PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

Title (Provisional)

A modified Delphi consensus regarding the clinical utility of triplet therapy in patients with metastatic hormone-sensitive prostate cancer patients in the UK

Authors

Glen, Hilary; Bahl, Amit; Fleure, Louisa; Clarke, Noel; Jain, Suneil; Kalsi, Tania; Khoo, Vincent; Mobeen, Junaid

---

VERSION 1 - REVIEW

Reviewer **1**

Name **Berrens, Anne-Claire**

Affiliation **The Netherlands Cancer Institute - Antoni van Leeuwenhoek Hospital**

Date **23-Jul-2024**

COI **None**

The authors have done a modified Delphi to specifically analyze the general consent in the UK on triple therapy with the aim to reach consensus on 4 main topics and formulate recommendations.

Abstract

Is well written and clear

Background

- It states there is no head-to-head comparison between treatments and some studies using triplet therapy show no prolongation of OS. Could you elaborate why this Delphi is (for now?) the answer when a prospective trial to study (and possibly head-to-head compare) the effect of this triplet therapy seems more appropriate?

Methods

The methodology is described in detail and the study is well-designed. A few comments/questions:

- Please add when the last update of the literature search was performed

- Was the steering committee the same during the initial meetings and the meetings to discuss the results?

- What software was used to distribute the statements?

- Was it agreed a priori what the level of consensus should be so further survey rounds were unnecessary?

- What exactly is considered a nominal fee and could this be a conflict of interest?

- What was it based on that a statement was considered an important statement (on which a recommendation was based)

- Please consider adding to the text where the 11-20 years of experience were in. (working in healthcare?/working in this particular role?)

- Please explain why the first 4 statements in Topic A are included. It reads like the participants are asked if they agree if other studies are proof of level 1 evidence or if they agree with outcomes of a phase III RCT. That seems a strange question to ask. Or is it to ask what they base their knowledge and decisions on?

Results

The tables and figures are clear and a nice visualization.

Discussion, limitations&strengths

Some limitations are discussed in the paragraph on the topic but general limitations are missing (except for the list)

- The 4-point likert scale is considered a strength according to the authors, but it does not take into consideration that sometimes the participant may be neutral and is now "forced" to slightly agree or disagree. Please add to the limitations.

- This is especially important as the overall agreement is based on both strongly agree and tend to agree. The authors may take into consideration to take the distribution of the answers into account before concluding there was strong agreement (as e.g. statement C27 and D35 both have an agreement of 98% but in C27 82% strongly agrees and in D35 only 52%).

- Gillesen et al. (Eur J Cancer, 2023) did a modified Delphi during the APCC conference. Although from 2023, the authors might consider to compare and add to the discussion section.

Reviewer	**2**
Name	**Pezzoli, Marta**
Affiliation	**University of Florence**

Date 20-Sep-2024

COI None

The paper presents the results of a well-structured Delphi consensus. The methods and results are clearly described and the discussion clarifies the essential aspects of the consensus.

VERSION 1 - AUTHOR RESPONSE

Reviewer 1		
Background		
It states there is no head-to-head comparison between treatments and some studies using triplet therapy show no prolongation of OS. Could you elaborate why this Delphi is (for now?) the answer when a prospective trial to study (and possibly head-to-head compare) the effect of this triplet therapy seems more appropriate?	This is an incredibly valid and prescient point, which was discussed by the group within their steering group meetings. A trial would be the ideal way to establish evidence to see the effects of triplet therapy, especially a head-to-head trial. However, trials are complex in terms of the funding, resources, and time required to complete them. This leaves clinicians treating patients presently, and in the immediate future, to wonder which treatments are the most appropriate and what optimal care should or could look like for their patients. A modified Delphi method was chosen as it was felt this method would provide more immediate answers to help guide decision making. The method would allow for the aggregation of opinion data from multiple clinicians across the UK to establish what should be done now for patients. The Delphi method is acknowedged as a reliable way to examine group consensus and to allows for the exploration of issues in the absence of appropriate empirical data[1–3]. It is not intended to replace clinical trial data, but act as a stop gap to aid practice until further evidence is generated.	This has been stated following the aim (p. 5)
Methods		
Please add when the last update of the literature search was performed	The literature review was performed between the 7th-9th of December 2022	This has been added to the methods (p. 5)

Was the steering committee the same during the initial meetings and the meetings to discuss the results?	Yes, the steering committee (the papers authors) were the same individuals in both meetings. This group helped develop the aim of the project and directed the project at each stage in terms of statement development, analysing the results, creating recommendations, and drafting the manuscript.	This has been clarified further in the methods (p. 5)
What software was used to distribute the statements?	The survey was initially distributed as an MS Forms survey by the steering group via email. The independent distributor (SERMO) use their own online platform for the survey, but the link was distributed via emails to potential responders on their database.	None
Was it agreed a priori what the level of consensus should be so further survey rounds were unnecessary?	Yes, the consensus threshold was set a priori as 75%. The survey was open for 2 months, with a target of 100 responses and 90% of statements passing the consensus threshold. It was agreed if these criteria were fulfilled then the no further rounds would be necessary.	This has been clarified further in the methods (p. 6)
What exactly is considered a nominal fee and could this be a conflict of interest?	The nominal fee for the respondents was based on “fair market value” and was set by SERMO in line with market research standards. The fee paid varied by role as follows: ONS £14, Hospital pharmacist £14, Geriatric Medicine £17, Consultant Urologist £17, Medical/Clinical Oncologists £19. The fee was considered to be an incentive to complete the survey, rather than a conflict of interest. The responders were not aware of the projects sponsor (until after completing the survey), no branded products were included in the survey and so they were not biased by this.	None

What was it based on that a statement was considered an important statement (on which a recommendation was based)	Statements were considered important based on the levels of consensus and disconsensus (those that achieved strong or very strong consensus vs those with lower agreement which may highlight educational needs within certain responder groups) and were selected by the group during their discussions in the second steering group meeting. The group then considered these statements in line with the literature, for example where high levels of agreement were supported by the literature and where low levels of agreement may indicate a need to disseminate certain literature to bridge knowledge gaps. Through their discussions, the group came to a consensus on which statements they felt could be developed into practical recommendations that would have the greatest impact on clinical practice and patient outcomes. The discussion section reflects the conversations had by the group in that second meeting and highlights the points which led to the selection of statements and development of the recommendations.	This has been clarified further in the methods (p 7.)
Please consider adding to the text where the 11-20 years of experience were in. (working in healthcare?/working in this particular role?)	Thank you for highlighting this. We defined this as years of experience in role – so it would be 11-20 years working as an oncologist etc.	This has been specified in the results (p. 7)
Please explain why the first 4 statements in Topic A are included. It reads like the participants are asked if they agree if other studies are proof of level 1 evidence or if they agree with outcomes of a phase III RCT. That seems a strange question to ask. Or is it to ask what they base their knowledge and decisions on?	These statements were added to set the scene for the project and highlight importance of these studies. They were also included as a way to test the knowledge base of the responders to see if there was awareness of the studies. Fluctuations were seen in agreement levels between responder roles for these statements. For example, geriatric medicine showed agreement below the consensus threshold for statement 3, while ≥88% of all other responders agreed. This could highlight lower awareness of the ARASENS data among clinicians in geriatric medicine.	None
Results		

The tables and figures are clear and a nice visualization.	N/A	None
Discussion, Limitations and Strengths		
The 4-point likert scale is considered a strength according to the authors, but it does not take into consideration that sometimes the participant may be neutral and is now "forced" to slightly agree or disagree. Please add to the limitations.	This is a really good point. Throughout the discussion in both of the steering group meetings it was considered if, for some statements, certain responder groups were more or less qualified to answer questions, and would therefore be “neutral” in their opinions. Based on the groups initial discussions it was felt it was better to avoid middle option bias than to include a neutral or don’t know option which could dilute the number of individuals providing their opinion. The group were keen to try and get as many opinions as possible and then analyse trends between role to see if there may have been certain responder types who showed variation as a whole. However, we acknowledge some individuals may have selected slightly agree or disagree based on trying to remain neutral.	This has been discussed further within the strengths and limitations (p. 14)
This is especially important as the overall agreement is based on both strongly agree and tend to agree. The authors may take into consideration to take the distribution of the answers into account before concluding there was strong agreement (as e.g. statement C27 and D35 both have an agreement of 98% but in C27 82% strongly agrees and in D35 only 52%).	Again this is an incredibly valid point which we acknowledge. The group did consider this when they were discussing the results and developing recommendations. When reporting Delphi results it is standard to primarily report descriptive statistics and the meand agreement[4,5], hence the focus on this is the body of the manuscript. However, for transparency the full distribution is provided clearly in Table 1.	The fact that distribution was considered by the group has been clarified (p. 7)
Gillesen et al. (Eur J Cancer, 2023) did a modified Delphi during the APCC conference. Although from 2023, the authors might consider to compare and add to the discussion section.	Thank you for highlighting this paper, it has some very interesting results which are very pertinent to this study. We have added some comparisons within the discussion as suggested.	A comparison has been included (p. 10 and 11)
Reviewer 2 (From Responses to Questions)		
The paper presents the results of a well-structured Delphi consensus. The methods and results are clearly described, and the discussion clarifies the essential aspects of the consensus.	Many thanks for you review of our paper, and for your kind comments.	None

References:

1. Shang Z. Use of Delphi in health sciences research: A narrative review. *Medicine (United States)* 2023; 102: E32829.
2. Woolley AW, Chabris CF, Pentland A, et al. Evidence for a collective intelligence factor in the performance of human groups. *Science (1979)* 2010; 330: 686–688.
3. Kurvers RHJM, Herzog SM, Hertwig R, et al. Boosting medical diagnostics by pooling independent judgments. *Proc Natl Acad Sci U S A* 2016; 113: 8777–8782.
4. von der Gracht HA. Consensus measurement in Delphi studies. Review and implications for future quality assurance. *Technol Forecast Soc Change* 2012; 79: 1525–1536.
5. Milevska-Kostova N, Dunn W. Delphi Analysis. In: Zaletel-Kragelj L, Bozikov J (eds) *Methods and Tools in Public Health: A Handbook for Teachers, Researchers and Health Professionals*. Lage, Germany: Hans Jacobs Publishing Company, 2010, pp. 423–436.