PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	The role of patient preferences in clinical practice guidelines: a multiple methods study using guidelines from oncology as a case
AUTHORS	Gärtner, Fania; Portielje, Johanneke; Langendam, Miranda; Hairwassers, Desiree; Agoritsas, Thomas; Gijsen, Brigitte; Liefers, Gerrit-Jan; Pieterse, A.H.; Stiggelbout, Anne

VERSION 1 - REVIEW

REVIEWER	Robert A. Smith American Cancer Society, United States
REVIEW RETURNED	04-Aug-2019

GENERAL COMMENTS	Bmjopen-20019-032483 Page 5, line 24: Would the authors describe the number of available cancer related CPGs, and how they chose the three CPGs to evaluate? The authors state that they selected modules that contained at least on preference-sensitive decision based on the presence of a weak recommendation, or perhaps a strong recommendation that should have been a weak recommendation? The question is, were there other CPGs that met these criteria, and if so, why were these selected over others? It may be that a convenience sample is quite OK as long as the CPG and modules meet the requirements, but some explanation would be appropriate. Page 5, line 39: I suggest that the authors make it clear that they've independently scored the scored the guideline, apart from any use of GRADE or other grading scheme that may (or may not) have been used by the guideline developer. The expression, "…from the Recommendation section, based on the phrasing used…." Threw me off. Did any of these guideline developers use GRADE? Doesn't mean that there couldn't be a difference of opinion, so it would be useful to mention that as well. Page 8: Regarding the endocrine therapy module, are the authors willing to say (if they know) why none of the guideline developers were willing to be interviewed? Page 13, last paragraph. Not a critique, but do the authors have any advice for the GRADE developers based on their findings? Would adherence to GRADE be more likely if there were a formal opportunity for the guideline developers to acknowledge the preferences and concerns about the state of the science and the limitations of GRADE rules and terminology? Boxes: Box 1: GRADE Approach. The authors might wish to acknowledge that guideline developers may feel some inherent tension between

	adherence to GRADE, which more often than not will leads to recommendation language that sounds more tepid and doubtful to end users than is intended by the guideline developers. Ironically, if most patients undergo a procedure, but there is no data on their personal preferences, by GRADE standards the recommendation for the procedure must be weak, not because it is, but because we don't have good data on patient preference. The guideline developers will be making their recommendations based on what they judge to be most effective for the patient. It is common for an intervention to get ahead of the evidence, but doctors may be reluctant to suggest to patients that not undergoing the intervention may be just as beneficial, and less harmful, when they are unclear about that as well. GRADE leans towards that being a patient decision, which is correct, but there is evidence that guideline developers may face difficulty with this. I note that this tension is described in your discussion with panel members. Also, the authors address some of the problems with GRADE terminology on page 12 line 16.

REVIEWER	Anna R Gagliardi Toronto General Hospital Research Institute and University of Toronto, Canada
REVIEW RETURNED	25-Aug-2019

GENERAL COMMENTS	I enjoyed reading this manuscript, which makes a needed contribution to the body of knowledge on how to make guidelines more implementable by incorporating information on uncertainties and preferences that will support patient-provider communication and decision-making. Some additional details will strengthen the reporting of methods and findings, and help readers understand the implications and how then can improve practice based on the findings. INTRODUCTION - Define and describe both uncertainties and preferences (i.e. pertaining to what, specifically?) METHODS - Rationalize choice of methods and provide citation for all three of mixed-methods, content analysis and qualitative interviews, plus reporting standards that were adhered to for all three - Unclear if this was a true mixed-methods study (see Fetters); if so, then essential details are lacking (i.e. was design convergent? Were quantitative and qualitative methods prioritized equally? Were data collection concurrent or sequential? What approaches/techniques were used to integrate data?). Perhaps instead it was a multiple methods study? - The 5 themes are noted in the first paragraph (Approach/Design) and again under Data extraction and analysis, which is repetitive. Suggest including these details only under Data extraction/analysis AND including this information in a table rather than text as a table would likely be more readily and easily digested by readers. Explain the origins of the themes/components of coding scheme. - Rationale for choice of CPGs is noted in the first paragraph (Approach/Design) and again under Selected CPG modules; suggest inclusion only in the latter section

	- Clarify if only the modules were examined (in which case, could important content have been missed) or also content that may have appeared in appendices or online? If not, this should be cited as a limitation - For qualitative methods, specify technique used for sampling, did participants provide informed consent, and who conducted interviews and what was their training? RESULTS Not sure if the information in boxes, which represents only partial data, offers readers only good examples that they can emulate? It would be ideal if the authors could compile all the good examples for all 5 themes DISCUSSION The first paragraph refers to users informing patients about trade-offs as part of the SDM process. While I believe the authors do mean that patients should be actively engaged, this phrasing reads as if clinicians tell patients what to do. Suggest reframing these points to emphasize patient engagement in decision-making based on patients who are fully informed about uncertainties and have ample opportunity to express and discuss preferences, and also to emphasize that patients may be directly accessing the guidelines and inclusion of this content makes guidelines more implementable to/for patients themselves. The Discussion reiterates results and notes some limitations. Suggest adding and discussion the implications, i.e. what should developers do to address the gaps identified by this research? Is it feasible for them to improve guidelines in this way? What support might they need to do so? What are suggestions for ongoing research? ABSTRACT Incorporate details about rationale/methods as noted above OVERALL SUMMARY BOX Clarity and grammar of the 3 bullets corresponding to strengths and limitations could be improved

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Reviewer Name: Robert A. Smith

Institution and Country: American Cancer Society, United States

Please state any competing interests or state ‘None declared’: None

Comments:

Thank you for your suggestions. We have included most of them in the manuscript, and trust that it is more clear now.

Page 5, line 24: Would the authors describe the number of available cancer related CPGs, and how they chose the three CPGs to evaluate? The authors state that they selected modules that contained at least on preference-sensitive decision based on the presence of a weak recommendation, or perhaps a strong recommendation that should have been a weak recommendation? The question is, were there other CPGs that met these criteria, and if so, why were these selected over others? It may be that a convenience sample is quite OK as long as the CPG and modules meet the requirements, but some explanation would be appropriate.

We indeed selected a convenience sample of guidelines for cancer types of which we knew they are prevalent; their treatment include preference-sensitive decisions, based on our own research and the literature; and treatments span various specialties. We agree that there other guidelines may have fitted these criteria as well. A priori we had planned to study head and neck tumours (treatment includes preference sensitive decisions due to lack of evidence) as well, but the guidelines were in the process of updating, which would have complicated the study. We have now added this information in the text:

“We selected three tumour-specific CPGs, and of each we selected all content of two modules to include in our analysis (i.e., the sections of the CPGs that address specific treatments or patient groups). We selected a convenience sample of modules for prevalent cancers that we expected to contain at least one preference-sensitive decision, calling for a weak recommendation. This expectation was based on earlier research from our group (11,13,15), views from the oncology experts on our research team, and/or on the availability of literature on SDM and decision aids for the treatment in that module. Each of the modules included more than one recommendation.” (p 5, lines 29-36).

Page 5, line 39: I suggest that the authors make it clear that they’ve independently scored the guideline, apart from any use of GRADE or other grading scheme that may (or may not) have been used by the guideline developer. The expression, “…from the Recommendation section, based on the phrasing used….” Threw me off. Did any of these guideline developers use GRADE? Doesn’t mean that there couldn’t be a difference of opinion, so it would be useful to mention that as well.

We realized that the way we formulated the sentence caused confusion. The categories mentioned between parentheses refer to the phrasing used in the Recommendations section of the guideline. The CPGs developers never explicitly used GRADE. We have now added a statement at the end of 2.1.1 to clarify this: “In none of the modules explicit reference was made to GRADE.”

We have also rephrased the paragraph in 2.1.2 that caused the confusion. It now reads: “First, we scored the strength of the recommendation (strongly in favour/ weakly in favour/ neutral / weakly against/ strongly against a specific option) for each treatment option described in the Recommendation section of the CPG. Scoring was solely based on the phrasing used in that section. The categories strong and weak that we used are in line with GRADE. We added the ‘neutral’ category if a weak recommendation for more than one option was given.” (p6, l 5-11) We hope that the text now is clear.

Page 8: Regarding the endocrine therapy module, are the authors willing to say (if they know) why none of the guideline developers were willing to be interviewed?

We have now added what information we have (p8, lines 20-24): “For one module (adjuvant endocrine therapy in breast cancer), only one of the clinician panel members indicated to have time to participate. After an interruption due to a clinical urgency she did not want to resume the interview because she found the questions too critical. Therefore only the IKNL supervisor and the patient panel member were interviewed.”

Page 13, last paragraph. Not a critique, but do the authors have any advice for the GRADE developers based on their findings? Would adherence to GRADE be more likely if there were a formal opportunity for the guideline developers to acknowledge the preferences and concerns about the state of the science and the limitations of GRADE rules and terminology?

We are not sure we understand this suggestion, for if, as we suggest, if panels “are more transparent about benefits and harms and their probabilities, as well as about the preferences of the guideline panel members, and their assumptions about patient preferences” they would adhere to GRADE. We have therefore left the statement as is. Nevertheless, also following the suggestion below regarding Box 1, and comments from reviewer 2, we have changed this final paragraph which now reads:

“In sum, our analysis points to a lack of transparency in the CPG development process about benefits, harms and their probabilities, the preferences of the guideline panel members, and their assumptions about patient preferences. Awareness needs to be created among CPG-developers that their judgments of the balance of benefits and harms are value-laden, and that variation exists in these judgments, among both clinicians and patients. Clear instructions and training to enhance knowledge and implementation of GRADE might improve the acknowledgement of preference-sensitive decisions in guidelines and support shared decision making. This will help avoid what McCartney feared in his 2016 Analysis in the BMJ: “there is the danger of guideline recommendations being applied to people who do not place the same values on those recommendations as their clinician (…)”.(23)

Box 1: GRADE Approach. The authors might wish to acknowledge that guideline developers may feel some inherent tension between adherence to GRADE, which more often than not will leads to recommendation language that sounds more tepid and doubtful to end users than is intended by the guideline developers. Ironically, if most patients undergo a procedure, but there is no data on their personal preferences, by GRADE standards the recommendation for the procedure must be weak, not because it is, but because we don’t have good data on patient preference. The guideline developers will be making their recommendations based on what they judge to be most effective for the patient. It is common for an intervention to get ahead of the evidence, but doctors may be reluctant to suggest to patients that not undergoing the intervention may be just as beneficial, and less harmful, when they are unclear about that as well. GRADE leans towards that being a patient decision, which is correct, but there is evidence that guideline developers may face difficulty with this. I note that this tension is described in your discussion with panel members. Also, the authors address some of the problems with GRADE terminology on page 12 line 16.

We now touched upon this point in our final paragraph, as stated above, and added a sentence in this respect to the Box, with a reference to a study by Alexander et al (2016): “Tension has been shown to occur between adherence to GRADE and the wish to make a strong recommendation out of

conviction that a treatment is beneficial, despite the evidence quality or certainty being (very) low (Alexander et al. 2016)".

We addressed preference for active treatment on p 13, lines 3-5.

Reviewer: 2

Reviewer Name: Anna R Gagliardi

Institution and Country: Toronto General Hospital Research Institute and University of Toronto, Canada

Please state any competing interests or state 'None declared': None declared

Comments:

I enjoyed reading this manuscript, which makes a needed contribution to the body of knowledge on how to make guidelines more implementable by incorporating information on uncertainties and preferences that will support patient-provider communication and decision-making. Some additional details will strengthen the reporting of methods and findings, and help readers understand the implications and how then can improve practice based on the findings.

Thank you for your judgment of our manuscript as being needed. And thank you for these suggestions, which helped us improve the clarity of our manuscript.

INTRODUCTION

- Define and describe both uncertainties and preferences (i.e. pertaining to what, specifically?)

We have now specified "preferences for treatments or outcomes of treatment" on p 4, line 7. We refer to uncertain benefits and uncertain evidence in the Introduction, and doubt whether this needs further explanation. We are unsure about what other use of uncertainty in the Introduction you refer to?

METHODS

- Rationalize choice of methods and provide citation for all three of mixed-methods, content analysis and qualitative interviews, plus reporting standards that were adhered to for all three .

Unclear if this was a true mixed-methods study (see Fetters); if so, then essential details are lacking (i.e. was design convergent? Were quantitative and qualitative methods prioritized equally? Were data collection concurrent or sequential? What approaches/techniques were used to integrate data?). Perhaps instead it was a multiple methods study?

We realize now that we did not truly provide a mixed-methods study, nor a formal content analysis. We performed a qualitative analysis of the content of guidelines, and validated this in semi-structured interviews with panel members. We have now used the term 'multiple methods' in the title and stated the following in the Abstract: "We performed a qualitative analysis of the content of the CPGs , and verified the results in semi-structured interviews with CPG panel members" and deleted the term 'mixed methods' entirely from the Methods. We replaced the term "content analysis" elsewhere with "qualitative analysis of the CPG content".

- The 5 themes are noted in the first paragraph (Approach/Design) and again under Data extraction and analysis, which is repetitive. Suggest including these details only under Data extraction/analysis AND including this information in a table rather than text as a table would likely be more readily and easily digested by readers. Explain the origins of the themes/components of coding scheme.

Thank you for this suggestion. We have now omitted listing the themes in the first paragraph of the Methods section, as suggested. We will put them in a Table if the Editor wishes so, they are listed under 2.1.2. We added a comment on the origins of the themes (p 5, l44 to p6, l 2)), which was mostly based on GRADE, apart from theme 4. CPG Panel Preferences. The origin of this theme was already explained in the text (p6, l36-38).

- Rationale for choice of CPGs is noted in the first paragraph (Approach/Design) and again under Selected CPG modules; suggest inclusion only in the latter section

We have deleted the information from the first paragraph as suggested.

- Clarify if only the modules were examined (in which case, could important content have been missed) or also content that may have appeared in appendices or online? If not, this should be cited as a limitation

We used all content of the modules, which was all online. Only one module had an appendix containing a Chapter on Shared Decision Making, but this was generic and unrelated to the recommendations. We have now added "all content" (p 5, line 29): "We selected three tumour-specific CPGs, and of each we selected all content of two modules to include in our analysis."

- For qualitative methods, specify technique used for sampling, did participants provide informed consent, and who conducted interviews and what was their training?

P 7, line 19: We have now added "all" to panel members, as stated in the manuscript, we had a minimum we needed. We added information on the training of the interviewer as well (p 7, line 40-41):

“One interviewer (FG) trained in qualitative research methods and highly experienced in interviewing carried out all the interviews.”

RESULTS

Not sure if the information in boxes, which represents only partial data, offers readers only good examples that they can emulate? It would be ideal if the authors could compile all the good examples for all 5 themes

We prefer to leave it at these examples, given the limited word count, and because GRADE provides many good examples. The examples shown are aimed at highlighting where things go awry.

DISCUSSION

The first paragraph refers to users informing patients about trade-offs as part of the SDM process. While I believe the authors do mean that patients should be actively engaged, this phrasing reads as if clinicians tell patients what to do. Suggest reframing these points to emphasize patient engagement in decision-making based on patients who are fully informed about uncertainties and have ample opportunity to express and discuss preferences, and also to emphasize that patients may be directly accessing the guidelines and inclusion of this content makes guidelines more implementable to/for patients themselves.

We have rephrased this (p 12, lines 5-9): “This makes it difficult for the users to judge the appropriateness of the strength of the recommendation. Further, it may hinder patient engagement in decision-making, which requires that patients are fully informed about the trade-offs. Moreover, patients may be directly accessing the guidelines, and inclusion of this information makes guidelines also more useful to them.”

The Discussion reiterates results and notes some limitations. Suggest adding and discussion the implications, i.e. what should developers do to address the gaps identified by this research? Is it feasible for them to improve guidelines in this way? What support might they need to do so? What are suggestions for ongoing research?

We have added the following, also in reply to reviewer 1 (p 13, lines 24-31):

“In sum, our analysis points out a lack of transparency in the CPG development process, both about benefits and harms and their probabilities and about the preferences of the guideline panel members and their assumptions about patient preferences. Awareness needs to be created among CPG-developers that their judgments of the balance of benefits and harms are value-laden, and that variation exists in these judgments, among both clinicians and patients. Clear instructions and training to enhance knowledge and implementation of GRADE might improve the acknowledgement of preference-sensitive decisions in guidelines and support shared decision making. This will help avoid what McCartney feared in his 2016 Analysis in the BMJ: “there is the danger of guideline

recommendations being applied to people who do not place the same values on those recommendations as their clinician (…)".(23)

We already had recommendations regarding further research (p13, line 14-16) and feel these describe our main recommendations.

ABSTRACT

Incorporate details about rationale/methods as noted above

We made changes accordingly.

OVERALL SUMMARY BOX

Clarity and grammar of the 3 bullets corresponding to strengths and limitations could be improved

We rephrased the strengths and limitation of the study.

VERSION 2 – REVIEW

REVIEWER	Robert A. Smith American Cancer Society, U.S.
REVIEW RETURNED	14-Sep-2019

GENERAL COMMENTS	Reviewer 1, Robert Smith In their response, the authors state, "Clear instructions and training to enhance knowledge and implementation of GRADE might improve the acknowledgement of preference sensitive decisions in guidelines and support shared decision making. This will help avoid what McCartney feared in his 2016 Analysis in the BMJ: "there is the danger of guideline recommendations being applied to people who do not place the same values on those recommendations as their clinician (…)".(23)" I agree that guideline developers need to be fluent in GRADE, and quite often there are several on the panel that are, while most are not, or there is a cognitive tug-of-war with GRADE methodology and their own interpretation of the balance of benefits and harms and the grading of recommendations. But regarding McCartney's quote, where the emphasis implies that testing is being pushed to a population that is not as enthusiastic about the intervention as the guideline developers, it also is true that guideline developers can impose their opinion that the balance of benefits to harms is not favorable, where the people the guideline is being applied to clearly disagree. Case in point, the USPSTF states that breast cancer screening in the 40s is a preference sensitive decision, but they add that the benefits of screening in that decade are low. Yet,

	the majority of U.S. women begin screening soon after their 40th birthday. Modern evidence clearly shows a substantially better benefit from today's service screening than is evident from the trial data, and systematic reviews have acknowledged this. So, not only is the balance of benefits and harms subject to the reference benefit and harms, but we would have to agree that judgment plays a much larger role in how the target population experiences the bundle of unexpected events that fall into the harm group. While the USPSTF relies on the RCTs, which have well documented limitations for this age group, they also judge being recalled for further imaging as a significant harm. Most women do not. The authors should consider noting, with respect to their final sentence in the excerpt above, that McCartney's "danger" only considered one side of the possibility that guideline developers could issue a recommendation that was at odds of the target population's preferences. By the way, I'm not insisting on this change. It is just something for the authors to consider.

VERSION 2 – AUTHOR RESPONSE

Dear Sir, Madam,

Thank you for your positive decision letter. The only suggestion provided in this second round was by prof. Smith, on qualifying the quote with which we summarized at the end. He feels the quote is one-sided, and refers to the USPSTF to show that patients may wish more healthcare than guidelines recommend.

We have reread the McCarthy piece, which we still feels does mostly pertain to clinical guidelines (her examples involve heart failure and depression, multimorbidity, and the author is not a screening physician, it seems). I have copied the entire paragraph, with our quote referring to the section in brackets. We still feel that the paragraph covers what we are saying in our conclusion .

"Finally, and most importantly, [there is the danger of guideline recommendations being applied to people who do not place the same values on those recommendations as their clinician], or indeed those intended by the guideline creators. Evidence reviews by organisations such as National Voices have found

that shared decision making engages people in their care and leads to decisions which patients find most appropriate to them.7 Surveys have shown that most patients wish either to share decision making with their clinicians or to take the decisions themselves.8 Guidelines should enable, not subvert, this process.