Journal of Clinical Epidemiology 128 (2020) 118−129

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

‘‘Vicarious thinking’’ was a key driver of score change in Delphi surveys for COS development and is facilitated by feedback of results

Rebecca Fish[a,b,1,*], Steven MacLennan[b,1], Bilal Alkhaffaf [a,c,d], Paula R. Williamson[e]

[a]*Division of Cancer Sciences, School of Medical Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Cancer Research Centre, Wilmslow Road, Manchester M20 4BX, UK*

[b]*Colorectal and Peritoneal Oncology Centre, Christie NHS Foundation Trust, 550 Wilmslow Road, Manchester M20 4BX, UK*

[c]*Academic Urology Unit, Institute of Applied Health Sciences, University of Aberdeen, Foresterhill, Aberdeen AB25 2ZD, UK*

[d]*Department of Oesophago-Gastric Surgery, Salford Royal NHS Foundation Trust, Stott Lane, Salford M6 8HD, UK*

[e]*MRC/NIHR Trials Methodology Research Partnership, Department of Biostatistics, Block F Waterhouse Building, University of Liverpool, 1-5 Brownlow Street, Liverpool L69 3GL, UK*

Accepted 9 September 2020; Published online 1 October 2020

Abstract

Objective: The objectives of this nested study were to (1) assess whether changes in scores between rounds altered the final degree of consensus achieved in three Delphi surveys conducted as part of COS development projects (anal, gastric, and prostate cancer), and (2) explore participants’ reasons for changing scores between rounds.

Study Design and Setting: All Delphi surveys were conducted online using DelphiManager software and included healthcare professionals and participating patients. Participants were invited to give a free-text reason whenever they changed their score across an important threshold on a 1−9 Likert scale (1−3 not important, 4−5 important, 7−9 critically important). Reasons for score change were coded by four researchers independently using an inductive-iterative approach.

Results: In all three Delphi surveys, the number of outcomes reaching criteria for consensus was greater in R2 than R1. Twelve themes and 23 subthemes emerged from 2298 discrete reasons given for score change. The most common reasons for the change were ‘‘time to reflect’’ (482 responses, 23%) and vicarious thinking (424, 21%), with 68% (291) of vicarious thinking attributed to seeing other participant’s scores.

Conclusion: Our findings support conducting a Delphi survey over the use of a single questionnaire where building consensus is the objective. Time to reflect and vicarious thinking, facilitated by seeing other participant’s scores, were important drivers of score change. How results are presented to participants between rounds and the duration of and time between rounds in a Delphi survey may, therefore, influence the results and should be clearly reported. © 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Keywords: Core outcome set development; Delphi survey; consensus; Stakeholders; Feedback methods; patient public involvement

1. Background and aims

Core outcome sets (COS) are increasingly being advocated as a means to ensure the relevance of research outcomes to stakeholders, reduce outcome heterogeneity, and minimize reporting bias [1−4]. Trial funding bodies, regulatory authorities, and guideline development groups, such as the (UK) National Institute for Health Research, the European Medicines Agency, and the (UK) National Institute for Health and Care Excellence, now actively endorse the use of COS [5]. The COMET Initiative promotes rigorous consensus methods involving key stakeholders for the development of COS [6], and the consensus derived COS-STAD (core outcome set standards for development) recommendations [7] describe a set of minimum standards for COS development projects. Consensus methodology appropriate to the context is advocated [8], although a 2019 systematic review of COS studies [9] found that 77% of all COS studies published in 2018 included a Delphi survey, increasing from 31% in 2013−2015 and 15% in

[1] Joint first authors.

* Correspondence author: Colorectal and Peritoneal Oncology Centre, Christie NHS Foundation Trust, 550 Wilmslow Road, Manchester M20 4BX, UK. Tel.: 0161 446 8051; fax: 0161 918 7078.

E-mail address: rebecca.fish-2@manchester.ac.uk (R. Fish).

https://doi.org/10.1016/j.jclinepi.2020.09.028

0895-4356/© 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

What is new?

Key findings

- In three Delphi surveys for COS development, most participants changed their score for at least one item between rounds, and in all three studies, the changes in score resulted in a greater number of outcomes reaching consensus in round two compared to round one.
- Time for reflecting and trying to understand the importance of an outcome from the perspective of another (vicarious thinking), facilitated by seeing other participant's scores, were important drivers of score change.

What this adds to what is known

- For core outcome set projects where building consensus is the objective, our findings support conducting a Delphi survey over the use of a single questionnaire
- The timing between rounds in a Delphi survey and the format of feedback presented may influence how participants think about their responses and subsequently impact the results.

What is the implication and what should change now

- We recommend that in studies including a Delphi survey, how results are presented to participants between rounds and the duration of and time between rounds are clearly reported so that the potential impact of these characteristics on participation and consensus can be explored.

1981–2013. A Delphi survey is a method of encouraging consensus, allowing participants to change their responses to a questionnaire after reviewing the anonymized summarized responses of other participants (Figure 1).

In a recent questionnaire study [10], Delphi participants reported considering the views of other participants when rescoring items, indicating that feedback of results in a Delphi can influence the scoring in subsequent rounds. However, studies exploring more broadly why Delphi participants choose to change their score between rounds is limited. There also remains uncertainty over whether a Delphi survey with multiple rounds is beneficial to reaching consensus or whether a single round questionnaire could produce the same result.

This nested study aimed to (1) assess whether changes in scores between rounds altered the final degree of consensus achieved and (2) explore participants' reasons for changing scores between rounds in three Delphi surveys conducted as part of COS development projects (anal, gastric, and prostate cancer).

2. Materials and methods

The scope and methodology of the CORMAC [11], GASTROS [12] and COMPACTERS [13] COS studies are summarized in Table 1.

2.1. Format of Delphi surveys

The Delphi surveys for all three COS development studies are summarized in Table 2. All were run using the online DelphiManager platform [14]. The CORMAC, GASTROS, and COMPACTERS Delphi surveys' were conducted in 2017, 2019, and 2014–15, respectively. The CORMAC and GASTROS Delphi surveys involved two rounds, whereas the COMPACTERS Delphi involved three rounds. In the interests of comparability, therefore, only data from R1 and R2 of the COMPACTERS Delphi has been included. In each Delphi round, participants were asked to rate the importance of including each outcome in the COS on a 1–9 scale described as not important (1–3), important but not critical (4–6), and critically important (7–9). In the second round, participants were shown a histogram (CORMAC and

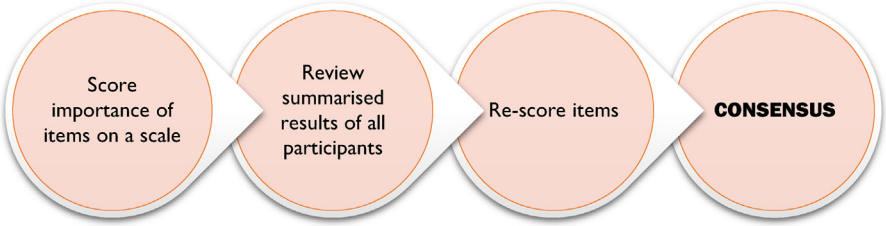

Fig. 1. Schematic representation of how consensus is reached through a Delphi survey.

Table 1. Scope and methodology of core outcome set projects

Element	CORMAC	GASTROS	COMPACTERS
Health condition	Squamous cell carcinoma of the anus/anal canal	Cancer of the stomach	Localized prostate cancer
Setting	Later phase clinical effectiveness trials that will inform clinical decision making	Later phase clinical effectiveness trials that will inform clinical decision making	Later phase clinical effectiveness trials that will inform clinical decision making
Population	Adults >18 years of age	Adults >18 years of age	Men >18 years of age
Types of intervention	Primary treatment with radiotherapy with or without concurrent chemotherapy	Surgery – total or partial gastrectomy	All primary treatments including active surveillance, watchful waiting, surgery, radiotherapy, brachytherapy, cryotherapy, high intensity focused ultrasound, and adjuvant hormonal therapy
Development steps	Systematic review; patient interviews; online e-Delphi; face-to-face consensus meeting	Systematic review; patient interviews; online e-Delphi; face-to-face consensus meeting	Systematic review; patient interviews, online e-Delphi; face-to-face consensus meeting

GASTROS) or a distribution of scoring percentages (COMPACTERS) of the previous round's scores together with their own score for each outcome, before being asked to consider the information presented and score each outcome again (Figure 2). In the CORMAC and GASTROS Delphi surveys, all stakeholders were shown the scores for each stakeholder group separately. In the COMPACTERS Delphi round two, healthcare professionals (HCP) and patients were randomized to receive round one feedback from peers only, multiple stakeholder groups separately, or multiple stakeholders combined. For comparability, therefore, only the group randomized to receive feedback from multiple stakeholder groups separately is included in this study. In the COMPACTERS Delphi, if a participant changed the score for any item from R1 to R2, a free-text pop-up box at the end of the survey asked the participant to describe their reasons for making the change. In the CORMAC and GASTROS Delphi surveys, a free-text pop-up box asked participants to describe the reason for changing their score every time a score was changed over an importance threshold (e.g., 3 to 4, 6 to 7).

2.2. Analysis of reasons for change

Free-text entries from the CORMAC and COMPACTERS Delphi surveys were coded by two researchers (RF, SM) independently using an inductive-iterative approach largely following the framework method outlined by Ritchie and Spencer [15]. The process was that each researcher independently familiarized themselves with the free text reasons for score changes within their own data sets first. Similar reasons were grouped together and assigned a code. The codes were discussed between researchers and the coding scheme refined, and the main themes identified were assigned to "parent" codes with subthemes assigned to "child" codes under each parent. The researchers then applied the coding framework to each other's data, working back and forth across the data and refining the framework until all responses were coded, and both researchers were in agreement. The coding framework was then applied to a sample of the GASTROS data by four researchers (RF, SM, BA, PW) independently followed by discussion and further minor refinement of the code descriptions and framework. The refined framework was then applied to all three data sets by four researchers (RF, SM, BA, PW).

3. Results

The characteristics of the three Delphi surveys are summarized in Table 2.

3.1. Change in score and consensus

The percentage of participants changing the score for at least one outcome between R1 and R2 was 97%, 84%, 92% (anal, gastric, prostate). The percentage of participants changing score across an important threshold for at least one outcome was 90%, 29%, and 60%. The median number of outcomes changed over a threshold per participant ranged from 18 (1–25) in HCPs in the CORMAC Delphi survey to 6 (1–40) in the GASTROS Delphi survey. The changes in score resulted in a greater number of outcomes reaching consensus in R2 than R1 in all three studies.

3.1.1. Coding framework

From 967 participants across the three Delphi surveys, 2,298 responses were received to the free-text box request for a reason for changing score. An overview of the coding framework is shown in Figure 3. Twelve "parent" codes describe the broad themes identified from the participants' free text reasons for the change. Five of the 12 parent codes are further expanded into "child" codes, which describe subthemes emerging within the parent code. Full

Table 2. Delphi survey characteristics

Characteristic	CORMAC	GASTROS	COMPACTERS
Languages	English	English, Chinese, Dutch, German, Italian, Portuguese, Spanish, Turkish	English
Participants counties of residence	Australia, Canada, France, Netherlands, New Zealand, Norway, Spain, Sweden, UK and Ireland, USA	Argentina, Australia, Austria, Azerbaijan, Belgium, Brazil, Cameroon, Canada, Chile, China, Colombia, Costa Rica, Czech Republic, Denmark, Ecuador, Egypt, Finland, Greece, Hong Kong, Hungary, India, Iran, Ireland, Italy, Japan, Jordan, Kenya, Luxembourg, Malaysia, Mexico, Mongolia, Morocco, Netherlands, New Zealand, Nigeria, Pakistan, Peru, Poland, Portugal, Romania, Russia, Saudi Arabia, Serbia, Singapore, South Korea, Sudan, Sweden, Switzerland, Taiwan, Thailand, Turkey, UK, Ukraine, USA, Vietnam	France, Germany, Italy, Netherlands, UK, USA
N participants completing R1; R2	Patients: 73; 54 HCP: 109; 93 Total: 182; 147	Patients: 268, 184 HCP: 684, 478 Total: 952, 662	Patients: 118; 109 HCP: 56; 49 Total: 174; 158
Types of HCP completing R1 and R2	Surgeon (coloproctologists) 36 (38%) Oncologist 26 (28%) Infectious diseases clinician 4 (4%) Pathologist 4 (4%) Radiographer 6 (6%) Radiologist 5 (5%) Radio-physicist 1 (1%) Specialist nurse 11 (13%) Missing 1 (1%)	Surgeons (esophago-gastric): 343 (71%) Cancer nurse specialists: 135 (29%)	Surgeons (urologist): 33 (68%) Oncologists: 8 (16%) Cancer nurse specialists: 8 (16%)
Delphi timeline	April 2017– September 2017 R1 open: 8 weeks Time between rounds: 4 weeks R2 open: 11 weeks	March 2019- October 2019 R1 open: 13 weeks Time between rounds: 8 weeks R2 open: 12 weeks	November 2014 – July 2015 R1 open: 6 weeks Time between rounds: 19 weeks R2 open: 6 weeks
Attrition rate R1-R2	Patients: 26.0% HCP: 14.6% Total: 19.2%	Patients: 31.1% HCP: 30.1% Total: 30.5%	Patients: 7.6% HCP: 12.5% Total: 9.2%
Total number of outcomes scored in R1	73	56	79
Number of outcomes reaching consensus in R1	12	11	9
Number of additional outcomes added to R2	5	1	5
Number of outcomes reaching consensus in R2; of which n additional outcomes	14; 1	13; 0	13; 0
N (%) of people changing score for at least one outcome	Patients: 52 (96.2%) HCP: 90 (96.7%) Total: 142 (96.5%)	Patients: 147 (79.9%) HCP: 410 (85.8%) Total: 557 (84.1%)	Patients: 102 (93.6%) HCP: 44 (89.8%) 146 (92.4%)
N (%) of people crossing a threshold for at least one outcome	Patients: 46 (85.1%) HCP: 86 (92.4%) Total: 132 (89.7%)	Patients: 129 (70.1%) HCP: 365 (76.3%) Total: 494 (74.6%)	Patients: 77 (75.5%) HCP: 18 (40.9%) Total: 95 (60.1%)
N (%) of eligible	Patients: 28 (61%)	Patients: 74 (40.2%)	Patients: 77 (75%)

(*Continued*)

Table 2. Continued

Characteristic	CORMAC	GASTROS	COMPACTERS
participants providing at least 1 reason for change	HCP: 43 (50%) Total: 71 (53%)	HCP: 117 (24.5%) Total: 191 (28.9%)	HCP: 18 (41%) Total: 95 (65%)
Median and range of number of outcomes with threshold change for those who had at least one such change	Patients: 10.5 [1–16] HCP: 18 [1–25] Total: 19 [1–26]	Patients: 8 [1–33] HCP: 6 [1-40] Total: 8 [1-40]	Patients: 9 [2-49] HCP: 9 [2–29] Overall: 9 [2–49]

descriptions of each code, including illustrative quotes from participants' responses are available in Online Appendix 1.

The coding framework developed initially from the CORMAC, and COMPACTERS data required minimal refinement when applied to the GASTROS data. Modifications included clarification of the definitions of codes and the addition of new child codes within parent codes, but no new major themes were identified.

3.1.2. Themes

The most frequently applied codes overall were "time to reflect" (482 responses, 23%), vicarious thinking (425, 21%), impact (394, 19%), and importance (311, 15%) (Table 3). Although there was some variation in the distribution of codes between the three studies, the three most common reasons for the change in the HCP group were common to all Delphi surveys (Table 4). Amongst HCP participants, "vicarious thinking" was the most frequently applied code in the CORMAC and COMPACTERS Delphi surveys and was the third most common reason for the change in the GASTROS Delphi survey.

There was greater heterogeneity between studies in the patient stakeholder group, with only one reason for the change ("personal experience") being common to the three most frequent codes in all three studies. Among patient participants, "time to reflect" was the most frequently applied code in the patient groups in the COMPACTERS and GASTROS Delphi surveys. However, in the patient group in the CORMAC Delphi survey, "time to reflect" accounted for only 5% of the reasons for the change.

3.1.3. Vicarious thinking

In all three studies, "vicarious thinking" was found more commonly in the HCP responses than the patient responses. Within the theme, four subthemes were identified, detailed in Table 5. The most common way participants described "vicarious thinking" was being influenced by the scores of other Delphi participants ("others' scores influenced me"), accounting for 68% (291) of the

Table 3. Reasons for change

	CORMAC				GASTROS			
	HCP		Patient		HCP		Patient	
Theme	Number of responses	% Of responses	Number of responses	% Of responses	Number of responses	% Of responses	Number of responses	% Of responses
Time to reflect	69	21%	9	5%	226	22%	144	31%
Vicarious thinking	160	48%	21	13%	187	18%	30	6%
Impact	40	12%	47	28%	228	23%	76	16%
Importance	29	9%	11	7%	186	18%	85	18%
Specificity/usefulness	26	8%	17	10%	129	13%	37	8%
Personal experience	0	0%	47	28%	0	0%	92	20%
Understand the survey/question differently	0	0%	1	1%	34	3%	1	0%
Error in previous round	3	1%	4	2%	10	1%	1	0%
How	3	1%	0	0%	10	1%		0%
Outcome not relevant to me		0%	6	4%	1	0%	3	1%
Covered by another item	2	1%	3	2%	0	0%		0%
Grand Total	**332**	**100%**	**166**	**100%**	**1011**	**100%**	**469**	**100%**
Unable to code	*28*	*8%*	*13*	*8%*	*72*	*7%*	*112*	*23%*

responses coded in this theme. Of these 291, 24% (103) referred to being influenced by patients' scores, 6% (27) by both patients' and HCPs' scores, and 4% (14) by HCPs' scores. HCP participants more frequently referred to patients scores' than the patient participants did.

4. Discussion

Most participants in all three Delphi surveys changed score for at least one outcome between rounds, and the number of items reaching consensus was subsequently increased. This finding supports conducting a Delphi survey over the use of a single questionnaire for core outcome set projects where building consensus is the objective.

The coding framework derived from two studies showed good applicability across a third, requiring only minimal refinements and no new major themes identified. We encourage other researchers to use and further refine our coding framework as necessary to further generate the much needed data in this field.

The reasons that participants gave for changing their scores provide useful data for researchers to consider when designing Delphi surveys. It is recognized that the duration of rounds and time between rounds are important factors, with a longer gap allowing for more change in an individual's circumstances, knowledge, and situational context [16]; however, at present, there is very little data available on this aspect of Delphi survey design. A 2019 study of the impact of design characteristics on response rates in COS Delphi surveys [17] found that insufficient data were reported on the duration of and time between rounds to allow analysis. Time to reflect on the importance of outcomes between rounds was the most common reason for a change in the current study. Furthermore, it was most frequently identified in the COMPACTERS Delphi survey, which had a 19-week interval between R1 and R2 compared to the 8-week and 4-week intervals in the GASTROS and COMRAC Delphi survey respectively. This finding is not sufficient to draw any conclusions about a direct correlation between the time between rounds and the degree of consensus. However, it does raise a question of whether the timing between rounds in a Delphi survey might influence participants thinking and their responses and potentially impact the results. We recommend that the duration of and time between rounds in Delphi surveys is reported to facilitate further work to explore the potential impact of this characteristic on participation and consensus.

Vicarious thinking, or trying to understand the importance of an outcome from the perspective of another participant, emerged as a major theme and was the second most common reason for change overall. Vicarious thinking appears to be facilitated by seeing other participants' scores, with nearly 70% of participants in the vicarious thinking group making direct reference to the scores of other participants as a reason for their change in score. This finding is consistent with the findings of Turnbull et al. [10], who reported that 83% of respondents to a post-Delphi survey questionnaire reported considering the scores of other participants when scoring in subsequent rounds. Furthermore, we found that of participants being influenced by others' scores, just over 40% cited being influenced by one particular stakeholder group, consistent with the differential weighting to scores from different stakeholder groups reported by Turnbull. This is also consistent with the findings from a nested randomized study by Brookes et al. [18], who

COMPACTERS					
HCP		Patient			
Number of responses	% Of responses	Number of responses	% Of responses	Total number of responses	Total % of responses
4	22%	30	42%	482	23%
12	67%	15	21%	424	21%
1	6%	2	3%	394	19%
0	0%	0	0%	311	15%
1	6%	1	1%	211	10%
0	0%	16	22%	156	8%
0	0%	7	10%	43	2%
0	0%	0	0%	18	1%
0	0%	0	0%	13	1%
0	0%	1	1%	11	1%
0	0%	0	0%	5	0%
18	**100%**	**72**	**100%**	**2068**	**100%**
0	*0%*	*5*	*7%*	*230*	*11%*

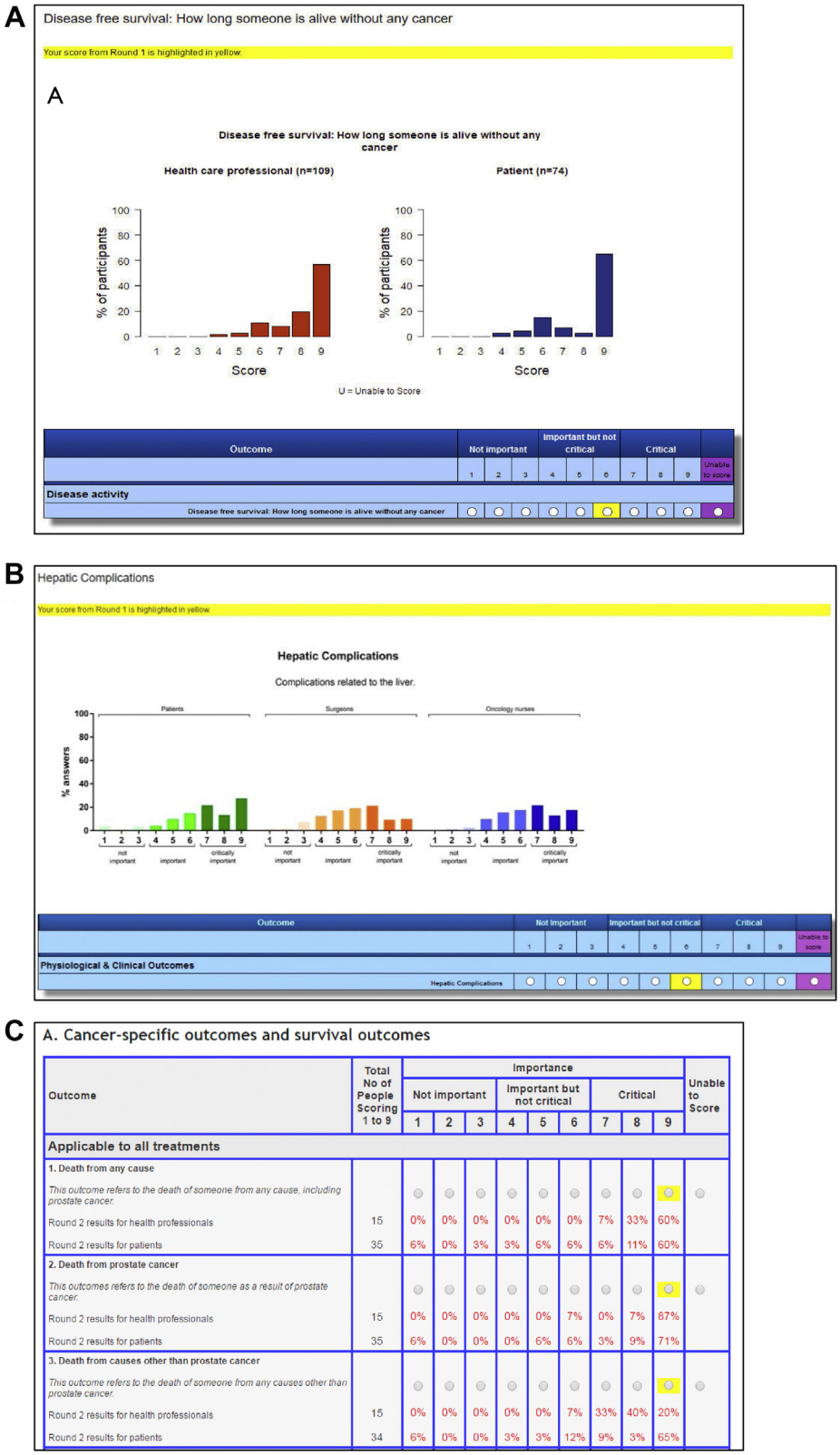

Fig. 2. Screenshots from round 2 of the CORMAC (A), GASTROS (B), and COMPACTERS (C) Delphi surveys showing how participants were shown the summarized results from round 1.

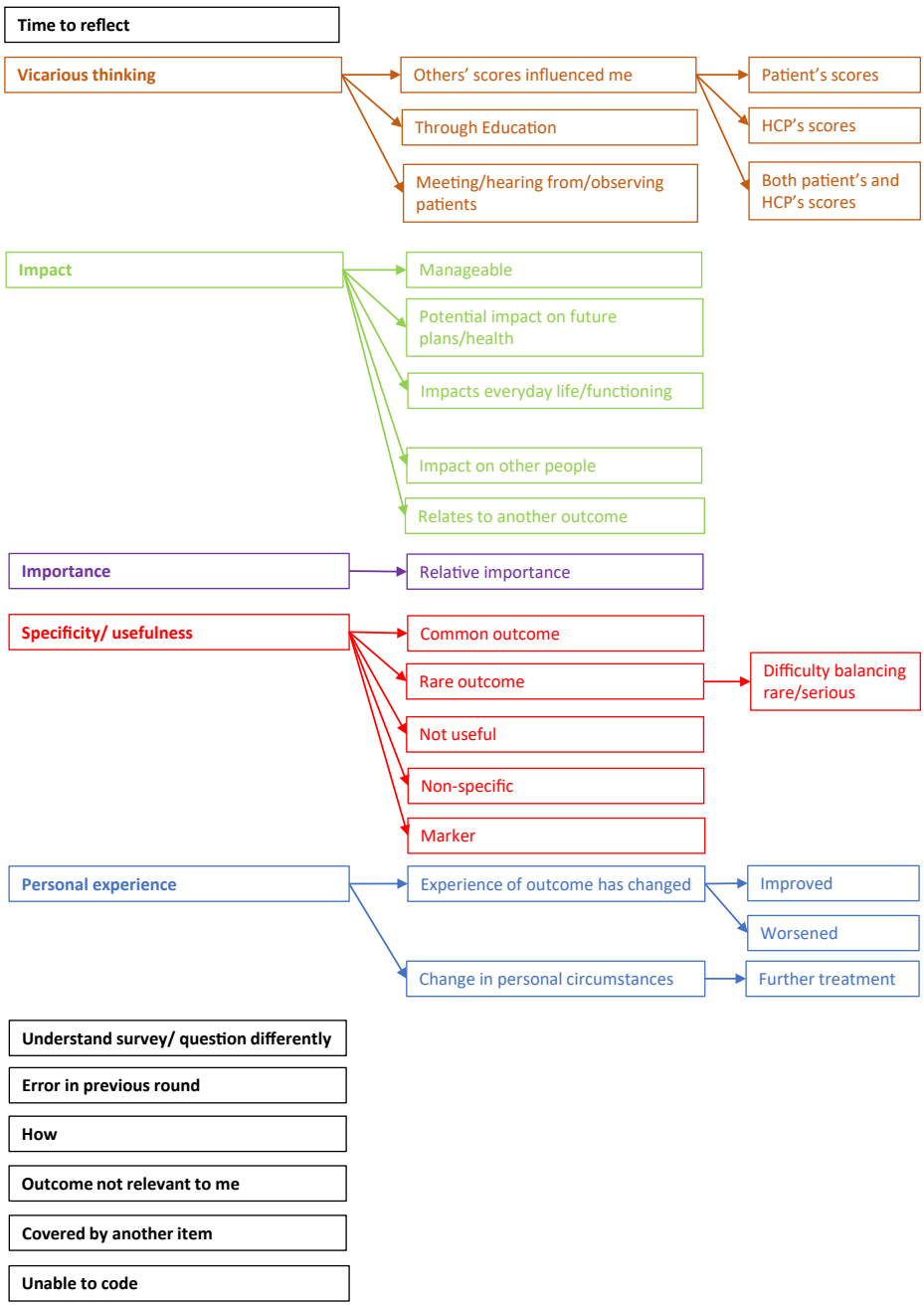

Fig. 3. Coding framework showing the relationship of major themes (''parent codes'') and their related subthemes (''child codes'').

reported that compared to participants randomized to receive pooled feedback from all participants, there was a small increase in the number of outcomes reaching consensus in the group randomized to receive separate feedback from each stakeholder group. These findings suggest that how results are presented to participants between rounds is a key factor in Delphi design that could influence the result. This is an important area for future methodological research.

Vicarious thinking and being influenced by the scores of others was more common in the HCP group than the patient group across all three Delphi surveys. This corroborates the post-Delphi survey findings of Turnbull et al. who also found patients considered the results from other stakeholder groups less frequently than all other stakeholder groups. Where others' scores were considered, however, they found that HCPs prioritized the results from both patients and other HCPs in contrast to our results, which show HCPs

Table 4. Most frequent reasons given

	CORMAC			
	HCP		Patient	
Rank	Code	% (n) of responses	Code	% (n) of responses
1	Vicarious thinking	48 (160)	Personal experience	28 (47)
2	Time to reflect	21 (69)	Impact	28 (47)
3	Impact	12 (40)	Vicarious thinking	13 (21)

more commonly referenced being influenced by patients' scores than by the scores from other HCPs. The difference in vicarious thinking that we observed between patient and HCP participants may help to explain the findings of Maclennan et al. in the nested randomized study of feedback composition on consensus. They observed no evidence of difference between groups receiving peer-only, multiple-separate or combined feedback. This may be explained by a high level of agreement in the first round of the Delphi; however, it may also be explained if the majority of patients (accounting for two-thirds of the participants in that study) do not use the results of others in their decision making.

We have identified only one other study examining the comments submitted by participants in a Delphi survey for COS development. Sautenet et al. [19] identified five broad themes from a thematic analysis of free-text comments submitted by participants in a renal transplant COS. The themes they describe align conceptually with the themes identified in this study. For example, their theme ''Understanding and awareness of risks'' describes patients and caregivers increasing their scores in response to comments from health professionals, as well as health professionals gaining increased respect for the impact of outcomes on patients, which maps to our theme of vicarious thinking. Their description of ''personal relevance'' maps to our theme of ''personal experience,'' ''capacity to control'' and ''debilitating repercussion'' to ''Impact'' and ''importance.'' The three Delphi surveys included in our study were for cancer COS so it is encouraging to see broadly similar themes in a noncancer study. However, it is important to acknowledge that different disease contexts may generate diverse and unique reasons for score change.

This study has some limitations. To allow comparison, a significant proportion of data from the COMPACTERS study had to be excluded; the randomization to various forms of feedback in that study meant that the sample of participants included in this analysis should, however, be representative of those in the wider study. In addition, participants in the COMPACTERS Delphi survey were given a single opportunity to provide reasons for change at the end of the survey, whereas in the CORMAC and GASTROS Delphi surveys, participants were asked to give a reason each time a score threshold was crossed. These differences resulted in there being only a low number of responses from the COMPACTERS Delphi survey so caution must

Table 5. Vicarious thinking as reason for change in score

	CORMAC				GASTROS			
	HCP		Patient		HCP		Patient	
Subtheme	Number of responses	% Of responses	Number of responses	% Of responses	Number of responses	% Of responses	Number of responses	% Of responses
Others' scores influenced me	128	80%	3	14%	109	58%	30	100%
Not further specified	*50*	*31%*	*3*	*14%*	*54*	*29%*	*25*	*83%*
Patient's scores	*46*	*29%*	*0*	*0%*	*49*	*26%*	*3*	*10%*
Both patient's and HCP's scores	*26*	*16%*	*0*	*0%*	*1*	*1%*	*0*	*0%*
HCP's scores	*6*	*4%*	*0*	*0%*	*5*	*3%*	*2*	*7%*
Meeting/hearing from/observing patients	1	1%	8	38%	69	37%	0	0%
Vicarious thinking- not further specified	31	19%	10	48%	6	3%	0	0%
Through education	0	0%	0	0%	3	2%	0	0%
Grand Total	**160**	**100%**	**21**	**100%**	**187**	**100%**	**30**	**100%**

GASTROS				COMPACTERS			
HCP		Patient		HCP		Patient	
Code	% (n) of responses	Code	% (n) of responses	Code	% (n) of responses	Code	% (n) of responses
Impact	23 (228)	Time to reflect	31 (144)	Vicarious thinking	67 (12)	Time to reflect	42 (30)
Time to reflect	22 (226)	Personal experience	20 (92)	Time to reflect	22 (4)	Personal experience	22 (16)
Vicarious thinking	18 (187)	Importance	18 (85)	Impact; specificity/usefulness	6 (1)	Vicarious thinking	21 (15)

be employed when drawing any conclusions from comparison of results across the three studies.

Qualitative analysis of the free text is inherently subjective, and there is likely to be an element of interrater and intrarater variability; we attempted to minimize variability through coding by multiple researchers and frequent coding discussion meetings. It is also important to acknowledge that human decision-making behavior is complex and participants written responses to a single direct question about their reasons for changing score can provide only a one-dimensional view of what is a multidimensional process. The responses can only reflect why the participant believes they changed their score when it is likely that unconscious factors also play a role. Responses also only represent what participants deem as an acceptable responses or be influenced by what they believe researchers want to hear.

Vicarious thinking was identified relatively infrequently in the GASTROS Delphi survey in both patients (6%) and HCPs (18%) compared to the CORMAC and COMPACTERS studies. The reasons for this difference are not clear. It is notable that the GASTROS Delphi survey was translated into multiple languages with reverse translation of participants' reasons for change back into English. It is possible that nuances of language or cultural differences were not appreciated during coding, which may have affected the results. We did not find any relationship between the types of HCP participating in the Delphi surveys and the frequency of vicarious thinking. Further the responses from specific types of HCP also varied widely between the Delphi surveys; for example 64% of responses from surgeons in the CORMAC study cited vicarious thinking compared to 15% of responses from surgeons in the GASTROS study. It is interesting to note that the magnitude of the difference between patients and HCPs citing vicarious thinking is similar in all three studies (about three times greater from HCPs than patients). This suggests that there may be more in common between different stakeholder groups in the same study population than between the same stakeholder groups in different populations. The degree to which cultural factors might influence attitudes toward consensus methods and health outcomes is an area that should be explored in future work.

The validity and generalizability of the coding framework derived through this study should be further assessed and refined in a broader context including a variety of

COMPACTERS					
HCP		Patient			
Number of responses	% Of responses	Number of responses	% Of responses	Total number of responses	Total % of responses
10	83%	11	73%	291	68%
7	*58%*	*6*	*40%*	*145*	*34%*
2	*17%*	*3*	*20%*	*103*	*24%*
0	*0%*	*0*	*0%*	*27*	*6%*
1	*8%*	*2*	*13%*	*16*	*4%*
0	0%	1	7%	79	19%
1	8%	3	20%	51	12%
1	8%	0	0%	4	1%
12	**100%**	**15**	**100%**	**425**	**100%**

health conditions and populations. The framework could also be tested for content validity by Delphi participants through the creation of a questionnaire where respondents are invited to select the codes from the framework that best describe their reasons for changing score along with an option to provide a different option of their reason was not represented.

4.1. Summary and conclusion

Our findings suggest that within a Delphi, showing participants summarized results and completing a second round are directly beneficial to reaching consensus. Time to reflect on the importance of outcomes between rounds and the opportunity to try to understand the experience of an outcome from the perspective of another are helpful to this process.

5. Funding

This work was supported by a National Institute for Health Research (NIHR) Senior Investigator Award (NF-SI_0513-10,025). The views expressed in this article are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

6. GASTROS

This study is supported by the National Institute for Health Research (NIHR) Doctoral Research Fellowship Grant (DRF-2015-08-023). The views expressed in this article are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

7. COMPACTERS

This work was funded by the Cancer Research Aberdeen and North East (CRANES) charity. Registered charity number: SC034542.

8. CORMAC

This work was supported by the National Institute for Health Research (NIHR) under its Research for Patient Benefit (RfPB) Programme (Grant Reference Number PB-PG-1013-32,064). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. The work was carried out as part of a PhD, funded in part by MRC North West Hub for Trials Methodology Research (Grant reference number MR/K025635/1). We acknowledge funding from the Christie charitable PMP research fund, and the kind donations to this fund from patients, families, and friends.

The funding sources had no role in the design of this study and will not have any role during its execution, analyses, interpretation of the data, or decision to submit results.

9. Conflicts of interest

RF, SM, and BA have no conflicts of interest.

PW is a member of the COMET Initiative Management Group, and PW developed the DelphiManager software described in this article.

CRediT authorship contribution statement

Rebecca Fish: Funding acquisition, Conceptualization, Methodology, Investigation, Formal analysis, Writing - original draft. **Steven MacLennan:** Funding acquisition, Conceptualization, Methodology, Investigation, Formal analysis, Writing - review & editing. **Bilal Alkhaffaf:** Funding acquisition, Investigation, Formal analysis, Writing - review & editing. **Paula R. Williamson:** Funding acquisition, Conceptualization, Methodology, Investigation, Formal analysis, Writing - review & editing, Supervision.

Acknowledgments

We would like to thank and acknowledge the participants in the Delphi surveys used here. We also thank individuals who contributed significantly to the wider COS development projects: Professor Andrew G. Renehan and Professor Caroline Sanders (CORMAC); Dr. Thomas B. L. Lam (COMPACTERS); Professor Iain A Bruce and Professor Jane M Blazeby (GASTROS).

Appendix A

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2020.09.028.

References

[1] Williamson P. Core outcome sets will improve the quality of obstetrics research. BJOG: An Int J Obstet Gynaecol 2014;121(10):1196.
[2] Williamson P, Altman D, Blazeby J, Clarke M, Gargon E. Driving up the quality and relevance of research through the use of agreed core outcomes. J Health Serv Res Policy 2012;17:1−2.
[3] Williamson P, Clarke M. The COMET (core outcome measures in effectiveness trials) Initiative: its role in improving cochrane reviews [editorial]. Cochrane Database Syst Rev 2012;4:1465−858.
[4] Heneghan C, Goldacre B, Mahtani KR. Why clinical trial outcomes fail to translate into benefits for patients. Trials 2017;18:122.
[5] COMET. COS. Uptake and endorsement 2020 [cited 2020]. http://www.comet-initiative.org/COSEndorsement. Accessed August 1, 2020.

[6] Williamson PR, Altman DG, Bagley H, Barnes KL, Blazeby JM, Brookes ST, et al. The COMET Handbook: version 1.0. Trials 2017;18(3):280.

[7] Kirkham JJ, Davis K, Altman DG, Blazeby JM, Clarke M, Tunis S, et al. Core outcome set-STAndards for development: the COS-STAD recommendations. PLoS Med 2017;14(11):e1002447.

[8] Williamson PR, Blazeby JM, Brookes ST, Clarke M, Terwee CB, Young B. Comments on Chevance et al.; 'Improving the generalizability and credibility of core outcome sets (COS) by a large and international participation of diverse stakeholders'. J Clin Epidemiol 2020;125:229−31.

[9] Gargon E, Gorst SL, Williamson PR. Choosing important health outcomes for comparative effectiveness research: 5th annual update to a systematic review of core outcome sets for research. PLoS One 2019; 14:e0225980.

[10] Turnbull AE, Dinglas VD, Friedman LA, Chessare CM, Sepulveda KA, Bingham CO 3rd, et al. A survey of Delphi panelists after core outcome set development revealed positive feedback and methods to facilitate panel member participation. J Clin Epidemiol 2018;102:99−106.

[11] Fish R, Sanders C, Adams R, Brewer J, Brookes ST, DeNardo J, et al. A core outcome set for clinical trials of chemoradiotherapy interventions for anal cancer (CORMAC): a patient and health-care professional consensus. Lancet Gastroenterol Hepatol 2018;3(12):865−73.

[12] Alkhaffaf B, Glenny A-M, Blazeby JM, Williamson P, Bruce IA. Standardising the reporting of outcomes in gastric cancer surgery trials: protocol for the development of a core outcome set and accompanying outcome measurement instrument set (the GASTROS study). Trials 2017;18(1):370.

[13] MacLennan S, Williamson PR, Bekema H, Campbell M, Ramsay C, N'Dow J, et al. A core outcome set for localised prostate cancer effectiveness trials. BJU Int 2017;120:E64−79.

[14] COMET DelphiManager. COMET initiative. 2018. [cited 2018]. Available at: http://www.comet-initiative.org/delphimanager/index.html. Accessed August 1, 2020.

[15] Ritchie J, Lewis JIn: Qualitative research practice: a guide for social science students and researchers, 2003. London; Thousand Oaks, Calif: Sage Publications; 2003:336.

[16] Day J, Bobeva M. A generic toolkit for the successful management of delphi studies. Electron J Business Res Methods 2005;3(2): 103−16.

[17] Gargon E, Crew R, Burnside G, Williamson PR. Higher number of items associated with significantly lower response rates in COS Delphi surveys. J Clin Epidemiol 2019;108:110−20.

[18] Brookes ST, Macefield RC, Williamson PR, McNair AG, Potter S, Blencowe NS, et al. Three nested randomized controlled trials of peer-only or multiple stakeholder group feedback within Delphi surveys during core outcome and information set development. Trials 2016;17(1):409.

[19] Sautenet B, Tong A, Manera KE, Chapman JR, Warrens AN, Rosenbloom D, et al. Developing consensus-based priority outcome domains for trials in kidney transplantation: a multinational delphi survey with patients, caregivers, and health professionals. Transplantation 2017;101(8):1875−86.