PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	International multi-phase mixed methods study protocol to develop a cross-cultural patient-reported outcome and experience measure for hand conditions (HAND-Q)
AUTHORS	Sierakowski, Kyra; Dean, Nicola; Pusic, Andrea; Cano, Stefan; Griffin, Philip; Bain, Gregory; Lalonde, Donald; Klassen, Anne

VERSION 1 - REVIEW

REVIEWER	Edith Poku School of Health and Related Research University of Sheffield United Kingdom
REVIEW RETURNED	23-Aug-2018

GENERAL COMMENTS	Manuscript ID: bmjopen-2018-025822 This submission covers an interesting and important issue – developing a reliable and acceptable culturally-relevant patient-reported outcome measure (PROM) for a wide range of conditions affecting the hand. In general the methodology adopted for the study is sound. However, more clarity about the methods described in the submitted protocol will immensely improve the quality of the submission. General revisions • Please check syntax to improve understanding of the submission. • Kindly consider providing a clear distinction between aspects of the study which have been completed and those phases which are being planned, conducted or yet to be implemented. • It is good practice to avoid starting sentences with abbreviations. • Please clarify whether you mean ‘clinical experts’ when you refer to ‘experts’ throughout. If this is the case, please use the term ‘clinicians’ or ‘clinical experts’ more consistently. • Please minimise the use of the term ‘the modern psychometric method’. • Please use Phase I, II, II consistently. • Citations may be needed in some sections, e.g. Introduction Lines 46 to 54 • It may be useful to provide the estimated sample size for the different aspects of the study or information about how the sample sizes will be determined. Comments and suggested revisions 1. Strengths and limitations of this study: Lines 11 to 12 Please explain - Independently functioning scales will allow tailoring of scales to the patient, study or clinical setting, which will reduce patient and administrative burden 2. Strengths and limitations of this study: Lines 16 to 17

	Please justify - Use of a modern psychometric approach will produce HAND-Q scales that are both scientifically sound and clinically meaningful. 3. Introduction: Lines 28 to 29 Consider replacing 'forces' with 'factors' These forces have led to an increasing amount of hand surgery performed outside of operating theatres with a fully awake patient and… 4. Introduction: Lines 55 to 56 The authors refer to the worldwide prevalence of hand conditions without providing available statistics. Relevant information would be useful to estimate the burden of illness of major hand conditions. 5. Methods and Analysis Generally, the use of the various subheadings limits the clarity of the description of the study's methodology. The submission could improve if authors would consider the combining or deleting some headings used of this section. For example, from the submission it can be inferred that the systematic review aspect of Phase I is completed but the conceptual framework and qualitative study are yet to be conducted. This and subsequent aspect of the study could be more clearly set out. a. Overview of PROM development This section may be improved by presenting the information here using the most appropriate tense. E..g 'We will use…' or 'We have used …' .instead of 'We use …' b. Phase I: Line 53 Suggestion – Delete 'and we proceeded with this study' because it sounds slightly colloquial c. Rigor: Lines 5 to 10 Authors assert to the robustness of the qualitative interview by stating that 'Several strategies have been put in place to ensure rigor of this qualitative study. A single interviewer will perform all of the qualitative interviews. One team member will perform all of the coding of the transcripts, which will then corroborated by a second team member.' Rigor may be limited if there is no piloting of interview guide or checking of discrepancies regarding the interpretation of questions. Expert Clinical input: Lines 10 to 15 Please clarify whether experts referred to here also include researchers and academics. Please consider revision here to improve comprehension: 'Expert input will be obtained before the final round of participant cognitive interviews in order to show any changes made from expert input to participants.' e. Translation: Lines 28 to 30 Please consider revision here to improve comprehension: 'Any discrepancies are resolved at each step and the resultant version is then shown to a small group of patients to ensure that the translation is valid and ready for use' 6. Ethics For this section, it will be important to focus on key ethical issues relating to the study, especially for aspects that will be conducted in multiple jurisdictions. 7. Table 1 Interview guide for qualitative interviews to be performed in Phase 1. Please consider re-phrasing the following questions to improve comprehension. (The ethical implication of other questions have not been considered as these beyond the scope of this review) - What was good or bad about the treatment?

	- What are the people like who care for you? Probe: friendly, made you feel comfortable, easy to talk to, listened to you - Does your condition create any functional problems?

REVIEWER	Isam Atroshi Lund University, Sweden
REVIEW RETURNED	30-Sep-2018

GENERAL COMMENTS	Thank you for the opportunity to review this interesting manuscript. This is a study protocol for a project that aims to develop a new patient-reported outcome measure for patients with hand conditions. The project seems well-designed and follows many current guidelines concerning development of patient-reported outcome measures and uses methodology based on modern measurement theory. A number of outcomes measures intended for assessing hand conditions, such as the DASH/QuickDASH and Michigan Hand Questionnaire, are currently available. The main limitation of these measures has not been lack of reliability or validity but rather inadequate responsiveness in certain conditions, mainly when comparing the efficacy of two treatment methods (a good example is distal radius fracture). Low responsiveness has also been observed when assessing effectiveness of treatments not expected to have a large effect size (such as non-surgical treatments) and treatments of diseases that usually do not cause substantial symptoms and activity limitations (such as Dupuytren disease). A new measure with higher responsiveness will be very useful for clinical research in hand surgery. Another related issue limiting the usefulness of current measures is lack of clear definition of what score difference/change constitutes a clinically important improvement. Responsiveness and definition of clinically important change will however be assessed only after the measure has been developed and tested for reliability and validity. Therefore, it remains to be seen whether this new measure will solve the problems encountered with the current measures. Introduction: Many readers may not be familiar with concepts like “modern psychometric theory” and “classical test theory”. The third paragraph in the introduction involves a very specific issue. Is the main purpose of this new measure to address patient experiences (not addressed in current measures) rather than symptoms/function? Methods: How will the final scales be chosen? Does Table 1 list the actual preliminary scales? Is it reasonable to include symptoms as part of a physical function scale? There seems to be a disproportionate emphasis on type of anesthesia. The most important outcomes of treatment are the degree of improvement in symptoms and in activity limitations caused by the disease, and usually type of anesthesia is not a major determinant of outcome. In most cases if you have an accurate diagnosis the outcome is determined by the efficacy of the treatment rather than if you do the surgery in local or regional anesthesia. Besides, an outcome measure should perform well even when used to evaluate the effectiveness of non-surgical treatments. Similarly, the focus on appearance of the hand as an

	independent scale is somewhat surprising given that symptoms and function are combined in one domain. The burden on patients is an important issue to consider when choosing an outcome measure especially in clinical research as patients may be asked to complete the scales multiple times. In these situations the most responsive scales will be chosen as the primary outcome to minimize sample size and patient burden. In most clinical trials comparing treatment effectiveness this will be symptoms and function rather than experience or appearance, although they may also be important and added as secondary outcomes. Participants: Why include only patients that have undergone surgery? Considering the very large number of different hand diagnoses that may have been treated surgically during 12 months, it would be interesting to know how many patients per diagnosis would be needed to adequately cover important symptoms and activities as well as appearance, experiences and other aspects related to a specific diagnosis. How do you ensure that this sample is representative of all hand conditions? Validity: The authors plan to use hypothesis testing as part of assessing construct validity. However only 3 hypotheses have been presented. The first hypothesis is to compare rheumatoid arthritis patients (a disease that often causes hand and wrist deformities) with patients with carpal tunnel syndrome (a disease that in the vast majority of patients does not have any effect on hand appearance except a possible surgical scar after surgery). Any measure can be validated using such a hypothesis. The second hypothesis involves patients “requiring further intervention”, how do you define that? It is unclear what is meant by “quality of life scores”, is “quality of life” a specific scale in this new outcome measure? There are established widely used measures of quality of life (such as the SF-36 and EQ-5D) and devising a new QoL measure for patients with hand conditions does not seem to be an urgent issue. Translation: Does a “prospective” translation preclude the need for these translated versions to be evaluated for reliability and validity, as is the case for “retrospectively” translated scales? Cognitive debriefing interviews: What are the inclusion criteria for the participants? Are these the same as in Phase 1 or will it include others? Analyses: Will confirmatory factor analysis be part of the analyses? Scoring: How will the new measure be scored? For this measure to be useful in clinical practice to monitor individuals (rather than on group level in research) can we assume that clinicians will be able to easily perform scoring based on RMT?

REVIEWER	Dr. Eloise Carr Faculty of Nursing, University of Calgary Canada

REVIEW RETURNED	10-Oct-2018

GENERAL COMMENTS	This is an important protocol describing an ambitious international multi-phase mixed methods study. It is evident that the team have developed a robust methodological approach and generally describe the details clearly. It is pleasing to see the patient experience so central to the development of a measurement tool. My comments relate specifically to the mixed method approach and qualitative and aspects of the protocol. The title states that the study is a multi-phase mixed methods study but there is no clear description of the research design or the reason for using a mixed methods approach. Although the authors state it is a multiphase study there is synergy with Cresswell & Plano's (2018) typology for an exploratory sequential design (the first phase is a qualitative study to inform the constructs for survey development). The relationship between the qualitative and quantitative phases are important to understand within the study design. A statement regarding the design typology and rationale would strengthen the protocol. The choice of interpretive description (ID) as the framework for the qualitative phase would seem entirely appropriate, as it puts the person's experience of healthcare at the centre. ID permits the exploration of what is known about a topic and also what is not known and the focus being on the human experience of health. This would seem entirely suitable given the previously completed systematic review and their desire to further understand the patients' experience of hand surgery. The approach to sampling and data analysis are in keeping with the qualitative approach. Recognizing that it is not possible to provide an exact sample size for the qualitative interviews, perhaps the authors can provide a general idea, based on their prior experience perhaps. There is a short section on rigor which is important in qualitative research. The authors use the term 'member-checking' (Page 9: line10) but refer to their research team. This term usually refers to the process of sharing the analysis with the participants rather than the study team. Please clarify the meaning here.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1: This submission covers an interesting and important issue – developing a reliable and acceptable culturally-relevant patient-reported outcome measure (PROM) for a wide range of conditions affecting the hand. In general, the methodology adopted for the study is sound. However, more clarity about the methods described in the submitted protocol will immensely improve the quality of the submission

Response: Thank you for feedback. We have made changes to the manuscript to aid clarity of the methodology used. We hope that these improvements address your concerns.

Reviewer 2: Thank you for the opportunity to review this interesting manuscript. This is a study protocol for a project that aims to develop a new patient-reported outcome measure for patients with

hand conditions. The project seems well-designed and follows many current guidelines concerning development of patient-reported outcome measures and uses a methodology based on modern measurement theory.

Response: Thank you.

Reviewer 2: A number of outcomes measures intended for assessing hand conditions, such as the DASH/QuickDASH and Michigan Hand Questionnaire, are currently available. The main limitation of these measures has not been lack of reliability or validity but rather inadequate responsiveness in certain conditions, mainly when comparing the efficacy of two treatment methods (a good example is distal radius fracture). Low responsiveness has also been observed when assessing the effectiveness of treatments not expected to have a large effect size (such as non-surgical treatments) and treatments of diseases that usually do not cause substantial symptoms and activity limitations (such as Dupuytren disease). A new measure with higher responsiveness will be very useful for clinical research in hand surgery. Another related issue limiting the usefulness of current measures is lack of clear definition of what score difference/change constitutes a clinically important improvement. Responsiveness and definition of clinically important change will however be assessed only after the measure has been developed and tested for reliability and validity. Therefore, it remains to be seen whether this new measure will solve the problems encountered with the current measures.

Response: Thank you for your insight about what is lacking in the field of hand outcome measures currently, we agree that there is a need for an instrument with increased responsiveness to ensure that we can measure the effects of the full variety of treatments offered to patients with hand conditions. The HAND-Q is being designed to be a clinically meaningful, discriminative instrument to allow for outcome assessment at the individual level. The use of Rasch Measurement Theory (RMT) to construct the scales will create scales that are able to differentiate between disease states with more accuracy than the legacy patient reported outcome measures (PROMs) that you referred to.

As you are aware, the responsiveness and clinically important change are only able to be established once the HAND-Q has been completed, so this does remain to be proven. However, by employing the techniques of PROM development that we have proven to be successful in other areas, we anticipate that this will be shown in due course.

Reviewer 2: Introduction: Many readers may not be familiar with concepts like “modern psychometric theory” and “classical test theory”.

Response: To address your concerns about using terms unfamiliar to the reader, we have included definitions of the concepts of “modern psychometric theory” and “classical test theory”.

Reviewer 2: The third paragraph in the introduction involves a very specific issue. Is the main purpose of this new measure to address patient experiences (not addressed in current measures) rather than symptoms/function? … There seems to be a disproportionate emphasis on type of anesthesia. The most important outcomes of treatment are the degree of improvement in symptoms and in activity limitations caused by the disease, and usually type of anesthesia is not a major determinant of outcome. In most cases if you have an accurate diagnosis the outcome is determined by the efficacy of the treatment rather than if you do the surgery in local or regional anesthesia. Besides, an outcome measure should perform well even when used to evaluate the effectiveness of non-surgical treatments. Similarly, the focus on appearance of the hand as an independent scale is somewhat surprising given that symptoms and function are combined in one domain.

Response: The HAND-Q is intended to provide a suite of independently functioning scales that measure concepts that matter to patients regarding their experience and their perception of their quality of life or outcome. Of course, we agree with you that in terms of measuring outcome the primary concepts of interest to this population are function and symptoms. There are multiple PROMs

that have purely focussed on these concepts, without acknowledgement of other factors. Each HAND-Q scale will measure a single concept. Therefore symptoms and function will be measured with separate scales.

In creating a suite of scales, the HAND-Q will allow for measurement of secondary concepts of interest such as hand appearance, patient experience and satisfaction. We believe that the patient experience is important to consider and understand to improve the services that are offered to this patient cohort. Hand appearance is acknowledged as an important motivator for surgical intervention in patients with rheumatoid arthritis. However, there is currently no widely accepted measure of hand appearance(5, 6).

There are often multiple procedures that can be used to treat a condition of the hand, without significant differences in clinically measured or patient reported outcome. An example of such a condition is that of 1st CMC arthritis; there are multiple surgical methods of managing this condition, none have been proven to be superior when using clinical measurements or legacy PROMs(7, 8). We propose that in this situation that we should be considering the patient's preference in terms of experience to guide decision making and informed consent.

Reviewer 2: The burden on patients is an important issue to consider when choosing an outcome measure especially in clinical research as patients may be asked to complete the scales multiple times. In these situations the most responsive scales will be chosen as the primary outcome to minimize sample size and patient burden. In most clinical trials comparing treatment effectiveness this will be symptoms and function rather than experience or appearance, although they may also be important and added as secondary outcomes.

Response: We agree that patient burden is a critical issue that should always be considered and minimised where possible. An asset to the design of the HAND-Q is that it is composed of independently functioning scales. This means that there is the ability to tailor which scales are used to the individual, the research project or the condition at hand. By using only the scales that measure concepts of interest the burden on patients is substantially minimised. We also concur that measurement of what you have referred to as ‘secondary outcomes’ – such as the patient experience and hand appearance, may be very important factors in some patient cohorts and of interest in research applications.

Reviewer 2: Methods: How will the final scales be chosen? Does Table 1 list the actual preliminary scales? Is it reasonable to include symptoms as part of a physical function scale?

Response: The field test scales were developed based on the qualitative data attained from patients regarding the outcome and experience issues that matter to them (therefore they are not listed in Table 1 as this is only the initial version of the interview guide). The items included in each scale will be chosen based on the results of the Rasch Analysis of the field test results. The items that make up each of the scales will be determined by the psychometrics of each item and the effects on the overall scale characteristics. Scales will be designed to measure only one concept to meet the requirements of unidimensionality that Rasch is based upon. Therefore, concepts such as symptoms and function would be measured using separate scales. When the instrument is finalised and made available to researchers and clinicians, they will be able to pick and choose the concepts to be measured for their application, as each scale of the HAND-Q can function independently.

Reviewer 2: Participants: Why include only patients that have undergone surgery? Considering the very large number of different hand diagnoses that may have been treated surgically during 12 months, it would be interesting to know how many patients per diagnosis would be needed to adequately cover important symptoms and activities as well as appearance, experiences and other aspects related to a specific diagnosis. How do you ensure that this sample is representative of all hand conditions?

Response: The HAND-Q will be suitable for use for patients who are receiving surgical or non-operative treatments. The design of the HAND-Q allows for only the applicable scales to be administered. Thus patients receiving non-operative treatments would not be asked about their experience of hand surgery. By measuring separate concepts with individual scales, this allows for improved measurement characteristics and clarity in the interpretation of scores and change over time.

Our participant recruitment for the qualitative interviews was limited to include only those who had undergone surgery because we wanted to explore their experience both before and after hand surgery.

Due to the extreme heterogeneity of hand conditions, it would be impractical if not impossible to include a representative with every condition. However, efforts were made to ensure that the group interviewed included a range of conditions commonly treated in the hand clinic setting. Qualitative interviews were conducted until saturation was met, that is there were no new concepts identified in a series of 3 interviews. In total we performed 62 interviews, each lasting approximately 60 minutes. The recruitment criteria for the field testing in Phase II is broad; it includes any patient being seen at the participating hand clinics that can read and understand the questionnaire. The items selected for each scale will be based on the data collected from this widely heterogenous international and multilingual cohort. We believe these efforts will ensure that the HAND-Q is suitable for use in all patients with a hand condition internationally.

Reviewer 2: Validity: The authors plan to use hypothesis testing as part of assessing construct validity. However only 3 hypotheses have been presented. The first hypothesis is to compare rheumatoid arthritis patients (a disease that often causes hand and wrist deformities) with patients with carpal tunnel syndrome (a disease that in the vast majority of patients does not have any effect on hand appearance except a possible surgical scar after surgery). Any measure can be validated using such a hypothesis. The second hypothesis involves patients “requiring further intervention”, how do you define that?

It is unclear what is meant by “quality of life scores”, is “quality of life” a specific scale in this new outcome measure? There are established widely used measures of quality of life (such as the SF-36 and EQ-5D) and devising a new QoL measure for patients with hand conditions does not seem to be an urgent issue.

Response: The validity hypotheses performed at the end of Phase II are not trying to prove the responsiveness of the scales. These hypotheses are to ensure that the scales are functioning in a logical and expected manner. Thus, the au priori hypotheses are quite general in their intent. More clinically specific hypotheses will be tested in Phase III which includes responsiveness, reliability and validity testing of the final HAND-Q scales. Detail of this has not been included in this protocol paper as this stage is yet to receive funding and ethical approval.

The variable of whether patients require further intervention or not is a self-reported variable that is included in the field-testing, as is the description of clinical severity as mild, moderate or severe. The hypothesis regarding “quality of life” scores has been rephrased for clarity.

Reviewer 2: Translation: Does a “prospective” translation preclude the need for these translated versions to be evaluated for reliability and validity, as is the case for “retrospectively” translated scales?

Response: Incorporating translation and cultural validation into the development of the HAND-Q means that the HAND-Q scales will be inherently reliable and valid in the included populations. Since the translations will then be included in the field-testing, we will be able to examine differential item function by language to determine if the scales work the same across language. Also, the combined

dataset will provide evidence of validity and reliability and whether or not we can develop a common scoring algorithm that will work internationally.

Reviewer 2: Cognitive debriefing interviews: What are the inclusion criteria for the participants? Are these the same as in Phase 1 or will it include others?

Response: The Cognitive debriefing interviews were conducted with a subgroup of participants from the Qualitative interviews conducted in Australia and Canada. The additional cohort from the United States was subjected to the same inclusion criteria as the Qualitative study. This has been clarified in the manuscript.

Reviewer 2: Analyses: Will confirmatory factor analysis be part of the analyses?

Response: Confirmatory factor analysis will not be a formal part of the analysis. The unidimensional nature of each scale will be confirmed with Rasch Measurement Theory (RMT) analysis.

Reviewer 2: Scoring: How will the new measure be scored? For this measure to be useful in clinical practice to monitor individuals (rather than on group level in research) can we assume that clinicians will be able to easily perform scoring based on RMT?

Response: Scoring of the HAND-Q will be based on lookup tables that will allow for easy conversion of the raw scores to a 0 to 100 score based on the Rasch logit scores. This allows for interval level measurement to be used which is more robust and suitable for application to the individual level in the clinical setting.

Reviewer 3: This is an important protocol describing an ambitious international multi-phase mixed methods study. It is evident that the team have developed a robust methodological approach and generally describe the details clearly. It is pleasing to see the patient experience so central to the development of a measurement tool. My comments relate specifically to the mixed method approach and qualitative and aspects of the protocol.

Response: Thank you.

Reviewer 3: The title states that the study is a multi-phase mixed methods study but there is no clear description of the research design or the reason for using a mixed methods approach. Although the authors state it is a multiphase study there is synergy with Cresswell & Plano's (2018) typology for an exploratory sequential design (the first phase is a qualitative study to inform the constructs for survey development). The relationship between the qualitative and quantitative phases are important to understand within the study design. A statement regarding the design typology and rationale would strengthen the protocol.

Response: Thank you for your insights and feedback about the qualitative and mixed methods used in this study. As you suggested, we have added a statement to explain the design typology and rationale.

Reviewer 3: The choice of interpretive description (ID) as the framework for the qualitative phase would seem entirely appropriate, as it puts the person's experience of healthcare at the centre. ID permits the exploration of what is known about a topic and also what is not known and the focus being on the human experience of health. This would seem entirely suitable given the previously completed systematic review and their desire to further understand the patients' experience of hand surgery.

Response: Thank you.

Reviewer 3: The approach to sampling and data analysis are in keeping with the qualitative approach. Recognizing that it is not possible to provide an exact sample size for the qualitative interviews, perhaps the authors can provide a general idea, based on their prior experience perhaps.

Response: We have also added an estimate of the expected sample size that would be needed to reach saturation in the qualitative study. This is based on the experience of our team in the development of the other PROMs.

Reviewer 3: There is a short section on rigor which is important in qualitative research. The authors use the term ‘member-checking’ (Page 9: line10) but refer to their research team. This term usually refers to the process of sharing the analysis with the participants rather than the study team. Please clarify the meaning here.

Response: We have amended the paragraph on rigor to provide greater clarity with regards to the processes of member checking and peer debriefing that were used.

VERSION 2 – REVIEW

REVIEWER	Isam Atroshi Lund University Sweden
REVIEW RETURNED	31-Dec-2018

GENERAL COMMENTS	The authors have responded to the issues raised in the review. No "limitations" are mentioned in the "Strengths and Limitations" box, does it mean this study will produce the "perfect" hand questionnaire?

REVIEWER	Dr. Eloise Carr Faculty of Nursing/Adjunct Research Professor, Cumming School of Medicine (Community Health Sciences) University of Calgary, 2500 University Dr NW Calgary, AB Canada T2N 1N4
REVIEW RETURNED	21-Dec-2018

GENERAL COMMENTS	Thank you for the revisions to the manuscript. Your responses and revisions are very helpful. This is a very nice manuscript and the study will be a very worthwhile contribution to the field.

VERSION 2 – AUTHOR RESPONSE

Reviewer 2: The authors have responded to the issues raised in the review. No "limitations" are mentioned in the "Strengths and Limitations" box, does it mean this study will produce the "perfect" hand questionnaire?

Response: Thank you for your comments. While the HAND-Q will be a robust instrument, we recognise that all research has limitations. We have altered the “Strengths and Limitations” field to include a limitation of our protocol. “The qualitative components of this study have only been performed with English speaking patients from Australia, the United States and Canada.”

Reviewer 3: Thank you for the revisions to the manuscript. Your responses and revisions are very helpful. This is a very nice manuscript, and the study will be a very worthwhile contribution to the field.

Response: Thank you for your encouragement, we hope that this work will be a valuable contribution to the field.