Contents lists available at ScienceDirect

European Journal of Surgical Oncology

journal homepage: www.ejso.com

The Delphi and GRADE methodology used in the PSOGI 2018 consensus statement on Pseudomyxoma Peritonei and Peritoneal Mesothelioma

R.J. Lurvink^{a,1}, L. Villeneuve^{b,c,1}, K. Govaerts^d, I.H.J.T. de Hingh^{e,f}, B.J. Moran^g, M. Deraco^h, K. Van der Speeten^d, O. Glehenⁱ, V. Kepenekian^{i,c}, S. Kusamura^{j,*}, List of PSOGI collaborators

^a Department of Surgery, Catharina Hospital, Eindhoven, the Netherlands

^b Department of Public Health, Clinical and Epidemiological Research, Hospices Civils de Lyon, Lyon, France

^c EMR 3738, Lyon 1 University, Lyon, France

^d Department of Surgical Oncology, Hospital Oost-Limburg, Genk, Belgium

^e Department of Surgical Oncology, Catharina Cancer Institute, Eindhoven, the Netherlands

^f Department of Epidemiology, GROW School for Oncology and Developmental Biology, Maastricht University, Maastricht, the Netherlands

^g Peritoneal Malignancy Institute, North-Hampshire Hospital, Basingstoke, United Kingdom

^h Department of Surgery, Peritoneal Surface Malignancy Unit, Fondazione IRCCS Istituto Nazionale Dei Tumori di Milano, Milan, Italy

ⁱ Department of Digestive Surgery, Centre Hospitalier Lyon Sud, Lyon, France

^j Department of Surgery, Peritoneal Surface Malignancy Unit, Fondazione IRCCS Istituto Nazionale Dei Tumori di Milano, Via Giacomo Venezian 1, Milano, Milan Cap, 20133, Italy

ARTICLE INFO

Article history:

Received 7 March 2019

Accepted 8 March 2019

Available online 23 March 2019

Keywords:

Peritoneal mesothelioma

Pseudomyxoma peritonei

Peritoneal surface malignancies

Delphi consensus

GRADE system

ABSTRACT

Pseudomyxoma Peritonei (PMP) and Peritoneal Mesothelioma (PM) are both rare peritoneal malignancies. Currently, affected patients may be treated with Cytoreductive Surgery and Hyperthermic Intraperitoneal Chemotherapy offering long-term survival or even cure in selected patients. However, many issues regarding the optimal treatment strategy are currently under debate.

To aid physicians involved in the treatment of these patients in clinical decision making, the PSOGI executive committee proposed to create a consensus statement on PMP and PM. This manuscript describes the methodology of the consensus process.

The Delphi technique is a reliable method for attaining consensus on a topic that lacks scientific evidence through multiple voting rounds which feeds back responses to the participants in between rounds. The GRADE system provides a structured framework for presenting and grading the available evidence. Separate questionnaires were created for PMP and PM and sent during two voting rounds to 80 and 38 experts, respectively. A consensus threshold of 51.0% was chosen.

After the second round, consensus was reached on 92.9%–100.0% of the questions. The results were presented and discussed in the plenary session at the PSOGI 2018 international meeting in Paris. A third round for the remaining issues is currently in progress.

In conclusion, using the Delphi technique and GRADE methodology, consensus was reached in many issues regarding the treatment of PM and PMP amongst an international panel of experts. The main results will be published in the near future.

© 2019 Published by Elsevier Ltd.

Abbreviations: CRS, Cytoreductive Surgery; DMPM, Diffuse Malignant Peritoneal Mesothelioma; GRADE, Grades of Recommendation, Assessment, Development, and Evaluation; EPIC, Early Post-operative Intraperitoneal Chemotherapy; HIPEC, Hyperthermic Intraperitoneal Chemotherapy; LoE, Level of Evidence; MCPM, MultiCystic Peritoneal Mesothelioma; PICO, Patient, Intervention, Control, Outcome; PM, Peritoneal Mesothelioma; PMP, Pseudomyxoma Peritonei; PSOGI, Peritoneal Surface Oncology Group International; RCT, Randomized Controlled Trial; SC, Steering Committee; SoR, Strength of Recommendation(s); WDPPM, Well Differentiated Papillary Peritoneal Mesothelioma.

* Corresponding author.

E-mail address: shigeki.kusamura@istitutotumori.mi.it (S. Kusamura).

¹ These authors contributed equally to this work.

<https://doi.org/10.1016/j.ejso.2019.03.012>

0748-7983/© 2019 Published by Elsevier Ltd.

Introduction

Pseudomyxoma peritonei (PMP) and Peritoneal Mesothelioma (PM) are both rare peritoneal malignancies, affecting about one to three in a million people each year [1,2]. Due to aspecific symptoms and diagnostic limitations, it is usually discovered at an advanced stage, limiting treatment options and decreasing prognosis.

Recently, Cytoreductive Surgery (CRS) followed by Hyperthermic Intraperitoneal Chemotherapy (HIPEC) has emerged as a potentially curative treatment of PMP and PM. The advent of this new treatment modality has significantly improved the prognosis of patients suffering from these diseases. However, many issues regarding the management of these diseases remain under debate.

Due to the very low incidence of PMP and PM, randomized controlled trials (RCT) addressing the various aspects of its management are currently lacking and highly unlikely to be performed in the near future. Instead, it is desirable to harmonize current clinical practice, bringing together expert opinions and experiences, as well as the available literature to build a consensus based guideline, especially in the most contentious issues. This may facilitate clinical decision making by physicians being confronted with this rare disease. The initiative for this guideline was proposed by members of the PSOGI Executive Committee and will include an update of the PMP consensus statement provided by the Peritoneal Surface Oncology Group International (PSOGI) in 2008 [3].

Purpose of the current manuscript is to describe the methodology used for this consensus statement and to discuss the main results.

Materials and methods

In this consensus statement for PMP and PM, we used two techniques to describe the level of evidence (LoE), to create questionnaires, and to achieve consensus. Here, we will describe the goals, approach, pros and cons, and reasons for choosing these methodologies.

The Delphi technique

The term ‘Delphi’ originates from the Oracle of Delphi in Greek mythology, which was frequently consulted in wars and before making important decisions, for she was believed to predict the future. The Delphi technique seeks to approximate the most valid answer by assuming superiority of the judgment of a group over the judgment of a single individual [4]. In contrast to a conventional group meeting, the Delphi technique guarantees anonymity, does not need to have all experts meet in person to give their opinion, and provides feedback on the responses to create discussion. Also, each participant can state their opinion independently of other participants, which is usually not the case in a conventional meeting where dominant participants will easily overrule less-dominant participants [5]. The Delphi technique is increasingly used over the past decades in various fields of medicine. Given these advantages, the Delphi technique was chosen over conventional questionnaires in the current project.

The Delphi technique is a reliable method for obtaining consensus on a specific topic that lacks scientific evidence and therefore relies on expert opinion. It consists of multiple voting rounds in which a multiple-choice questionnaire is sent to a group of participants who are considered experts on the topic [6,7]. In general, the more participants the better, as well as that it would be preferable to pick experts from different backgrounds to create a heterogeneous group with a broader knowledge level. This would obviously contribute to the quality of the consensus [6]. After each round, the results are summarized and returned to the participants

in a subsequent round. These results are returned to create a debate, through which participants may reconsider their response in the next round. There is no limitation to the number of rounds with the Delphi technique [6,7]. Ideally, this will, for each question, lead to reaching a pre-set threshold percentage of votes that is considered consensus.

The GRADE system

The GRADE system (Grades of Recommendation, Assessment, Development, and Evaluation) is used for setting up questions, rating the quality of scientific evidence and stating either strong or weak recommendations [5,8,9]. An advantage of the GRADE system is the separate rating of the quality of evidence and the strength of the recommendation (SoR), e.g. allowing in some cases a strong recommendation despite low LoE [8]. The process is facilitated with the structured framework that the GRADE system provides. The superiority of the GRADE system as compared to other systems is based on the fact that the GRADE system uses clear cut and transparent rules for rating the evidence and stating recommendations [9]. These explicitly defined requirements are lacking in other systems, reducing their scientific value.

According to the GRADE system, questions should be set up according to the PICO-model: Patient, Intervention, Control, Outcome [10]. This model is recommended by The Cochrane Handbook for Systematic Reviews of Interventions to develop a review question, thus ensuring that the relevant components of the question are well defined [11]. The PICO framework is primarily centred on therapy questions, and although it can be adapted to formulate research questions related to prognosis or diagnosis, it is less suitable for other types of clinical information needs [12].

The LoE concerning each question was categorized into four degrees: ‘high’, ‘moderate’, ‘low’, and ‘very low’ LoE. These levels agree with a certain degree of confidence in the estimate of the effect [13]. In general, RCT’s are classified as ‘high’ LoE, whereas observational studies are considered ‘low’ or ‘very low’ LoE. However, a high risk of bias can decrease the LoE, while a low risk of bias can increase the LoE [8,14]. After the quality of evidence has been determined, balance between benefits and harms, as well as the patient’s preferences, will be considered as well before providing a recommendation [8]. Recommendations were either ‘strong’ or ‘weak’, and either ‘positive’ or ‘negative’, leading to four possible options [15,16]. So, only after all these factors have been considered, the expert panel was able to propose a certain recommendation [17]. An overview and description of possible combinations and their implications for recommendations are shown in Table 1.

Expert panel and questionnaire development

For the composition of the expert panel, the steering committee (SC) started identifying surgeons and medical oncologists worldwide with expertise in the management of PMP and/or PM, based on centres’ peritoneal surface malignancies volume and scientific production over the last two decades.

Concerning PM, more restrictive criteria were applied following the French RENAPE network standards. The centre was defined as expert when they treat at least 15 (rare) peritoneal surface malignancies per year and have experienced more than 100 CRS and HIPEC procedures [18].

The SC also decided on which topics were to be included in the questionnaires. The timeline of the meetings and voting rounds is shown in Fig. 1.

Main topics to be included in the PMP questionnaires:

- Pathology terminology for histological classification

Table 1
GRADE Strength of Recommendations and Quality of the Evidence (the combinations listed could be further subdivided in positive and negative categories according to the direction of the balance between desirable and undesirable effects).

Strength of Recommendation and Quality of Evidence	Clarity of balance between desirable and undesirable effects	Methodological Quality of Supporting Evidence (examples)	Implications
Strong recommendation, high-quality evidence.	Desirable effects clearly outweigh undesirable effects, or vice versa.	Consistent evidence from well-performed RCTs or exceptionally strong evidence from unbiased observational studies.	Recommendation can apply to most patients in most circumstances. Further research is unlikely to change our confidence in the estimate of the effect.
Strong recommendation, moderate quality evidence.	Desirable effects clearly outweigh undesirable effects, or vice versa.	Evidence from RCTs with important limitations (inconsistent results, methodological flaws, indirect, or imprecise) or exceptionally strong evidence from unbiased observational studies.	Recommendation can apply to most patients in most circumstances. Further research (if performed) is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Strong recommendation, low-quality evidence.	Desirable effects clearly outweigh undesirable effects, or vice versa.	Evidence for at least one critical outcome from observational studies, RCTs with serious flaws or indirect evidence.	Recommendation may change when higher quality evidence becomes available. Further research (if performed) is likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Strong recommendation, very low-quality evidence (very rarely applicable).	Desirable effects clearly outweigh undesirable effects, or vice versa.	Evidence for at least one critical outcome from unsystematic clinical observations or very indirect evidence.	Recommendation may change when higher quality evidence becomes available; any estimate of effect for at least one critical outcome is very uncertain.
Weak recommendation, high-quality evidence.	Desirable effects closely balanced with undesirable effects.	Consistent evidence from well-performed RCTs or exceptionally strong evidence from unbiased observational studies.	The best action may differ depending on circumstances or patients or societal values. Further research is unlikely to change our confidence in the estimate of the effect.
Weak recommendation, moderate-quality of evidence.	Desirable effects closely balanced with undesirable effects.	Evidence from RCTs with important limitations (inconsistent results, methodological flaws, indirect, or imprecise) or exceptionally strong evidence from unbiased observational studies.	Alternative approaches likely to be better for some patients under some circumstances. Further research (if performed) is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Weak recommendation, low-quality of evidence.	Uncertainty in the estimates of desirable effects, harms, and burden; desirable effects, harms and burden may be closely balanced.	Evidence for at least one critical outcome from observational studies, RCTs with serious flaws or indirect evidence.	Other alternatives may be equally reasonable. Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Weak recommendation, very low-quality evidence.	Major uncertainty in the estimates of desirable effects, harms, and burden; desirable effects may or may not be balanced with undesirable effects.	Evidence for at least one critical outcome from unsystematic clinical observations or very indirect evidence.	Other alternatives may be equally reasonable. Any estimate of effect, for at least one critical outcome, is very uncertain.

- Pre-operative work-up
- Use of systemic chemotherapy
- Contra-indications for CRS and HIPEC
- Treatment of an (unexpected) appendiceal neoplasm at appendectomy
- Treatment of premenopausal women
- Follow-up after CRS and HIPEC

Main topics to be included in the PM questionnaires:

- Diffuse Malignant Peritoneal Mesothelioma (DMPM):
 - o Elements required in the pathology report
 - o Pre-operative work-up
 - o Contra-indications for CRS and HIPEC
 - o Choice of HIPEC drug
 - o Treatment of premenopausal women
 - o Use of systemic chemotherapy
 - o Follow-up after CRS and HIPEC
- Well Differentiated Papillary Peritoneal Mesothelioma (WDPPM):
 - o Indications for CRS with/without HIPEC
 - o Treatment of premenopausal women
 - o Use of systemic chemotherapy
- MultiCystic Peritoneal Mesothelioma (MCPM)
 - o Indications for CRS with/without HIPEC
 - o Treatment of premenopausal women

o Use of systemic chemotherapy

Literature concerning these topics was searched in Pubmed and Cochrane Library and summarized. We used the GRADE system to build questionnaires for both PMP and PM.

Each question required two answers: one for rating the quality of evidence (high, moderate, low, very low), and another for rating the SoR (strong positive, weak positive, weak negative, strong negative). To facilitate participants to give an evidence-based answer, an overview of the literature as well as a summary regarding the Delphi and GRADE methodology have been provided.

Although the first round of a Delphi generally consists of an 'open response' questionnaire, it was decided that all possible and relevant clinical scenarios were included in multiple choice questions. The questionnaires were available with an online questionnaire tool (SurveyMonkey®), of which the weblinks were sent to the participants by email (Appendix 1). Reminder e-mails were sent to stimulate responses. After the first round, the results were summarized for each question and shown to the participants in the second voting round.

The results of the second round were shown and discussed at the PSOGI 2018 meeting in Paris. Here, a third round was proposed to reach consensus on the few remaining questions, and to include two rephrased questions from the PMP questionnaire due to lack of clarity in the way they were previously formulated.

DMPM: Diffuse Malignant Peritoneal Mesothelioma; MCPM: MultiCystic Peritoneal Mesothelioma; PMP: Pseudomyxoma Peritonei; WDPPM: Well-Differentiated Papillary Peritoneal Mesothelioma

Fig. 1. Flowchart and timeline of meetings and voting rounds.

Results

Expert panel

The PM expert panel consisted of 37 surgeons and 1 medical

oncologist, originating from 17 countries. The PMP expert panel consisted of 69 surgeons and 11 medical oncologists, originating from 28 countries (Table 2).

Table 2
Number of experts and origin in each expert panel.

	PM	PMP
Specialty		
Surgical Oncology	37	69
Medical Oncology	1	11
Country		
Australia	1	5
Belgium	2	3
Brazil	0	2
Canada	1	2
China	2	2
Denmark	1	1
France	4	6
Germany	3	6
Greece	1	2
India	0	5
Israel	2	2
Italy	4	7
Japan	1	1
Netherlands	1	5
New Zealand	0	1
Norway	0	2
Poland	0	1
Saudi Arabia	0	1
Singapore	0	1
South Korea	0	1
Spain	1	4
Sweden	1	2
Switzerland	1	2
Taiwan	0	1
Turkey	0	1
UK	5	5
Ukraine	0	1
USA	7	8

Questionnaires and voting rounds

In total, 42 questions regarding PM and 69 questions regarding PMP were formulated. PM questions were distributed over 3 questionnaires, regarding DMPM (27 questions), WDPPM (8 questions), and MCPM (7 questions). The PMP questions were divided over 2 questionnaires; part I (34 questions) considered terminology and pathological classification, preoperative evaluation and diagnosis, CRS and HIPEC, HIPEC regimen, management of diaphragmatic disease and thoracic extension, Early Postoperative Intraperitoneal Chemotherapy (EPIC), maximal tumour debulking, and follow-up. Part II (35 questions) mainly focused on the use of systemic chemotherapy and the management of the unexpected finding of various types of appendiceal neoplasms at appendectomy.

The first voting round of both PM and PMP took place in July 2018 over a period of 19 consecutive days. According to the Delphi methodology, these results were shown in the questionnaires during the second voting round, which was held in August 2018 and lasted 22 days.

Completion rates ranged from 67% to 82% during the first round and from 61 to 69% during the second round (Fig. 1). Questionnaires that had not been fully completed, were not taken into account in the analyses.

Achieving consensus

Like in the 2006 PSOGI consensus process [3], we defined that consensus was reached when the cut-off point of 51.0% of the votes at a question was obtained.

During the first round, the participants graded the available LoE on a 4-point scale from 'very low' to 'high'. The interpreted LoE was very variable, and an agreement was only achieved in a small percentage of the questions (0.0%–17.6%). This level of agreement

changed drastically during the second round, in which the participants agreed on the LoE in 71.4%–96.9% of the questions.

A similar, but smaller, shift between the first and second round was seen in the questions regarding recommendations. During the first round, consensus regarding the recommendations was reached for 60.0%–71.4% of the questions. This number had risen to 92.9%–100.0% during the second round. The amount of questions on which consensus was reached per questionnaire and per round are shown in Table 3.

Finally, we also analysed the voting behaviour of high-volume PMP centres as compared to all PMP centres, since we hypothesised that a high-volume centre could judge the LoE and SoR differently than the other centres. These results are shown in Table 4 and show no significant differences in voting behaviour between those two groups.

Discussion

This paper will be the first of a series concerning the 2018 PSOGI consensus statement for PM and PMP. With this consensus statement, we seek to support other clinicians and (starting) HIPEC centres worldwide in the treatment of these rare diseases. This work provides a broad outline of the first Delphi consensus statement in peritoneal surface malignancies on such a large scale.

We used the Delphi technique and GRADE system to create a structured and well-founded consensus statement. Although these methods have numerous advantages, some limitations should be taken into consideration and might affect validity, which we will discuss below.

Advantages of using the Delphi technique and GRADE method

The Delphi technique provides anonymity, both when answering the questionnaires as when receiving back the answers, allowing all participants to express their opinion independently. Giving feedback to responders intends to broaden knowledge and stimulate new ideas, so participants might reconsider their answer in the next voting round(s). This will ideally lead to achieving consensus.

The GRADE system provides a structured framework for the presentation and rating of scientific evidence that can be used in systematic reviews and guidelines. An important advantage of the GRADE system over other rating systems is that quality of evidence and the SoR could eventually be rated separately, which means that a strong recommendation can be given even though high quality scientific evidence is lacking [8]. We considered this very helpful for our consensus statement, since the quality of the evidence was expected to be low; using the GRADE system, this low-quality evidence does not rule out a strong recommendation.

Limitations

With the Delphi technique, the goal is to reach a pre-defined consensus threshold. However, it is unclear how to define this threshold: scientific evidence on this matter is lacking, although it is generally believed that increasing the threshold increases validity [6,7].

We chose a 51.0% threshold, based on the previous PSOGI consensus experience in 2006 [3]. Intuitively, a consensus statement of 51.0% might be less valid than a statement of for example 71.0%, so we will take this into account when publishing the consensus statements and recommendations. By reporting the percentages of votes at each recommendation that achieved consensus, we hope to be sufficiently transparent and to allow readers to draw their own conclusions regarding each

Table 3

Amount of questions per questionnaire on which consensus was reached during the first two voting rounds.

Questionnaire	Round 1, Consensus on Level of Evidence	Round 2, Consensus on Level of Evidence	Round 1, Consensus on Recommendation	Round 2, Consensus on Recommendation
PM				
WDPPM	0/8 (0.0%)	7/8 (87.5%)	5/8 (62.5%)	8/8 (100%)
MCPM	1/7 (14.3%)	5/7 (71.4%)	5/7 (71.4%)	7/7 (100%)
DMPM	9/51 (17.6%)	47/51 (92.2%)	38/56 (67.9%)	53/56 (94.6%)
PMP				
PMP-I	9/56 (16.1%)	50/56 (89.2%)	42/70 (60%)	65/70 (92.9%)
PMP-II	5/32 (15.6%)	31/32 (96.9%)	21/35 (60%)	34/35 (97.1%)

Table 4

Voting behaviour of expert centres as compared to all participating centres.

Level of Evidence		Round 1				Round 2			
		High	Moderate	Low	Very Low	High	Moderate	Low	Very Low
Expert Centres	DMPM	23.3%	38.8%	31.7%	6.2%	19.8%	45.8%	32.3%	2.0%
	MCPM	27.6%	30.5%	23.3%	18.6%	22.8%	49.7%	21.2%	6.3%
	WDPPM	21.7%	31.7%	32.1%	14.6%	13.4%	41.7%	44.0%	0.9%
	PMP-1	21.9%	39.3%	32.4%	6.3%	12.7%	54.0%	31.2%	3.1%
All Centres		19.4%	37.4%	35.0%	8.2%	12.7%	53.7%	31.3%	3.2%
Expert Centres	PMP-2	11.7%	39.7%	43.1%	5.6%	2.0%	46.4%	50.6%	1.1%
All Centres		10.9%	37.9%	44.8%	6.4%	2.3%	47.3%	49.3%	1.0%
Expert Centres									
Strength of Recommendation		Round 1				Round 2			
		High	Moderate	Low	Very Low	High	Moderate	Low	Very Low
Expert Centres	DMPM	32.5%	40.4%	8.1%	4.6%	33.1%	49.0%	4.9%	2.7%
	MCPM	57.6%	29.0%	7.6%	5.7%	65.1%	19.0%	6.9%	9.0%
	WDPPM	58.3%	30.4%	9.6%	1.7%	70.4%	21.8%	6.9%	0.9%
	PMP-1	29.3%	33.5%	11.4%	5.5%	31.8%	43.2%	8.7%	4.7%
All Centres		28.6%	34.9%	11.3%	6.4%	32.0%	40.7%	9.0%	5.1%
Expert Centres	PMP-2	24.4%	33.4%	25.4%	17.3%	23.6%	36.7%	21.6%	18.0%
All Centres		22.2%	32.5%	25.4%	20.4%	23.1%	36.5%	20.9%	19.4%
Expert Centres									

recommendation.

The first round of a Delphi consensus consists of an open-response questionnaire, to reduce tunnel vision among the SC and allow all possible answers in the next voting round [4]. We deviated from this by starting immediately with the regular multiple-choice questionnaire, although we tried to provide all relevant answers, and added the option 'other, please specify' in some questions. However, not starting with an open-response questionnaire may have caused limited answers and could thus reduce validity.

Evidence regarding the criteria for selection of experts and the number of experts that are required for a valid Delphi consensus statement is also lacking [6,7]. In general, the criteria for selection are left to the SC, as well as the number of experts invited. The available literature describes expert panels varying from only 4 to 3000 experts [4]. However, in contrast to popular belief that a higher number of participants increases the reliability of the consensus, there is very little evidence available that proves this [6]. Therefore, the decision about the size of the expert panel is pragmatically chosen by the SC. Obviously, quality of the experts is crucially important, so a large expert panel does not guarantee a high-quality consensus statement [4]. Our expert panels consisted of 38 (PM) and 80 (PMP) experts, and we consider this to be a sufficient number of experts, especially since our strict selection criteria for the experts. Due to the low prevalence of these diseases, only centres/experts with multiple years of experience with these diseases were included in the expert panel, increasing the quality of the consensus statements.

A downside of our expert panels is the limited number of medical oncologists: the PM expert panel contained only 1 medical oncologist, and the PMP expert panel contained 11 medical

oncologists. Since also questions regarding the use of systemic treatment were included in the questionnaires, the opinion of more medical oncologists could have affected the results concerning these questions.

Finally, although both the Delphi technique and GRADE system are useful methods for generating a consensus statement, it should not be forgotten that both methods require a judgment or opinion of the participants [6–8,10]. Although it is assumed that the judgment of a group is more valid than the judgment of an individual, neither judgment is necessarily absolutely correct.

Observations during voting

A decrease in the participation rate between the first and second rounds was observed. This is most probably due to the time consuming process given the length of the questionnaires and the summarized scientific literature that was added.

To stimulate responses, multiple e-mail reminders were sent to the participants during each voting round. This is recommended by other researchers that have used the Delphi technique as well, since the Delphi technique is more time consuming than a standard survey [7]. Other researchers that used the Delphi technique describe varying degrees of participation, from 75% to 100% [7]. The one study achieving a 100% response rate performed face-to-face interviews in the first round, which gives a personal touch and increases the likelihood of participants to complete the questionnaires. In our case, this was unfortunately not possible, since our experts originate from HIPEC-centres worldwide.

Consensus was reached during the first voting round in 60.0%–71.4% of the questions. This high level of disagreement shows us how disharmonized clinical practice of PM and PMP currently is,

and thus proves the requirement for this consensus-based guideline.

Next, it was noted that the quality of the scientific literature concerning the topics in the questionnaires was in general overestimated by the participants. The participants were asked to grade the scientific literature for each topic on a 4-point scale: 'high', 'moderate', 'low', 'very low'. According to the GRADE system, to rate the evidence as 'high' quality, it should contain well-performed RCT's or exceptionally strong evidence from unbiased observational studies. 'Moderate' quality of evidence would consist of RCT's with some limitations or strong evidence from unbiased observational studies. 'Low' quality of evidence is based on observational studies or RCT's with indirect evidence; 'very low' quality is based on unsystematic clinical observations or very indirect evidence. Most of the available scientific evidence consisted of cohort/observational studies, which had some scientific flaws and generally contradicted each other. Therefore, the SC considered most of the evidence as 'low' or even 'very low' quality of evidence. However, in many cases, the available evidence was rated by the participants as 'moderate' or even 'high' quality of evidence. So, there is a discrepancy between the perceived quality of evidence and the actual quality of evidence, which might have influenced the SoR given by the participants. We considered two possible explanations for this phenomenon. In the first place, we hypothesised that high volume centres, who are undoubtedly perfectly competent to phrase a recommendation regarding the clinical practice of treating PM and PMP, tend to overrate the evidence and the SoR due to difficulties in accurately perceiving the quality of evidence. This would be even more difficult for clinicians exclusively working with peritoneal surface malignancies, due to the general lack of high-quality evidence in this field. This hypothesis was proven to be unlikely, as the voting behaviour of the high-volume centres is comparable to lower volume centres, as shown in Table 4. Our second hypothesis is related to the panellist perspective on how much further research could advance in clarifying contentious issues in PM and/or PMP. The best study design is well-known to be the RCT that is unlikely to be successfully completed in PM and PMP due to their extremely low prevalence. In Table 1, the last column contains information regarding the implications of various possible combinations between SoR and LoE. The higher the LoE, the lower the chance that future research could change the estimate and consequently the SoR. Therefore, the panellist might have perceived that the SoR/LoE in the context of PM and/or PMP as something more related to the feasibility of better-designed future studies, rather than the type of study design. Consequently, the panellist might have overrated the SoR/LoE in several questions, considering the slim chance of studies of better quality being conducted in the years to come that will be able to change the current knowledge on PM and/or PMP.

Consensus was not yet reached on 15 questions after the two voting rounds, mainly due to defaults in the design of the questions. These questions were altered and will be included in the third voting round, after which we aim to achieve consensus as well. Four questions were added in the PMP – Unexpected Appendiceal Finding section, since we realized that we had not included all possible scenarios in the questionnaire.

Conclusion

To our knowledge, this consensus statement is currently the only consensus statement regarding PM and PMP. With this consensus statement, we seek to aid clinicians and HIPEC centres in

the care and treatment for patients with PM or PMP. Although there is some caveat in the execution of the Delphi and GRADE methods, we reckon the results of this consensus statement very successful, especially after seeing the disharmony in the recommendations during the first round, proving the requirement for this consensus statement. Therefore, we aim to provide a valid guideline in which consensus is reached in most of the questions. These results will be published in multiple manuscripts over the course of 2019 and 2020.

Declarations of interest

None.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ejso.2019.03.012>.

References

- [1] Rizvi SA, Syed W, Shergill R. Approach to pseudomyxoma peritonei. *World J Gastrointest Surg* 2018;10:49–56. <https://dx.doi.org/10.4240/wjgs.v10.i5.49>.
- [2] Boffetta P. Epidemiology of peritoneal mesothelioma: a review. *Ann Oncol* 2007;18:985–90. <https://dx.doi.org/10.1093/annonc/mdl345>.
- [3] Kusamura S, Baratti D, Younan R. The Delphi approach to attain consensus in methodology of local regional therapy for peritoneal surface malignancy. *J Surg Oncol* 2008;98:217–9. <https://doi.org/10.1002/jso.21059>.
- [4] Thangaratinam S, Redman CWE. The Delphi technique. *TOG* 2005;7:120–5. <https://doi.org/10.1576/toag.7.2.120.27071>.
- [5] Rowe G, Wright G, Bolger F. Delphi: a re-evaluation of research and theory. *Technol Forecast Soc Change* 1991;39:235–51. [https://doi.org/10.1016/0040-1625\(91\)90039-I](https://doi.org/10.1016/0040-1625(91)90039-I).
- [6] Powell C. Methodological issues in nursing research, the Delphi technique: myths and realities. *J Adv Nurs* 2003;41:376–82. <https://doi.org/10.1046/j.1365-2648.2003.02537.x>.
- [7] Keeney S, Hasson F, McKenna H. Consulting the oracle: ten lessons from using the Delphi technique in nursing research. *J Adv Nurs* 2006;53:205–12. <https://doi.org/10.1111/j.1365-2648.2006.03716.x>.
- [8] Guyatt GH, Oxman AD, Kunz R. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924–6. <https://doi.org/10.1136/bmj.39489.470347.AD>.
- [9] Grades of Recommendation. Assessment, development, and evaluation (GRADE) working group. Education and debate – grading quality of evidence and strength of recommendations. *BMJ* 2004;328:1490–4. <https://doi.org/10.1136/bmj.328.7454.1490>.
- [10] Guyatt GH, Oxman AD, Akl EA. GRADE guidelines: 1. Introduction – GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 2011;64:383–94. <https://doi.org/10.1016/j.jclinepi.2010.04.026>.
- [11] Higgins JPT, Green S. *Cochrane Handbook for systematic reviews of Interventions*. Wiley Online Library; 2008. ISBN 978-0-470-69951-5.
- [12] Huang X, Lin J, Demner-Fushman D. Evaluation of PICO as a knowledge representation for clinical questions. *AMIA Annu Symp Proc* 2006:359–63.
- [13] Guyatt GH, Oxman AD, Kunz R. GRADE guidelines: 2. Framing the question and deciding on important outcomes. *J Clin Epidemiol* 2011;64:395–400. <https://doi.org/10.1016/j.jclinepi.2010.09.012>.
- [14] Balshem H, Helfand M, Schünemann HJ. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol* 2011;64:401–6. <https://doi.org/10.1016/j.jclinepi.2010.07.015>.
- [15] Guyatt GH, Oxman AD, Vist G. GRADE guidelines: 4. Rating the quality of evidence – study limitations (risk of bias). *J Clin Epidemiol* 2011;64:407–15.
- [16] Andrews JC, Guyatt GH, Oxman AD. GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. *J Clin Epidemiol* 2013;66:719–25. <https://doi.org/10.1016/j.jclinepi.2012.03.013>.
- [17] Andrews JC, Schünemann HJ, Oxman AD. GRADE guidelines: 15. Going from evidence to recommendation – determinants of a recommendation's direction and strength. *J Clin Epidemiol* 2013;66:726–35. <https://doi.org/10.1016/j.jclinepi.2013.02.003>.
- [18] Villeneuve L, Passot G, Glehen O, Isaac S, Bibeau F, Rousset P, et al. And on behalf of the RENAPE Network, the RENAPE observational registry: rationale and framework of the rare peritoneal tumors French patient registry. *Orphanet J Rare Dis* 2017;12. <https://doi.org/10.1186/s13023-017-0571-y>.