract introduction “we describe generalisable method…” and conclusions “the study addresses a gap in the literature”.

*We very much appreciate the reviewer’s advice. Our priority is to present a method for developing instruments that are fit for validation. The illustration of the method using the AGQ as an example of the method is secondary. We have taken the reviewer’s advice and have substantially revised the title and background section of the manuscript accordingly.*

The second major limitation in the work – which is related to the one mentioned above – is about the term “theory”. In qualitative research, a “theory” is frequently referring to “grounded theory”, which claims that the theoretical conceptual framework should emerge from data coming directly and essentially from the information coming from the study subjects. This is not the case here, as
the theory is pre-existing: the authors use the ICF theory as the basis for the interpretation and organization of items, rather than a theory which would have emerged from specific qualitative research with patients. We found no justification in the manuscript explaining why the authors did consider a preexisting model would be superior to a grounded theory model. Moreover, the validity of the preexisting model itself is in our opinion not well justified: why this model rather than another one; how was this model developed, why is it a good start for developing a new scale?

We agree that when reporting research methodology and presenting study findings the theoretical framework of a study should be reported, regardless of the theoretical perspective adopted. Unfortunately, studies reporting the development of instruments using qualitative methods (e.g. focus groups) rarely make reference to any underpinning theoretical framework. Instead, such studies tend to report only the methodological approach, data collection methods used and the fact that the items generated using qualitative methods represent the patients’ views of what is important.

This is not a paper about generating new theory. It does not pretend to be a synthesis of qualitative research in that sense. It is merely a synthesis of the products of such research (ie validated instruments developed from qualitative research). As such, we decided to use established theory to ensure that the item coverage represented a broad (bio-psycho-social) view of health outcomes rather than representing the medical model. We accept that there is a range of views about generating new theory versus using established theory. Given the breadth and depth of research in the field we decided to use established theory. We have clarified this point in the manuscript on pages 4 and 7.

The third major limitation is about the use of this new instruments: “screening” suggests this instrument is aimed at detecting specific patients among a more general population, who would be eligible for a specific action or decision. What is the clinical or epidemiology question that the new questionnaire is intended to address? Is it treatment, is it referral, is it communication, or anything else? If there is an intention to screen for specific patients, what is the next step: threshold, decision rule, classification… And, what is the targeted population: general population, population already showing some risk factors, population complaining about the something, population visiting specialists?

The target population is general population in both arms (intervention and control) of a proposed glaucoma screening trial (RCT). The questionnaire was developed as a patient-reported outcome measure for use, in addition to clinical outcome measures, in the trial. The aim is to assess vision
related disability, from the perspective of participants, as a primary trial outcome (i.e. in both the screened and unscreened populations). It is not designed as the screening instrument; rather, it is designed to assess whether vision related functioning and disability in people in the screening intervention group (some of whom will have been detected, explored further, diagnosed and treated) is better than people in the control group, several years after the screening has taken place. We have revised the manuscript on page 4 to address the reviewer’s comment.

The fourth limitation is in the selection of information and questionnaires used for the development of the current scale: one of the selection criteria is availability of the source questionnaires in English. How can the authors claim this is generalisable, if they did not assess the contents of other existing questionnaires developed with non English-speaking patients?

We claim only that the pre-validation methods (i.e. Steps 1 - 5) are generalisable, not the AGQ itself. We appreciate the reviewer’s point that there may be validated instruments in foreign languages; however, our past research in multi-lingual contexts has made us aware that translated instruments require re-validation in the new language context. Hence, non-English language instruments, once translated, would not meet our criterion of having been validated.

In addition, the following discretionary revisions are suggested:

Page 4: “the optimal glaucoma-specific PRO instrument evaluating a screening programme would comprise items relevant to disability associated with glaucoma and its treatment and items capturing the impact of any disability on quality of life”. Why so? How do authors justify this statement?

We have revised the manuscript on page 4 to more clearly communicate the aims of the AGQ.

Table 1: Step 3D: how do you justify that items are not relevant to the aim of the instrument?

We thank the reviewer for their comment and have revised Table 1 Step 3D.

Step 3 in text (page 8): item reduction. Step D: not really informative. Step E: how did the authors ensure reproductibility?

We thank the reviewer for their comment and have revised the manuscript on page 8.