

A novel method for observing proportional group awareness and consensus of items arising from list-generating questioning

Ravi Jandhyala (✉ ravi@medialis.co.uk)

Technical advance

Keywords: novel consensus method, X-linked hypophosphatemia, prompted opinion, knowledge awareness, registries, survey design

Posted Date: August 26th, 2019

DOI: <https://doi.org/10.21203/rs.2.13006/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Current Medical Research and Opinion on March 11th, 2020. See the published version at <https://doi.org/10.1080/03007995.2020.1734920>.

Abstract

Background: Methodologies used to gain consensus among healthcare professionals, including variations of the Delphi technique and the RAND/UCLA appropriateness method, all force a consensus, corrupting the original opinion through the consensus-generating process. Furthermore, none assess knowledge awareness of the experts prior to the consensus process. Methods: Four case studies about X-linked hypophosphatemia (XLH) are reported to demonstrate the principle of group 'awareness' of items, consensus and the concept of prompted agreement. The novel methodology consisted of two surveys: Round 1 was an item-generation round in which participants were asked an open-ended question. Responses to Round 1 were collated into themes and developed into mutually exclusive items. Item generating was also performed using systematic literature reviews when appropriate. Items generated were used to develop a structured questionnaire (Round 2) comprising statements for which each participant identified their level of agreement using a five-point Likert scale. All responses were analysed anonymously. Item awareness, observed agreement consensus and prompted agreement were objectively measured. Results: The free-text responses to the item-generation round tested the awareness of specific concepts or items regarding setting up a European registry for XLH, the limitations of empirical treatment for XLH in children and adults, and triggers for treatment of XLH in adults. The four case studies showed different levels of item awareness, observed consensus and various degrees of prompted agreement. All participants agreed or strongly agreed with statements based on the most frequent items listed in Round 1. Less frequent items generated during Round 1 had various degrees of prompted agreement consensus, and some did not reach the consensus threshold of >50% agreement by the participants. Conclusions: Observed proportional group awareness and consensus is a relatively quick process compared with the Delphi technique and its variants, providing objective assessment of expert knowledge and standardized categorization of items with regard to awareness, consensus and prompting. It offers the opportunity for tailored management of each item or concept in terms of educational need and further investigation.

Introduction

In healthcare, consensus aligns opinion of experts and enables decisions to be made about a specific need or question to be answered, particularly when published evidence is either too extensive, confusing, or contradictory, or there is a dearth of evidence. Without evidence from well-conducted clinical trials, consensus of expert opinion, however, is the lowest quality of evidence. Nevertheless, expert opinion consensus often bridge the gap between no evidence and 'some' evidence consisting of anecdotal or case study reports, and subject to bias interpretation.

Methodologies or procedures used to gain consensus among healthcare professionals include the Delphi technique, modified Delphi techniques, and the RAND/UCLA appropriateness method. The original Delphi technique, developed in the 1950s by the RAND Corporation to make predictions or forecasting future events, has been refined and modified for use in many different areas, including government and healthcare [1]. The core components of the Delphi technique, including many of its variations, comprise anonymity, iteration with controlled feedback, statistical group response, and expert input [2]. Where the

Delphi technique and modified Delphi techniques differ is that the original Delphi technique starts with an open-ended questionnaire that is given to a panel of selected experts to solicit specific information about a subject or content area [1]. In contrast, a modified Delphi technique begins with a list of carefully pre-selected items, often using a systematic literature review as a starting point [3]. In this way, a modified Delphi consensus is based upon previously published evidence, the robustness of which can be readily assessed and categorized. The RAND/UCLA appropriate method was developed to combine the best available scientific evidence with the collective judgment of experts to yield a statement regarding the appropriateness of performing a procedure at the level of patient-specific symptoms, medical history, and test results thus finding consensus, thus the resulting consensus is specific to one disease of an explicit severity [4].

There are a number of potential problems with the current procedures of consensus building in a healthcare setting. The first being that these methods all force a consensus, and as such, the original opinion may be corrupted through the consensus-generating process. Often there is no way of knowing how closely the final consensus represents the original opinion of experts. Secondly, these methods do not measure or indicate awareness of items at the beginning of the process. Information or knowledge is required on which to base a consensus, and a person cannot hold an opinion on something that they have no knowledge of. Furthermore, the selection of the expert panel may influence the final consensus, particularly as experts are usually invited to participate based on an assumption of knowledge prior to starting the consensus process and is normally based on the experts' published research and/or current research interests. As such, an assessment of knowledge awareness prior to the consensus process would inform the reliability of the final consensus, yet this is rarely included as part of consensus reporting in a healthcare setting. Thirdly, these consensus processes do not indicate prompting of item awareness, for example, the interaction between expert consensus and subject matter prompting.

Any process to obtain consensus from a group of experts would benefit from understanding awareness of items on which a later consensus is to be formed. The four case studies about X-linked hypophosphatemia (XLH) are reported to demonstrate the principle of group 'awareness' of items and its relationship with consensus on whether those items are retained. Here, the novel methodology is described and the concept of proportional group knowledge awareness, prompted and observed consensus of items arising from list-generating questioning are further explored using these case studies.

Methods

Case studies

Four case studies were used to define and confirm the methodology for assessing proportional group awareness and consensus of items arising from list-generating questioning. These case studies represented a range of questions and scenarios to fully explore the novel methodology (Table 1).

Participants

Participants for each case study were recruited from a pool of experts identified according to specific pre-specified criteria to confirm their experience in the area of interest relating to the research question (Table 1). For case study 1, each participant had to have relevant patient registry experience. Recruitment was based on participants having a similar level of knowledge regarding an XLH registry, and exposure to the same project information prior to participation. Participants for case study 1 were recruited from the Medialis Real World Evidence Asset Management Division team of clinical study managers. The participants for case studies 2, 3 and 4 were recruited from a cohort of practicing doctors who treat patients with XLH in Europe. For each participant, relevant experience with regards to the inclusion criteria was recorded.

Study procedure and design

Item-generation questionnaire (Round 1)

Surveys were constructed and distributed electronically via email using Survey Monkey [5]. During Round 1, the item-generation round, participants were asked to complete an open-ended question, using up to 10 blank fields for their answers. For example, in Case study 1, participants were asked to identify specific points which they would recommend other teams focusing on when running a registry for patients with XLH by answering a single open ended question: *'With specific reference to your experience in setting up and implementing the XLH Registry what recommendations would you make to other groups wishing to implement a similar European rare-disease registry?'* The survey allowed for free text responses, and participants submitted their recommendations or responses using terms and phrases of their own choosing. Each participant was given 6 hours to complete the exercise and asked to submit up to 10 recommendations. For the other case studies, participants were allocated a pre-agreed amount of time to complete Round 1.

For each case study, responses to the item-generation round were collated into themes and developed into mutually exclusive items by the investigator. Responses to Round 1 were analysed using a process of content analysis and open coding to categorize items into themes.

Item generating using literature reviews

Where appropriate, literature reviews were used to supplement item generation by the participants. Specific terms were used to search an online database (MEDLINE via PubMed; Table 1). The search results were manually reviewed and the abstracts of relevant articles were coded using the same technique as the participant responses to the list-generation question. The frequency of each item generated from the literature review was not recorded. Items generated from the literature review that were not identified by the participant responses to the Round 1 question were included in Round 2.

Structured questionnaire to observe consensus (Round 2)

The compiled list of items/themes were reviewed and refined by the investigator and included in Round 2 as structured questionnaires (Supplementary Materials). The investigator had the opportunity to add items/themes not originally included as part of the response to Round 1 in order to test prompted agreement. Similarly, items from a systematic literature review could be added if not identified by the participants in their responses to the Round 1 question. Participants who completed Round 1 were asked to complete the Round 2 questionnaire. Each participant was asked to state his or her level of agreement with each statement on a five-point Likert scale (strongly agree, agree, neither agree or disagree, disagree or strongly disagree). Responses to the structured questionnaire in Round 2 were used to determine observed consensus, proportional group awareness and the effect of prompting.

Analysis plan

Participants were asked to include their names in the surveys for administration purposes. This was to ensure that all participants completed the relevant surveys. Responses to surveys were anonymized before analysis.

Four case studies were considered sufficient to test the methodology. Consensus was defined as agreement of more than half of the participants, i.e., a simple majority was used to confirm consensus. Data from the item-generation round and the consensus round were exported from SurveyMonkey to a Microsoft Office Excel spreadsheet for analysis.

Awareness Index

The responses to the Round 1 questionnaire were used to assess knowledge awareness by calculating the frequency of each coded item in relation to the overall most frequently occurring coded item (the Awareness Index). A minimal awareness threshold was defined as an Awareness Index >50%. If items generated from a literature review were not mentioned by the participants' responses in Round 1, then there was no awareness of the item by the participants.

Consensus Index

The *Consensus Index* was calculated as a percentage of participants who agreed or strongly agreed with each statement.

Index Score to measure prompting

The *Index Score* was used to measure prompting during Round 2. The concept of prompting was pre-specified to have occurred if the absolute difference between the *Awareness Index* (calculated using the

item-generation responses in Round 1) and the *Consensus Index* (calculated using the percentage of participant agreement with each statement during the structured questionnaire in Round 2) was greater than or equal to 0.05 or 5% whereas if it was less than 0.05 (5%) then there was no prompting. Unprompted consensus was defined when the majority of participants suggested the item during Round 1, and the majority of participants agreed or strongly agreed that the item was important in Round 2. Any item that was not suggested by the participants during Round 1, but the participants agreed that the item was important in Round 2 was said to be completely prompted.

Results

The case studies were conducted between March 2017 and December 2019. The number of participants ranges between 3 and 10 for each case study. A description of the participants for each case study is summarized in Table 2.

In Round 1, all 10 blank fields were used by the participants in case study 1. For the other case studies, up to six blank fields were completed in each study.

Systematic literature review

Literature reviews were conducted for case studies 2 and 3 and no additional items were generated for these case studies.

Responses to Item-generation (Round 1) and structured (Round 2) questionnaires

In case study 1, recommendations for setting up a European XLH registry, Round 1, five key areas were identified by the participants: engagement, regulatory, communications, relationships and internal procedures. The most frequently mentioned coded items were 'positive initial investigator engagement' and 'positive relationship with sites' and were mentioned twice by each participant in their free-text responses to the item-generation round. The least frequently mentioned coded items were mentioned once each (Figure 2).

In case studies 2 and 3 (limitations of treatment for XLH in either children or adults), the participants also identified five key areas in response to the Round 1 question: persistence of treatment, convenience, adherence, efficacy and safety, although there were differences in the frequency of each item in each area depending whether treatment referred to children or adults (Figures 3 and 4). Case study 4 looked at triggers for treating XLH, which showed a spread of awareness of items generated and also a spread of consensus of each statement (Figure 5).

Awareness (item generation) and prompted agreement (consensus)

When presented with the structured questionnaire in Round 2, many of the participants were agreed or strongly agreed that the items generated during Round 1 were important. The majority of items had some degree of prompted agreement consensus. Participants were prompted to consider the majority of items, as can be seen from the gradients of each line visualizing the absolute difference (*Index Score*) between the *Awareness Index* and the *Consensus Index* in Figures 2–5. The *Index Score* for each item was also numerically calculated (data not shown).

Discussion

This series of case studies have shown how to methodically observe proportional group awareness and subsequent consensus of items arising from list-generating questioning. The free-text responses to the item-generation round tested the awareness of specific concepts or items regarding setting up a European XLH registry, the limitations of empirical treatment for XLH in children and adults, and triggers for treatment of XLH in adults. The purpose of this methodology is not to force a consensus, but to evaluate the current knowledge and opinions of the selected experts. It is unique in its objectiveness, its approach and interpretation compared with traditional forms of consensus generating methodologies.

This method of observing proportional group awareness and consensus consists of two rounds of questionnaires. The first survey has to consist of an open-ended question for the experts to answer. The answers then can be used to generate listed items and assess the awareness of the experts of a particular subject, recording the awareness of items in an unprompted way. The structured questionnaire generated from the items identified from the first round can also have items added to it by the investigator(s). As such, the process of observing proportional group awareness and consensus is flexible, allowing results from literature searches to input into the structured questionnaire in Round 2. This enables the participants to have the benefit of reviewing items that may not have been suggested in their first round.

In case study 1, setting up a European XLH registry, recommendations were elicited from small number of experts with specific reference to a single project. The unprompted consensus items of a positive initial engagement with a country chief investigator and developing positive relationship with sites show that these are at the forethought of clinical study managers minds when thinking about setting up a successful registry. Nevertheless, participants had minimal awareness of the majority of items identified during Round 1. The high number of items that were elicited with minimum observed frequency maybe driven by the low number of advisors ($n = 3$) and variation in individual opinion. Nevertheless, this case study demonstrated the ability to utilize the minimum of three advising participants as well as the response to the 'test' item that was introduced.

The case studies for empirical treatment of XLH in both children and adults demonstrated the different levels of knowledge or awareness of treatment options. For treating children, the highest level of awareness of items generated indicated that the panel of experts focused on the same items, including adherence to treatment, problems of variable serum phosphate levels and inadequate response to healing

rickets. The participants' responses enabled a small number of clear items generated for empirical treatment of XLH in children. This may be due to good alignment in level of experience of the experts or that the question is well defined and enables a small number of items to be generated. Items with an Awareness Index <50% would benefit from effort to increase awareness through education. In contrast, the same question asked about the treatment of adults demonstrated a low level of awareness of items identified by the expert participants. This may represent the diffuse knowledge of the subject, or genuinely disparate positions on the items to be included. However, during Round 2 in both case studies of the limitations of empirical treatment of XLH, the participants showed a high level of agreement consensus. In case study 4, awareness of triggers for treatment of XLH in adults was low, except for pain, which had unprompted consensus as being an important trigger for treatment. A wide variation in opinion through divergent local practices or low familiarity with the subject matter may be responsible for the low awareness of other items generated.

In the case studies reported here, the first had participants with specific knowledge and experience setting up an XLH registry, whereas the other case studies all had participants who were experts in treating patients with XLH and had published articles about XLH in peer-review journals. Expertise required depends on the clinical question being asked. As such, experts could be defined in a clinical setting by the number of patients they treat per year with a specific disease, or the number of patients enrolled in a clinical trial, or other discerning attribute that makes the person an expert in their field and able to provide value insights to the question in hand.

In order to have an opinion from which to form a consensus, a group of experts would first need to be aware of the key aspects of the subject of interest on which to base that opinion. Without one there cannot be the other. A threshold level of awareness of their subject matter would, by definition, qualify them as an expert in it. However, accepting that no expert can omniscient, it would follow that being made aware of new material would inevitably require him or her to form an opinion on it that may influence the overall consensus. As such, there is necessarily internal (within the expert group) and external prompting (external sources outside the expert group) when it comes to forming an opinion.

The expert responses to the structured questionnaire allows the investigator(s) to observe any consensus that arises, regardless of whether it is prompted or unprompted. In these case studies, prompted agreement was pre-specified as a difference of 5% or more in the *Index Score* (the absolute difference between the *Awareness Index* and the *Consensus Index* for an item. Traditionally, 5% has been considered a threshold for statistical significance. It was recognized that the prompting was going to be 'coarse' given the small number of participants, particularly in case study 1; however, the observation of prompting could still be made on data collected from Round 1 and Round 2. Some items may have significantly more support after prompting in Round 2 but fail to secure majority agreement by the participants. Nonetheless, the advantage that anonymity of the participating experts is never lost, therefore there is no effect of dominant individuals, manipulation or compulsion to confer to certain viewpoints [6, 7]. Moreover, anonymity preserves the independence in item generation during Round 1.

In these case studies, the lack of face-to-face meetings denied the opportunity to discuss pertinent topics, which may have been seen as a limitation; however, there is no reason why the process could not include a face-to-face discussion meeting to be held after Round 1 and before the experts complete the Round 2 questionnaire.

The ability to categorize items on the basis of the difference between the *Awareness Index* and *Consensus Index* enables appropriate handling of these items (Figure 6). For example, an item with both an *Awareness Index* and *Consensus Index* of 100% denotes complete awareness of an item which does not undergo prompting and resulted in complete consensus. Such an item can be considered to be well understood or to be common knowledge and beyond challenge. However, recognising that differences in awareness of suggested items across the experts will exist identify items of lower awareness. For items with an *Awareness Index* <50%, the degree of prompting observed in the performance of these items in Round 2 indicates how these items are appreciated when uniformly presented, i.e., when all experts are made equally aware of the item (as seen by the gradient of the slope between Round 1 and Round 2, or the *Index Score*). The performance of these items around the simple majority threshold of >50% provides insight into the 'tipping' point of broad agreement. Items which are categorized as having a *Consensus Index* <50% indicates a degree of education may be required in order to increase awareness of an accepted item (Figure 7). When an item is not suggested by the experts and introduced into Round 2 as either a test suggestion or items gleaned from a literature review which are then prompted to <50% may indicate the experts lack a threshold level of knowledge or awareness of the subject matter. Alternatively, items falling into this category may constitute a fundamental 'challenge' to the norm and indicate an opportunity to redefine these norms. The consensus list can allow items to be utilized in further discussions without further investigation or used with a parallel investigation, or alternatively not used without formal investigation. Finally, in situations where there is no consensus in Round 2, the threshold for active investigation may not be met and the research team may simply opt to monitor these items and defer decision on active management until more is known about them.

Observed proportional group awareness and consensus is a relatively quick process compared with the Delphi technique and its variants, which can be long and drawn out depending on the pre-specified level of agreement required. The Delphi technique often shows a change in participants' views towards consensus and stability with each subsequent round as indicated by an increase in percentage agreements, convergence of importance rankings, an increase in Kappa values and a decrease in comments as rounds progressed [8].

A potential limitation of the case studies reported here is that the Round 2 questionnaires were constructed in a way that encouraged agreement rather than disagreement, possibly leading to response bias, in particular, acquiescence bias. To mitigate this and test for acquiescence bias, in future studies Round 2 surveys could have a selection of questions that repeat the same concept or item but are reworded so that if the participant strongly agreed with the first statement, then they would be likely to strongly disagree with the repeated statement. For example, 'Pain reported by the patient is the most important deciding factor with regards to treatment' and 'Pain reported by the patient does not affect the

type of treatment given'. Nevertheless, it might not always be appropriate to phrase statements in this way as the whole process is directed towards agreement consensus; the responses to the open-ended question in Round 1 are coded and lead to the statements listed in Round 2. As such, the participants who have suggested specific items in Round 1 are likely to agree with statements based on these items. In addition, the open-ended question asked in Round 1 needs to be carefully considered to ensure that with the objectives of the study will be able to be fulfilled. Similarly, expert participants need to be chosen carefully.

The process of observing proportional group awareness and consensus takes account of knowledge awareness of the experts at the beginning of the consensus process, unlike the Delphi technique, modified Delphi and RAND/UCLA. There was no forced consensus, no attrition of participants, no sending questionnaires to a wider group of academics who may not have the same level of knowledge and making the process more transparent. Although the Delphi technique is useful for determining judgement convergence, this does not necessarily mean that the resulting consensus is accurate [9]. If participants fail to answer a question or questions in any particular round in Delphi, the results of that round may show a misleading bias towards agreement [8]. Participants in the case studies reported here had similar levels of knowledge so it was expected that they would have similar levels of prompted agreement for each item. Participants from different specialties would be more likely to have different levels of prompted agreement for each item identified during the list-generating round. Although the number of participants was relatively small in each case study, other studies have shown that only a small number of experts is required when assessing consensus [10].

The method of observing proportional group awareness and consensus could be used to answer most questions, for example, in therapy areas where information is lacking, making us reliant on expert opinion (level 5 evidence). Observing proportional group awareness and consensus would also be useful when a specific subjective interpretation is required to answer the research question, when assessing differing levels of knowledge and awareness of key information in order to develop education programmes, when assessing the time delay between expert knowledge and when such knowledge becomes known by the wider community, and in therapy areas in which randomized controlled trials are considered unethical (for example, some areas of obstetrics or oncology) thus limiting the quality of evidence available. These kinds of surveys could also be used to assess the level of knowledge of different groups of advisors who answer the same Round 1 question, which could be useful when developing regional education programmes. Similarly, repeating the Round 1 or Round 2 surveys would measure how knowledge about a specific subject changes over time. The value of observing proportional group awareness and consensus is that it elucidates unmet educational need, knowledge gaps, and the time-lag between expert knowledge (unpublished) and when it leaches to the broader academic community. Thus, the clear benefit of the unprompted list generation and group consensus methodology is that it helps assess the current level of understanding and the temporal lag between research and expert knowledge in a specific therapy area or topic. The difference between the unprompted awareness and the prompted agreement consensus identifies items or areas either where more education is required (in the case of items not

listed by the participants, or listed by very few of them in Round 1) and where more research is required (in the case of where there is no observed agreement consensus for a statement in Round 2).

Conclusions

In summary, the methodology described here to observe proportional group awareness and consensus provides objective assessment of expert knowledge, which is applicable to research areas where opinions may be subjective or based on limited evidence, especially in research areas where other consensus techniques may not be appropriate to apply. Standardized categorization of items with regard to awareness, consensus and prompting, offer the opportunity manage each item or concept in a more tailored manner in terms of education and further investigation.

Abbreviations

XLH, X-linked hypophosphatemia.

Declarations

Ethics approval and consent to participate

All participants were specialist doctors and gave informed written consent to provide expert opinion about X-linked hypophosphatemia. No patient data were used in the study, thereby negating the need for ethics approval.

Consent for publication

Not applicable.

Availability of data and material

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

RJ is the founder of Medialis, Ltd. Medialis conducted the Case Studies observing proportional group awareness and consensus. All study managers (AD, JM and JT) who were experts for the case study 1, are employees of Medialis. All other participants who took part in case studies 2, 3 and 4 were paid for their time.

Funding

The case studies were conducted alongside but independent to paid services Medialis, Ltd was contracted to perform by one of its clients and the case studies themselves remain wholly funded by Medialis Ltd. Medical writing and editorial support were funded by Medialis, Ltd.

Author contributions

RJ was completely responsible for the conception, design of the work, the acquisition, analysis, interpretation of data and has drafted the work and substantively revised it. RJ has also approved the submitted version and has agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved, and the resolution documented in the literature.

Acknowledgements

Celia J Parkyn, PhD, provided medical writing and editorial support for manuscript preparation, and was funded by Medialis, Ltd.

References

- 1.Hsu C-C, Sandford BA: *The Delphi Technique: making sense of consensus. Practical Assessment, Research & Evaluation* 2007, *12*(10).
- 2.Goodman CM: *The Delphi technique: a critique. J Adv Nurs* 1987, *12*(6):729–734.
- 3.Custer RL, Scarcella JA, Stewart BR: *The Modified Delphi Technique - A Rotational Modification. Journal of Career and Technical Education* 1999, *15*(2).
- 4.Fitch K, Bernstein SJ, Aguilar MD, Burnand B, LaCalle JR, Lazaro P, Loo Mvh, McDonnell J, Vader JP, Kahan JP: *The RAND/UCLA appropriateness method user's manual. Available from https://www.rand.org/pubs/monograph_reports/MR1269.html. Accessed May 2, 2019.* In.; 2001.
- 5.SurveyMonkey, Inc. San Mateo, California, USA. Available from: <http://www.surveymonkey.com>.
- 6.Kirkley A, Alvarez C, Griffin S: *The development and evaluation of a disease-specific quality-of-life questionnaire for disorders of the rotator cuff: The Western Ontario Rotator Cuff Index. Clin J Sport Med* 2003, *13*(2):84–92.
- 7.Walker A, Selfe J: *The Delphi method: a useful tool for the allied health researcher. . Int J Ther Rehab* 1996, *3*(12):677–681.

- 8.Holey EA, Feeley JL, Dixon J, Whittaker VJ: *An exploration of the use of simple statistics to measure consensus and stability in Delphi studies. BMC Medical Research Methodology* 2007, 7(1):52.
- 9.Cousien A, Obach D, Deuffic-Burban S, Mostafa A, Esmat G, Canva V, El Kassas M, El-Sayed M, Anwar WA, Fontanet A *et al: Is expert opinion reliable when estimating transition probabilities? The case of HCV-related cirrhosis in Egypt. BMC Medical Research Methodology* 2014, 14(1):39.
- 10.Akins RB, Tolson H, Cole BR: *Stability of response characteristics of a Delphi panel: application of bootstrap data expansion. BMC Medical Research Methodology* 2005, 5(1):37.

Tables

Table 1. Summary of case studies

Case study	Open-ended question asked in Round 1 (list-generating round)	Criteria for expert participation	Systematic literature review (search terms)*
1. Setting up a European registry for patients with XLH	With specific reference to your experience in setting up and implementing the XLH Registry what recommendations would you make to other groups wishing to implement a similar European rare-disease registry?	<p>Recruitment was from the Medialis Real World Evidence Asset Management Division team of clinical study managers.</p> <p>Relevant patient registry experience.</p> <p>Participants have a similar level of knowledge regarding rare-disease registries, specifically an XLH registry, and exposure to the same project information.</p>	No
2. Limitations of empirical treatment of XLH in children	What are the limitations of empirical treatment for XLH in children?	<p>Experience in treatment of children with XLH.</p> <p>Management of XLH in a recognized specialist centre. Author of publication in paediatric XLH. Participation in paediatric clinical trials in XLH.</p>	<p>X-linked rickets; XLH; X-linked hypophosphatemia; Familial hypophosphatemia; Familial hypophosphatemic rickets; X-linked hypophosphatemic rickets; Vitamin D-resistant rickets; VDRR; Genetic rickets; Hereditary rickets</p>

			Phosphate(s) and vitamin D (calcitriol, 25-hydroxycholecalciferol, 1,25-dihydroxycholecalciferol, cholecalciferol, vitamin D3/D2/D)
3. Limitations of empirical treatment for XLH in adults	What are the limitations of empirical treatment for XLH in adults?	Experience in treatment of adults with XLH. Management of XLH in a recognized specialist centre. Author of publication in adult XLH. Participation in adult clinical trials in XLH.	X-linked rickets; XLH; X-linked hypophosphatemia; Familial hypophosphatemia; Familial hypophosphatemic rickets; X-linked hypophosphatemic rickets; Vitamin D-resistant rickets; VDRR; Genetic rickets; Hereditary rickets Phosphate(s) and vitamin D (calcitriol, 25-hydroxycholecalciferol, 1,25-dihydroxycholecalciferol, cholecalciferol, vitamin D3/D2/D)
4. Triggers for treatment of XLH in adults	What are your triggers for treatment of XLH in adults?	Experience in treatment of adults with XLH. Management of XLH in a recognized specialist centre. Author of publication in adult XLH. Participation in adult clinical trials in XLH.	No

*Online databases MEDLINE via PubMed were used to identify search terms. XLH, X-linked hypophosphatemia.

Table 2. Case study participants

Case study number	Description of participants	Number of participants who responded to:	
		Round 1	Round 2
1. Setting up a European XLH registry	The participants all worked on a European XLH patient registry. In addition, two of the three participants had more than 10 years of experience in study management and all three participants had at least 5 years of experience in national regulatory submission processes. One of the participants had additional experience in local site regulatory approval (3 years), site contract execution (8 years), and engagement of chief (n=8) and primary investigators (n=32).	3	3
2. Limitations of empirical treatment of XLH in children	The participants were all physicians with experience of treating children with XLH, in Europe. All worked in a national centre recognized for the management of paediatric XLH patients and had authored articles on XLH in the paediatric population	8	8
3. Limitations of empirical treatment for XLH in adults	The participants were all physicians with experience in treating adults with XLH in Europe. All worked in a national centre recognized for the management of paediatric XLH patients in their country and had authored articles on XLH in the paediatric population.	10	10
4. Triggers for treatment	The participants were all physicians with experience in treating adults with XLH in Europe. All worked in a national centre recognized for the	10	10

of XLH in adults	management of paediatric XLH patients in their country and had authored articles on XLH in the paediatric population.		
------------------	---	--	--

XLH, X-linked hypophosphatemia.

Figures

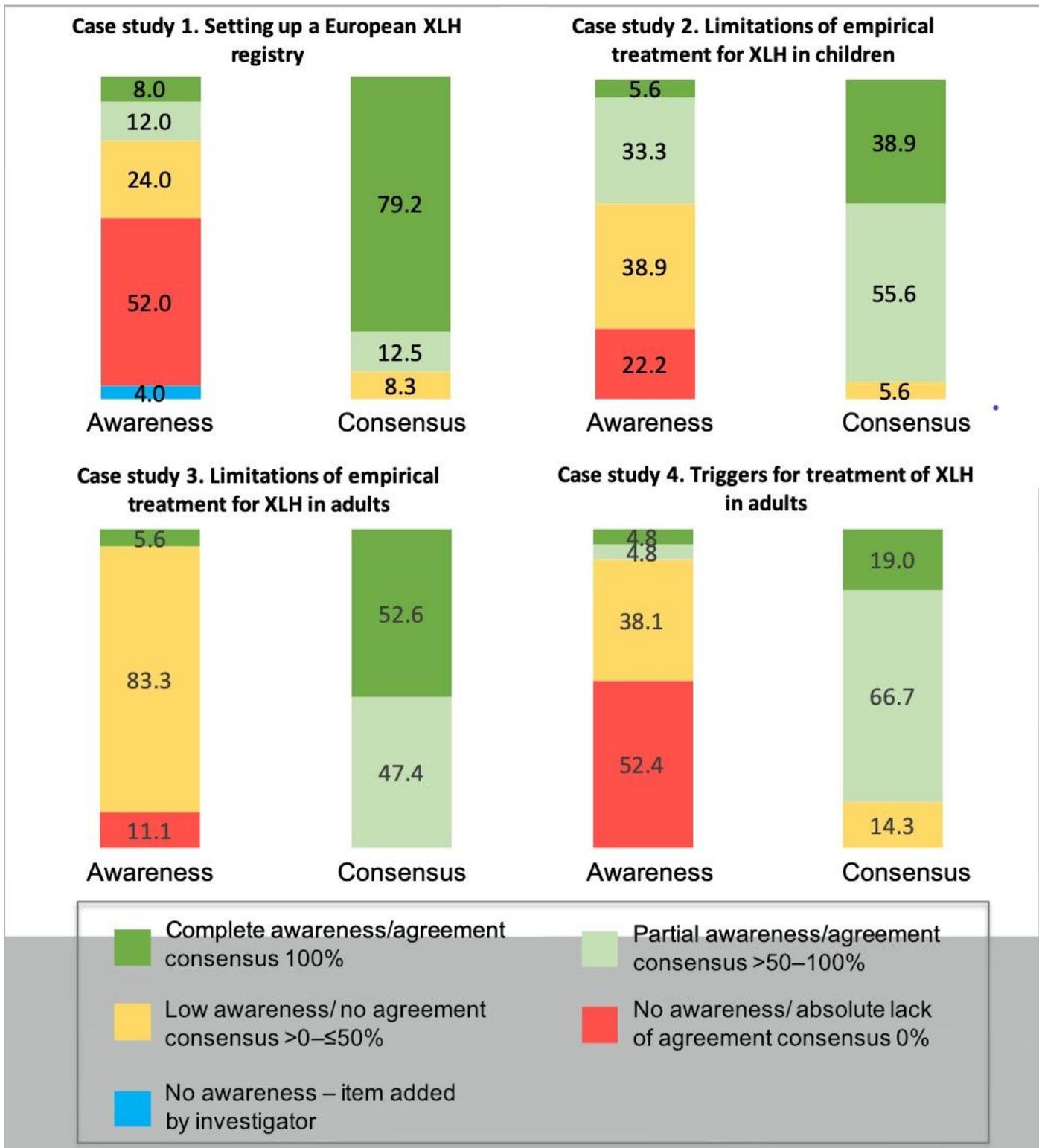


Figure 1

Figure 1. Overview of awareness and agreement consensus The awareness and agreement consensus overview is categorized according to the Awareness Index (frequency of each coded item in relation to the overall most frequently occurring coded item in Round 1) and the Consensus Index (calculated using the percentage of participant agreement with each statement during the structured questionnaire in Round 2). Complete awareness of an item has an Awareness Index of 100% and complete agreement consensus

has a Consensus Index of 100%. Partial awareness and consensus have indices >50–100%. Low awareness and no agreement consensus have indices 0–≤50%. XLH, X-linked hypophosphatemia.

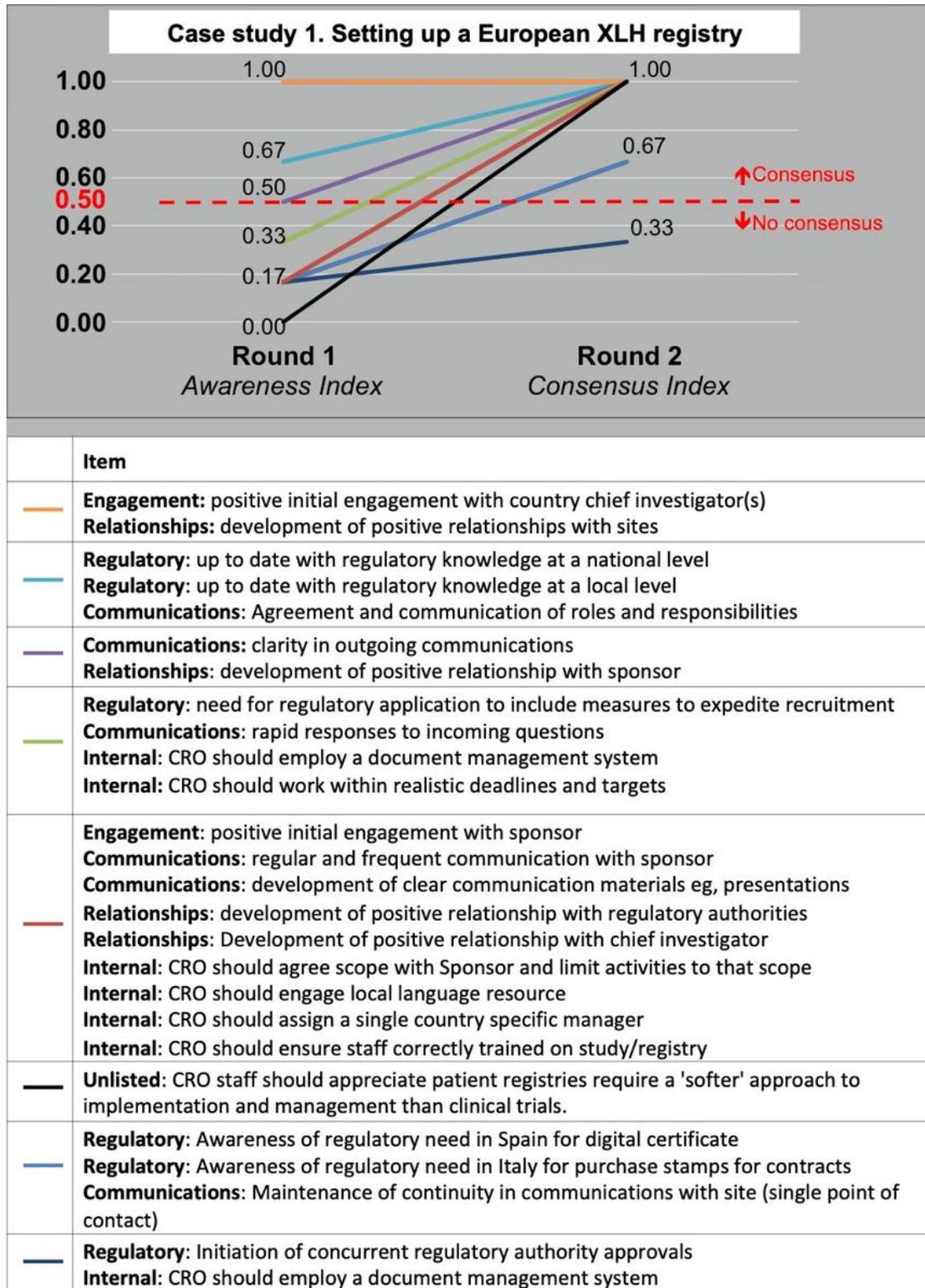


Figure 2

Figure 1. Overview of awareness and agreement consensus The awareness and agreement consensus overview is categorized according to the Awareness Index (frequency of each coded item in relation to the overall most frequently occurring coded item in Round 1) and the Consensus Index (calculated using the

percentage of participant agreement with each statement during the structured questionnaire in Round 2). Complete awareness of an item has an Awareness Index of 100% and complete agreement consensus has a Consensus Index of 100%. Partial awareness and consensus have indices >50–100%. Low awareness and no agreement consensus have indices 0–≤50%. XLH, X-linked hypophosphatemia.

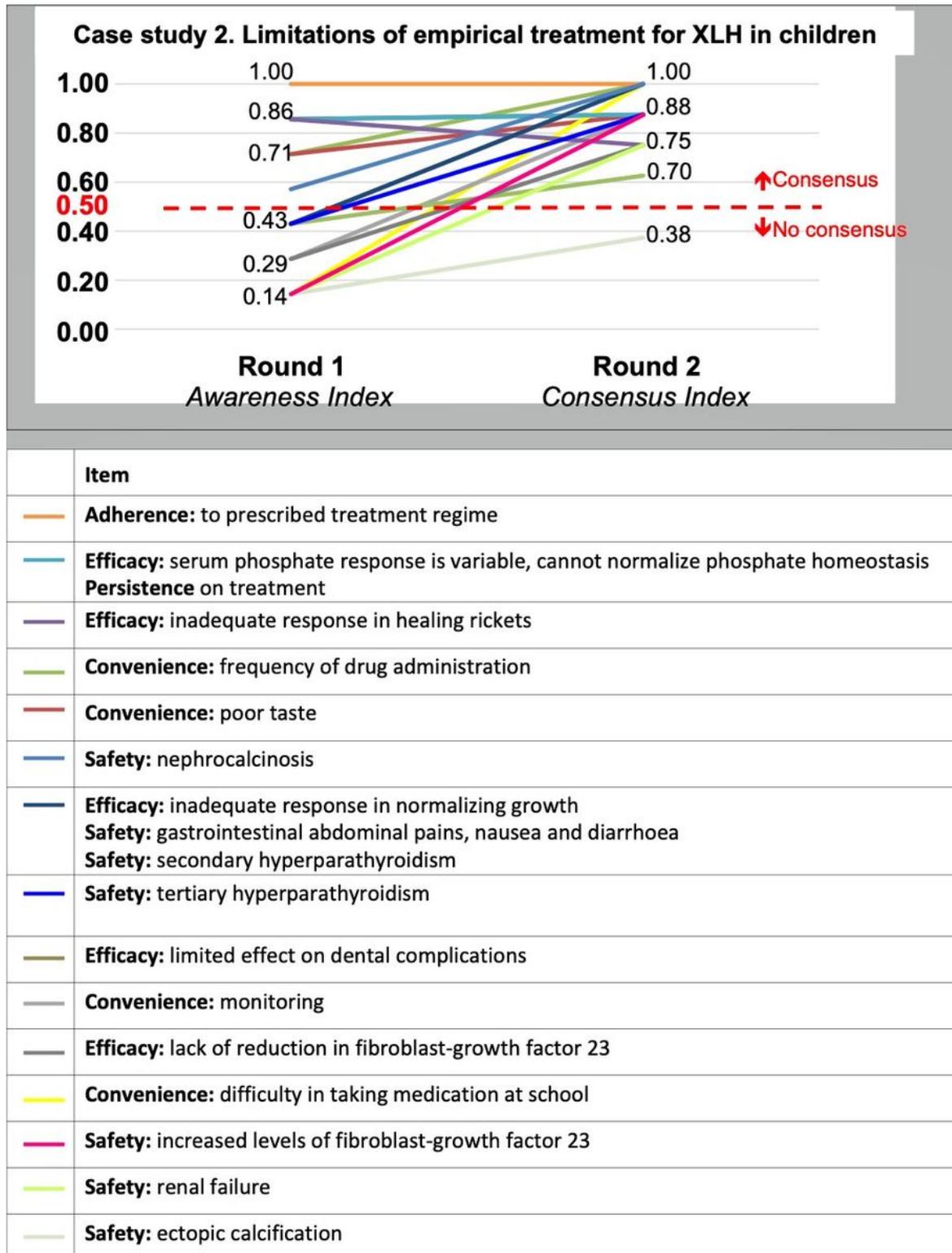
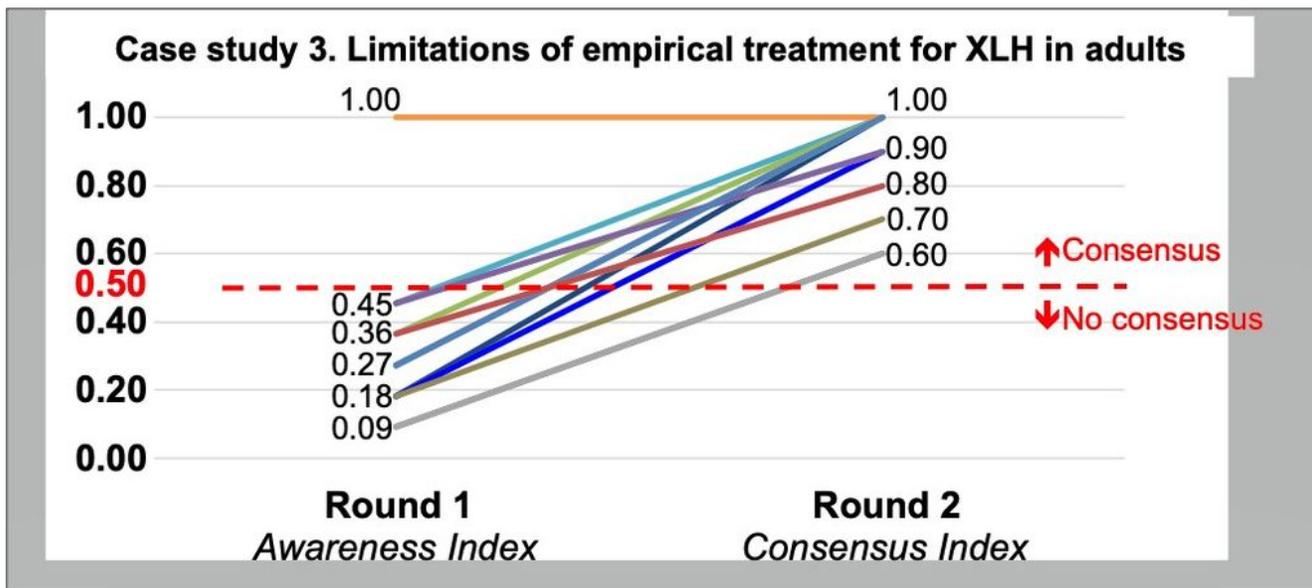


Figure 3

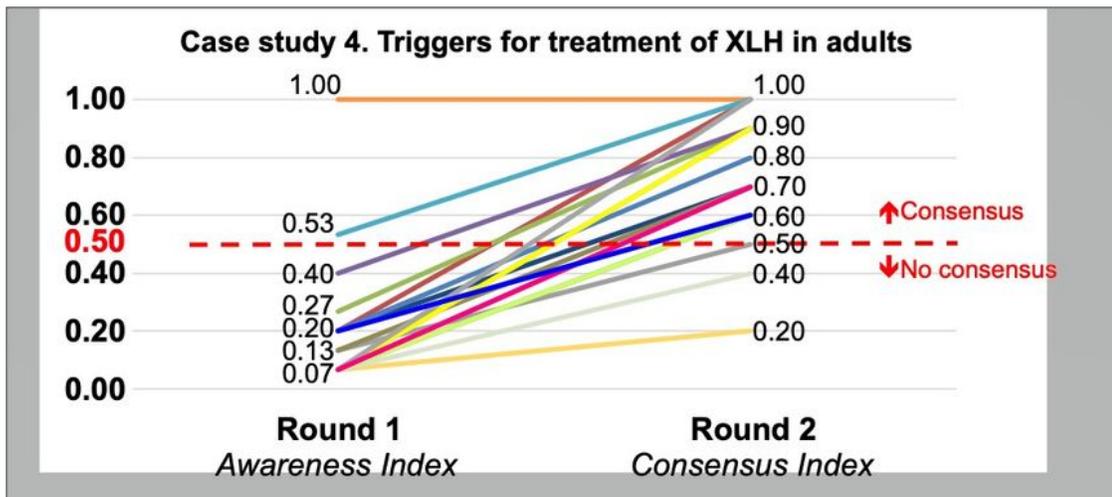
Figure 3. Awareness (item generation) and prompted agreement (consensus) for case study 2



Item
— Safety: gastrointestinal abdominal pains, nausea and diarrhoea
— Convenience: frequency of drug administration
— Safety: secondary hyperparathyroidism Safety: tertiary hyperparathyroidism
— Efficacy: serum phosphate response is variable, cannot normalize phosphate homeostasis
— Safety: renal failure Safety: hypercalcaemia
— Efficacy: inadequate response in fracture healing Efficacy: inadequate response in reducing pain Safety: nephrocalcinosis Adherence: to prescribed treatment regime Convenience: regular monitoring
— Persistence on treatment
— Efficacy: inadequate response reducing stiffness Efficacy: lack of reduction in fibroblast-growth factor 23
— Safety: ectopic calcification
— Convenience: poor taste Convenience: life-long treatment

Figure 4

Figure 4. Awareness (item generation) and prompted agreement (consensus) for case study 3



Item
Pain
Osteomalacia: fractures
Low quality of life
Planned orthopaedic surgery
Osteomalacia
Nephrocalcinosis
Hypophosphatemia
Enthesopathy Secondary hyperparathyroidism
High alkaline phosphatase levels
XLH diagnosis
Fracture non-union
Muscle weakness Pain that is unresponsive to simple analgesics
Fatigue Decreased physical function Stiffness
Skeletal deformity High fibroblast growth factor 23
Dental abscess
Short stature

Figure 5

Figure 5. Awareness (item generation) and prompted agreement (consensus) for case study 4

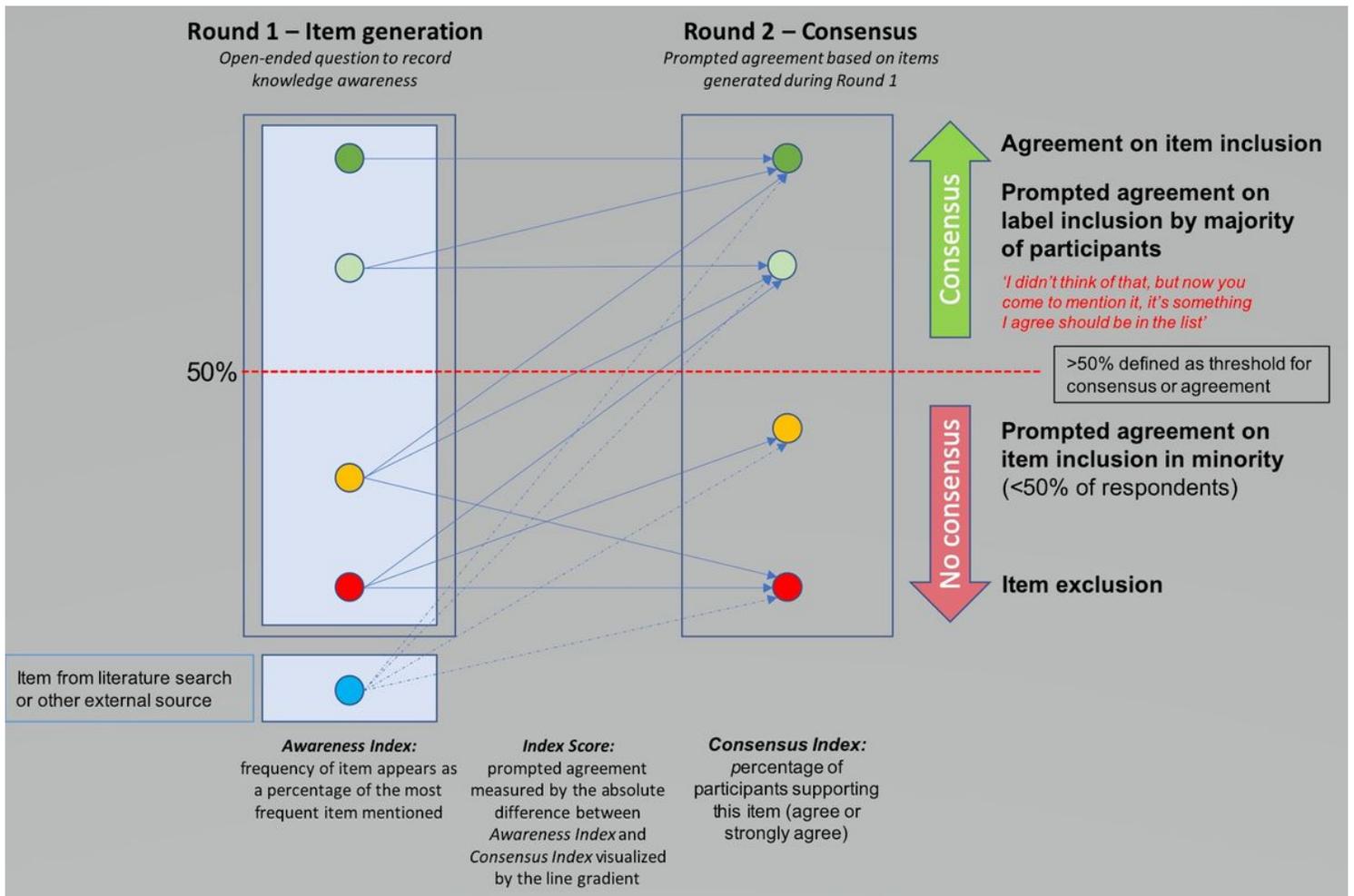


Figure 6

Figure 6. Insights into stratifying knowledge awareness and prompted agreement Interpreting differences between group awareness and consensus of items. Round 1 – Item Generation: Green circles represent group awareness of items identified during a list-generating question; orange circle represents items are possibly require more education about or represent a time-lag between expert awareness and the knowledge spreading to the wider academic community; red represents items that are possibly not as relevant to the study question; blue represents external source of item, for example, an item from a systematic literature review. Round 2 – Consensus: Green circles represent agreement or prompted agreement; orange circle represent prompted agreement but in the minority of participants, possibly because more research is required to provide more evidence; red circle represents agreement that item should be excluded. Arrows between Round 1 and Round 2 represent the level of prompted agreement with the steeper the slope the greater the level of prompted change of opinion. The Index Score represents the amount of prompted agreement between awareness (item-generating) and observed consensus.

Round 1: Awareness Awareness Index ($f_{\text{item}}/f_{\text{max}} \times 100$)		Potential future actions	Round 2: Consensus Consensus Index ($n_{\text{support}}/n_{\text{total}} \times 100$)		Potential future actions	Prompting Index Score, the absolute difference between Awareness Index and Consensus Index	
A1	Maximum awareness AI: 100%	<p>Item(s) require increasing amounts of education/promotion</p>	C1	Maximum consensus CI: 100%	Utilize in education programmes	P1	No prompting <5% difference in AI and CI
A2	High awareness AI: >50%–<100%		C2	High consensus majority in majority CI: >50–<100%	Utilize in education programmes and/or further investigate to provide more evidence for item(s)	P2	Prompting Difference between AI and CI $\geq 5\%$
A3	Low awareness AI: between $>[f/f_{\text{min}} \times 100]$ and $\leq 50\%$		C3	Low consensus in minority CI: >0–<50%	Further investigate to provide additional evidence for item(s)		
A4	Minimum awareness AI: $f/f_{\text{min}} \times 100$		C4	Minimum consensus CI: 0%	Monitor future changes in awareness. Complete lack of consensus suggests no evidence to support item		
A5	No awareness AI=0%					P3	Absolute prompting When item not suggested in Round 1, but introduced by investigator, prompting can be up to 100%

Figure 7

Figure 7. Planning future actions based on awareness and consensus categories Group awareness of items and prompted agreement consensus can be used to plan future actions according to pre-defined categories.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Round2QsCasestudy1.pdf](#)
- [Round2QsCasestudy2.pdf](#)
- [Round2QsCasestudy3.pdf](#)
- [Round2QsCasestudy4.pdf](#)