PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Development of 'Core Outcome Sets' for Meningioma in Clinical Studies (The COSMIC Project): Protocol for two Systematic Literature Reviews, eDelphi Surveys and Online Consensus Meetings.
AUTHORS	millward, Christopher; Armstrong, Terri S.; Barrington, Heather; Bell, Sabrina; Brodbelt, Andrew; Bulbeck, Helen; Crofton, Anna; Dirven, Linda; Georgious, Theo; Grundy, Paul; Islim, Abdurrahman; Javadpour, Mohsen; Keshwara, Sumirat M.; Koszdin, Shelli; Marson, Tony; McDermott, Michael W.; Meling, Torstein R.; Oliver, Kathy; Plaha, Puneet; Preusser, Matthias; Santarius, Thomas; Srikandarajah, Nisaharan; Taphoorn, Martin J.B.; Turner, Carole; Watts, Colin; Weller, Michael; Williamson, Paula; Zadeh, Gelareh; Zamanipoor Najafabadi, Amir H.; Jenkinson, Michael; ICOM, International Consortium on Meningiomas

VERSION 1 – REVIEW

REVIEWER	Hoffmann, Christin University of Bristol, Population Health Sciences
REVIEW RETURNED	08-Nov-2021

GENERAL COMMENTS	Thank you for the opportunity to review the manuscript which outlines a protocol for the development of two core out sets to be used in clinical effectiveness trials for i) meningioma (COSMIC: Intervention) and for incidental/untreated meningioma (COSMIC: Observation). The authors have developed a protocol to generate two core outcome sets for two different approaches to treatment and the reasons for doing so are clear. I think this is an interesting topic and the authors propose a well thought out and clearly presented approach to a review. Objectives, methods, and analyses are clear and robust. Appropriate PPIE involvement and oversight by a SAG are both well described throughout. Some suggestions for improvement: The writing is clear and concise given the complexity of the methods. There seemed repetition in the two presented systematic reviews and for the benefit of reducing length and complexity, the authors could consider reducing duplication and, for example, only highlight areas where Systematic Review (SR) 2 differs from SR 1. Similarly, there is considerable repetition of information in Table 1. Reduction in duplicate information would make it easier for the reader to spot differences. But this may be personal preference. The presented measures to ensuring robustness of methods are appropriate. I also agree with using 10% of included articles to check agreement in data extraction. In my view, it is unlikely that the two reviewers will extract 100% identical data and I would suggest defining "concordance" (p. 21 line 11) in this case.

	Analyses are mostly clear, but the authors can add a few lines about their plans for quantitative data synthesis from the eDelphi surveys. The authors clearly describe the generation of two long lists how data from trial/SR searches will inform the separate lists. It is less clear, however, how the supplemental data sources will feed into the separate lists. It would be good to have clarification as to whether the authors expect duplication in both lists or how decisions about outcome allocation will be made in this case. It is unusual to pick specific publications a priori as supplemental data source for the long list as opposed to performing a more comprehensive and systematic supplemental search for publications. What exactly are the reasons for picking only one qualitative paper and two specific systematic reviews? A bit more detail on how the qualitative paper will be analysed would be beneficial. It was unclear whether authors just extract the themes identified in the original publication or perhaps have access to transcripts to perform further secondary analyses? Some additional points for clarification: The authors describe that “In the second round of the eDelphi surveys, the round one response from each of the two panels will be presented separately for each outcome” (p.27, lien 3) Do the authors mean an average of the stakeholder groups’ score will be presented? If so, perhaps specify. Description of consensus meeting is less detailed. Perhaps, the authors can add a few lines about how exactly are presentation, discussion and voting be conducted. How or whether it will be evaluated and, for example, whether the authors considering virtual or face-to-face methods? I wondered whether this sentence is contradictory. “The primary focus of each meeting will be to discuss those outcomes without consensus. These outcomes will be presented for discussion amongst stakeholder representatives from both eDelphi survey panels and voted on in order to achieve consensus.” (p. 31, line17) Aren’t discussions are necessary to move towards agreement, yet without establishing formal quantifiable consensus at that stage? Best of luck to the authors with this work.

REVIEWER	Evans, Gareth The University of Manchester, and Central Manchester University Hospitals NHS Foundation Trust, Saint Mary's Hospital
REVIEW RETURNED	12-Dec-2021

GENERAL COMMENTS	The authors present a Development of ‘Core Outcome Sets’ for Meningioma in Clinical Studies (The COSMIC Project). The Protocol for two Systematic Literature Reviews, eDelphi Surveys and Consensus Meetings is presented. The document is overall well written. More clarity is needed on inclusion/exclusion criteria earlier in the document. For instance the important subgroup with NF2 appears to be excluded but this is still not entirely clear from the documentation. The authors should also provide more clarity about the Delphi process. Given several authors are on a very recent and similar piece of work recently published in Neuro-oncology they need to make clear this is an entirely separate piece of work. Specific comments 1. ‘Meningioma are more than twice as common in females (12 vs 5.3 per 100,000 population).[1]’ -Whilst this is true boys are also more likely to be affected by meningioma than girls 2. ‘Intracranial meningioma clinical effectiveness trials’ The INTUITT brigatinib trial in NF2 related meningioma should also be mentioned 3. There is no mention of inherited predisposition to meningioma or

	‘multiple meningioma’ patients. Does this work apply to these patients or not? The introduction should at least allude to these conditions given that treatment trials are being developed specifically for NF2 related meningioma. At least 1% of meningioma occurs within the context of an NF2 patient and around 4% additionally develop more than one meningioma. This is an important sub population that must be mentioned in the context of this study/ Rarer inherited predisposition such as SMARCE1 related clear cell meningioma is another example. 4. Several of the senior authorship are on very similar work just published. Although under ‘exclusion criteria’ the following is stated ‘Similarly, studies investigating meningioma only in patients with the genetic condition Neurofibromatosis type 2 (NF2) would not be included as again, this can be considered a different disease entity’ - This does not make it clear that NF2 patients are excluded from use of the resultant outcome measures. 5. https://pubmed.ncbi.nlm.nih.gov/34791428/. How is the present work different? This should be stated clearly in the manuscript. 6. ‘Observation COS requires a majority agreement from both panels, of its critical importance.[36] During round 1 of the eDelphi surveys, participants will rate the importance of each outcome presented using the 9-point Likert scale. It shall be explained to participants that the following scores represent outcome importance, whereby (1-3) is of limited importance, (4-6) is important but not critical, and (7-9) is critically important.’ -Please make this completely clear. Does this mean that a simple 51% majority scoring 7-9 will suffice? Most Delphi use a higher threshold. What if 30% score as of limited importance? As you state lower down that 80% need to score at 7-9 you need to make it clear that ‘large’ majority is required here 7. ‘All items from round 1 will be carried forward to the second round of each eDelphi survey. In the second round of the eDelphi surveys, the round one response from each of the two panels will be presented separately for each outcome’ -Will there be any change in wording of the items to build consensus between round 1 and round 2? 8. ‘Participants eligible to complete both eDelphi surveys, who subsequently complete both rounds of both eDelphi surveys, will be eligible to take part in both consensus meetings.’ -will these participants gain authorship on the final publication and if so how? A s named authors or as part of a group authorship? 9. Please provide an intended timeline for this piece of work and whether any elements are already underway/completed

VERSION 1 – AUTHOR RESPONSE

Reviewer 1 comments:

The writing is clear and concise given the complexity of the methods. There seemed repetition in the two presented systematic reviews and for the benefit of reducing length and complexity, the authors could consider reducing duplication and, for example, only highlight areas where Systematic Review (SR) 2 differs from SR 1.

Thank you for highlighting this. We agree and so the length of the text has been reduced considerably for SR2. We have maintained a degree of separation because PROSPERO does not allow the registration of COS outcome reviews such as these (although many can be found in the database), and we wanted to ensure our methods were recorded comprehensively a priori.

Similarly, there is considerable repetition of information in Table 1. Reduction in duplicate information would make it easier for the reader to spot differences. But this may be personal preference.

Thank you for highlighting this. I do agree that there is repetition, but 5 of the 11 standards do differ. I think it is very important to see that The COSMIC Project is developing two distinct but related COS, and I think that this warrants each having its own column within this table. It is a COS-STAD requirement that each standard is addressed, so for transparency, I would prefer to keep these separate.

The presented measures to ensuring robustness of methods are appropriate. I also agree with using 10% of included articles to check agreement in data extraction. In my view, it is unlikely that the two reviewers will extract 100% identical data and I would suggest defining “concordance” (p. 21 line 11) in this case.

Thank you for highlighting this. Outcome extraction can be highly variable in COS development studies and we agree that 100% concordance is highly unlikely. The main data point to be extracted is, without doubt, the verbatim outcome which is subject to subjective variability. Other data points can be extracted once this is known. Therefore, concordance on what outcomes exist within each study with variability of no more than 5% will be the accepted level, and this has been incorporated into the manuscript.

Analyses are mostly clear, but the authors can add a few lines about their plans for quantitative data synthesis from the eDelphi surveys.

A sentence has been added to explain how the consensus percentage is calculated but there is nil other quantitative analysis required.

The authors clearly describe the generation of two long lists how data from trial/SR searches will inform the separate lists. It is less clear, however, how the supplemental data sources will feed into the separate lists. It would be good to have clarification as to whether the authors expect duplication in both lists or how decisions about outcome allocation will be made in this case. It is unusual to pick specific publications a priori as supplemental data source for the long list as opposed to performing a more comprehensive and systematic supplemental search for publications. What exactly are the reasons for picking only one qualitative paper and two specific systematic reviews? A bit more detail on how the qualitative paper will be analysed would be beneficial. It was unclear whether authors just extract the themes identified in the original publication or perhaps have access to transcripts to perform further secondary analyses?

We have removed details of adding additional outcomes from a HRQoL SR and neurocognitive SR. The reason for this is that even if more outcomes were to be identified from such reviews, they would end up being collapsed back into the broader categories of HRQoL and Neurocognition. To not do this would produce long-lists that are excessively long, and far too specific for stakeholders during within our delphi survey. For completeness, the reason specific reviews were mentioned was because the HRQoL review is co-authored by myself – and I therefore know that it has robustly identified PROMs used in meningioma research. The neurocognition SR was authored by a SAG member, and I know that this is the most up-to-date and comprehensive review published to date, and would have more than served the purpose required for The CSOMIC Project. Either way, these methods are no longer required.

I have however, expanded the methods section of the eDelphi surveys to explain that the SAG can adjust/modify/add outcomes where appropriate, which allows for additional outcomes of interest to be incorporated. This includes those added by patient research partners before round 1, and by patient and healthcare professional participants after round 2 of the delphi. This includes those outcomes identified from previously published semi-structured interviews, of which only one paper has been published. A scoping review demonstrated this to us, and therefore a full systematic review of semi-structured interviews would have been futile. The authors of the paper describing the semi-structured interviews are also part of the SAG and can add these outcomes from these interviews at the SAG level. This is an accepted methodology as per COMET.

Some additional points for clarification:

The authors describe that “In the second round of the eDelphi surveys, the round one response from each of the two panels will be presented separately for each outcome” (p.27, lien 3) Do the authors mean an average of the stakeholder groups’ score will be presented? If so, perhaps specify.

Thank you for highlighting this. It is the distribution of scores that will be presented for each outcome and for each stakeholder group. The manuscript has been adjusted accordingly.

Description of consensus meeting is less detailed. Perhaps, the authors can add a few lines about how exactly are presentation, discussion and voting be conducted. How or whether it will be evaluated and, for example, whether the authors considering virtual or face-to-face methods?

The meetings will be virtual but this is a rapidly evolving methodology. Preliminary guidelines have been produced by COMET and currently concern considerations for designing such meetings. Reference to “online” have been added throughout, and the methods section for these meetings has been expanded. The analysis is already described and has not been adjusted.

I wondered whether this sentence is contradictory. “The primary focus of each meeting will be to discuss those outcomes without consensus. These outcomes will be presented for discussion amongst stakeholder representatives from both eDelphi survey panels and voted on in order to achieve consensus.” (p. 31, line17) Aren’t discussions are necessary to move towards agreement, yet without establishing formal quantifiable consensus at that stage?

I do not think the statement is contradictory but I have adjusted the sentences slightly to increase clarity of meaning. The eDelphi surveys will have achieved consensus on the definite inclusion and exclusion of a proportion of outcomes. Therefore, discussion of these outcomes is not needed per se, but they are often reviewed for completeness. The primary purpose of the consensus meeting is to discuss those that have not achieved consensus IN or OUT, and to hold repeat discussions and rounds of voting in order to build consensus and evaluate if a move towards the consensus threshold has taken place. If consensus cannot be reached on a number of outcomes, a further consensus meeting may be necessary.

Reviewer 2 comments:

1. ‘Meningioma are more than twice as common in females (12 vs 5.3 per 100,000 population).[1]’ - Whilst this is true boys are also more likely to be affected by meningioma than girls

Thank you for noting this. The opening paragraph serves to describe the basic epidemiology of the disease only. It is my understanding that boys are more likely to be affected in the context of Neurofibromatosis type 2, rather than sporadic meningioma, so this would not be a necessary comment to add I believe.

2. ‘Intracranial meningioma clinical effectiveness trials’ The INTUITT brigatinib trial in NF2 related meningioma should also be mentioned

This is, of course, an important trial. However, this trial is not of direct relevance, given the scope of the planned COS which does not include patients with NF2.

3. There is no mention of inherited predisposition to meningioma or ‘multiple meningioma’ patients. Does this work apply to these patients or not? The introduction should at least allude to these conditions given that treatment trials are being developed specifically for NF2 related meningioma. At least 1% of meningioma occurs within the context of an NF2 patient and around 4% additionally develop more than one meningioma. This is an important sub population that must be mentioned in the context of this study/ Rarer inherited predisposition such as SMARCE1 related clear cell meningioma is another example.

Thank you for raising this important issue. The scope of a COS i.e. the patients/context/intervention for which it has been developed is highly important. I have recently authored a review article that is under review with neuro-oncology that focuses on this topic specifically within neuro-oncology. The

study advisory group has unanimously agreed that NF2-associated meningioma should not be in either of the two COS that we are developing. As you have noted, it represents 1% of meningioma but is a fundamentally different disease entity, in the demography, other tumours it is associated with, as well as eye/skin manifestations and overall prognosis. The methodology of this study would be flawed should it try to include NF2 studies in the literature review and long-list, then NF2 patients in a Delphi and consensus meeting. Consensus would not easily be achieved on core outcomes. Very few delphi participants would have NF2, especially if participants were recruited in line with epidemiology, therefore, no one would agree that what is important to an NF2 patient, is core for all meningioma patients. If NF2 was included, outcomes of relevance to this patient population would be rejected, yet the COS would claim to be applicable to NF2 trials, having been developed with very few NF2 patients and without NF2 specific outcomes being carried forward. The COS would do a great disservice to this patient population as a result. The same is true for spinal meningioma. We have had discussions about developing a COS for NF2 and I would strongly suggest that a dedicated COS would better serve this patient group and clinical researchers working in this area. This is something we are keen to explore and would welcome collaboration on. Finally, on the issue of multiple meningioma and SMARCE1, we can find no reason to not include these patients. It would not positively or negatively influence the study.

4. Although under 'exclusion criteria' the following is stated 'Similarly, studies investigating meningioma only in patients with the genetic condition Neurofibromatosis type 2 (NF2) would not be included as again, this can be considered a different disease entity' -This does not make it clear that NF2 patients are excluded from use of the resultant outcome measures.

Thank you for identifying this. Again, you raise an important point. Building on my previous response in point 3, I would suspect that COSMIC intervention would not be all-encompassing for use in future NF2 trials. However, some of the outcomes may be core to NF2 patients, but this would not be known, as NF2 patients cannot be incorporated into this study. Uptake of COS is an interesting topic. Uptake is generally poor, and we know that researchers make individual decisions about whether a COS exists which is relevant/applicable to a trial that is to be performed. It is generally up to a triallist to scrutinize the protocol, systematic reviews, and final COS manuscript in order to make this assessment. I have added sporadic to the COS STAD table, to make the population clearer, so thank you for highlighting this. I would refrain from discussing the specifics of uptake within the protocol, but would certainly discuss this within the final COS manuscript, or better still, a protocol for the development of an NF2 COS. It is worth noting that some triallists use more than one COS for a trial should they exist.

5. Several of the senior authorship are on very similar work just published. https://pubmed.ncbi.nlm.nih.gov/34791428/. How is the present work different? This should be stated clearly in the manuscript.

There is indeed a piece of work with many of the same co-authors, but this is in no way related. A core outcome set is concerned with outcomes that are measured and reported in clinical effectiveness trials in response to the intervention of interest and this is described in the introduction. The piece of work mentioned concerns common data elements… essentially, baseline characteristics of trial participants that are independent of a trial intervention. I am yet to see a protocol that also describes CDEs and I fear this only serves to cause confusion, given the fact the purpose of our work is clearly defined. I would however, make note of this work at the point of publishing the COS when discussing implementation within a trial setting – by making mention that common data elements have also been described.

6. 'Observation COS requires a majority agreement from both panels, of its critical importance.[36] During round 1 of the eDelphi surveys, participants will rate the importance of each outcome presented using the 9-point Likert scale. It shall be explained to participants that the following scores represent outcome importance, whereby (1-3) is of limited importance, (4-6) is important but not critical, and (7-9) is critically important.' -Please make this completely clear. Does this mean that a simple 51% majority scoring 7-9 will suffice? Most Delphi use a higher threshold. What if 30% score as of limited importance? As you state lower down that 80% need to score at 7-9 you need to make it clear that 'large' majority is required here

The sentence referring to majority is an opening sentence to the paragraph concerning eDelphi data

analysis. We state our threshold of 80%, which is higher than most COS studies to date, so there should be no ambiguity as to what our definition of majority consists of. We have however, added the word ‘large’ to complement this in the opening paragraph.

7. ‘All items from round 1 will be carried forward to the second round of each eDelphi survey. In the second round of the eDelphi surveys, the round one response from each of the two panels will be presented separately for each outcome’ -Will there be any change in wording of the items to build consensus between round 1 and round 2?

I am not aware of this being a routine methodology for such studies. We plan for consensus building to be achieved by showing stakeholder opinion for that outcome, and allowing a participant to reconsider. Should consensus not be achieved, either in or out for each outcome, the outcome would end up being discussed at the consensus meeting, and in this forum, terms are often merged and adjusted based on such discussions.

8. ‘Participants eligible to complete both eDelphi surveys, who subsequently complete both rounds of both eDelphi surveys, will be eligible to take part in both consensus meetings.’ -will these participants gain authorship on the final publication and if so how? A s named authors or as part of a group authorship?

Thank you for raising this issue. We intended to add this to our protocol as we feel it is appropriate to acknowledge the contribution of all stakeholders who make the research possible. This has not been routine in COS development, but is something I do feel is necessary to improve recruitment and retention of healthcare professionals in particular. A section has been added to the eDelphi consent section and the dissemination section describing that a collaborative group will be created for those completing both rounds of the eDelphi surveys, one for each of COSMIC: Intervention and COSMIC: Observation.

9. Please provide an intended timeline for this piece of work and whether any elements are already underway/completed

This point was also raised by the editor. A brief description has been added to the methods.

VERSION 2 – REVIEW

REVIEWER	Hoffmann, Christin University of Bristol, Population Health Sciences
REVIEW RETURNED	11-Mar-2022

GENERAL COMMENTS	The authors have addressed comments sufficiently and satisfactorily. I agree with the proposed changes to the manuscript and particularly welcome the shortening of the manuscript and options of supplementation of the long lists. I have no reservations against publication of this protocol.

REVIEWER	Evans, Gareth The University of Manchester, and Central Manchester University Hospitals NHS Foundation Trust, Saint Mary's Hospital
REVIEW RETURNED	11-Mar-2022

GENERAL COMMENTS	The authors have addressed the majority of the reviewer comments well. A few minor issues remain. 1. ‘Meningioma are more than twice as common in females (12 vs 5.3 per 100,000 population).[1]’ -Whilst this is true boys are also more likely to be affected by meningioma than girls Thank you for noting this. The opening paragraph serves to describe

	the basic epidemiology of the disease only. It is my understanding that boys are more likely to be affected in the context of Neurofibromatosis type 2, rather than sporadic meningioma, so this would not be a necessary comment to add I believe. ***The author believes wrongly. This is the same in the general population! Please alter this statement*** 2. The authors should state up front that NF2 are excluded but multiple meningioma and SMARCE1 related familial meningioma are not.

VERSION 2 – AUTHOR RESPONSE

Thank you for re-reviewing our work so quickly and for inviting minor corrections.

Other than the edits described below, I have only added a point of clarification that for COSMIC: observation, untreated intracranial meningioma can be either asymptomatic or minimally symptomatic, and adjusted Table 1 to reflect changes previously agreed upon with the reviewers regarding the use of published systematic reviews of patient-reported outcomes.

Dear Review 2 - Prof Evans.

Apologies, I had misinterpreted the original comment and see now that you refer to the paediatric population. I have adjusted the sentence by adding ", although boys are more likely to be affected than girls."

Text has been adjusted accordingly for the latter point