Responses to stakeholder comments

Stakeholder	Comment	Response
Question 1 Are the objectives and methods of this evidence review clear?		
Howard Tracer, USPSTF	Yes	Thank you. No changes requested.
Somto Ibezi, Black Physicians Canada	Yes	Thank you. No changes requested.
Kathryn Boyd, Nova Scotia Health	Yes. Objectives clearly stated with clinical question identified. Methods were extensive and clearly defined including criteria, processes used and aim/goal of research.	Thank you. No changes requested.
Charlotte Yong-Hing, BC Cancer Agency	Yes.	Thank you. No changes requested.
Dee Anne Carol, Alberta Health	Yes	Thank you. No changes requested.
Julie Brunet, Québec INESSS (Institut national d'excellence en santé et en services sociaux)	No. The Background in the Abstract should be more elaborate if this paper is published as a stand-alone paper like the previous version in 2018 (Background and Purpose) Objective in the main section is clear. Methods in Abstract section need to be clarified. Please add a sentence to increase comprehension for non-familiar user of Health State Utility assessment and Health State utility values (HSUVs) literature type, since this kind of information was not in your previous report of 2018. ex: Utilities are measured on a scale of 0 to 1, in which 0 represents a health state equivalent to death and 1 represents a health state of perfect health. Ref:Chang et al. Explaining Health State Utility Assessment JAMA. 2020;323(11):1085-1086. doi:10.1001/jama.2020.0656 This will allow everybody to understand the numbers presented in the FINDINGS section of the Abstract. Main section Eligibility (p7) Table 1. Inclusion and exclusion criteria: why Intervention in the standard PICO was changes for Exposure? Under table 1 Eligibility criteria TTO, tine tradeoff?-time tradeoff? Literature Search (p10)	Thank you for the excellent comments/suggestions. We have expanded the background in the abstract as suggested. We added into the abstract and introduction information about interpreting utility values and what could be used as a minimally importance difference to the public. For our PICOs, we used exposure vs. intervention because for this review we are not interested in the effects of an intervention (e.g. effectiveness of a decision aid), but rather data on how an exposure to an outcome (anticipated or experienced) is valued. We corrected our typo for time trade-off. Our search update used a revised search to add terms for HSUVs but would have only found those published

	Searches were restricted by language to include full texts published in English and French, with a publication date of 2017 onwards. …. To capture studies on HSUVs published between 2000 and 2017 (for utilities related to screening outcomes) or between 2014 to 2017. Should be: 2000-current and 2014- current or 2000-2023 and 2014-2023? since the previous documents (2018) indicates that the search for the review of women's values and preferences was updated on December 2017? The explanation in this section doesn't match the information presented in the Methods section-Eligibility criteria TABLE 1-for timing (Non-HSUV studies & HSUVs in exposures 1-5: Published on or after 2000/ For HSUVs in exposures 6 & 7: published 2014 or later) Risk of Bias Assessments (p14) The validated tool used to assess the risk of bias is not clearly stated. GRADE guidance is normally used for systematic review not primary research evaluation? You refer to Zhang et al 2019 for the RoB assessment, but there is 9 questions in supplement 2&3 (table S2.7 & S3.7) and 7 in Zhang et al article. Also, there is no confounding factors question in your évaluation… Please elaborate on this way to perform a RoB analysis. Since you have several RCT in your selection, the use of a specific validated tool such as RoB2 should be more appropriate? Data analysis (p15) HSUV data "If variance measures were not reported we used one from a similar study" This strategy seems strange to me, but I'm not familiar with HSUV studies... Please clarify this decision.	2017 onwards, so we also searched references lists in systematic reviews on utilities for previously published studies from either 2000 (screening outcomes) or 2014 (for treatment related utilities). We have added some clarity around this. We have clarified that there is no published validated risk of bias tool for these questions, and how we modified the questions proposed in the GRADE guidance. Because we were not evaluating the effects between groups that could differ by potential confounders, we did not ask a question specific to this, though we did rate as a risk selected populations (all attending screening), and considered age differences and findings for at-risk groups during the synthesis. We've followed guidance from the Cochrane handbook for dealing with missing data (eg imputing data from similar studies), as long as not very many studies required this, and added this citation and mention that this was used during subgroup analysis (for risk of bias) and findings removed if they differed from others.
Davina Gallagher, BC Cancer Agency	Yes	Thank you. No changes requested.
Modupe Tunde-Byass, Black Physicians Canada	Yes. The objectives and methods were clearly stated , the review examined the relative importance placed by women age >/=35years on the potential outcomes of breast-cancer screening. Traditionally, women <50 years were not routinely screened. The results of the review revealed a population of women who would not have appreciated the importance of screening. However, such women could be getting their information from other "trusted" sources like social media or communities to make informed decisions.	Thank you. We found several studies enrolling women aged 35-50 and for this review it is not considered a limitation that the women haven't experienced the intervention since this has been shown to lead to belief perseverance, whereby the behavior/attitudes persist despite being

	The review in addition to the well-known databases, screened submissions by stakeholders and reference list. These may not capture the preferences of women between age 35 to 50.	given information that may otherwise contradict their values. The studies were all conducted in the context of making a decision about screening.
Heather Bryant, Canadian Partnership Against Cancer	Yes. The objective is very succinctly described, and it is of direct relevance to the current Task Force review. Selection of the sample is well-described, as is the search process and the methods of data analysis. It was reassuring to see that the sample included data from a wide range of countries and settings, as subjective weightings of perceived risks and benefits may be influenced by public health or media messages; the fact that the results reflect so many contexts adds to the robustness of the findings.	Thank you. No changes requested.
Jennifer Payne, Nova Scotia Health	Not selected: From the background section: 'This systematic review update examined the relative importance placed by patients aged ≥ 35 years on the potential outcomes of breast-cancer screening.' From the conclusion section: 'Evidence across a range of data sources on how informed patients value the potential outcomes from breast-cancer screening will be useful during decision-making for recommendations. Further, the evidence supports providing easily understandable information on possible magnitudes of effects to enable informed decision-making.' These two statements are not consistent. If the purpose was to determine the relative importance, and to separately determine the value of that type of information in decision-making, then the objective should indicate so. Given the multi-disciplinary nature of the audiences interested in this work, and presumably the 'stand-alone' nature of the document in question, it would help if the methods statement elaborated on the approach (utilities) so that when the reader moves to the results section, there is some basis for interpreting the numerical values (suggestion: insert a sentence prior to the sentence 'Our main analysis for utilities). I've said previously I have concerns with the lower age threshold in the 30s given that these individuals are not currently eligible for average risk screening anywhere in the country so their insight is less helpful (ie it's theoretical). Similarly, the results for individuals in their 80s is also less of interest.	Thanks for your comments. We agree that our objective did not include assessment of use/effectiveness of information provision during decision making. We have revised our conclusions to state that findings support that all outcomes are important and that information on the likelihood of their occurrence may be necessary to enable informed decision making. We have added details to the abstract and main text to help with interpretation of utility values. We valued your previous input about age of eligibility and considered this during finalization of our protocol. We kept 35-40 year olds (and 80+) as eligible for the review since we were not wishing to impose any limits on what could be practiced elsewhere and because we wanted to include studies looking at attitudes/intentions about starting to screen in ones 40s for which there is an interest in evaluating. For the studies among people older than 70 we differentiated between those 70-

		71, 70-79 and 75%+ to make sure any differences were examined.
Margo Wilson, Society of Rural Physicians of Canada	From email reply: Overall I really enjoyed this review of such an important topic. My criticism of the paper is that it seemed to extend beyond the primary objective of the paper by including patient preferences regarding cancer treatment. While this is an extremely important topic, in my mind it was beyond the scope of this paper and would be better treated in a separate discussion. Yes. 1) The objectives of the paper were clearly stated, however, the paper went beyond the stated objectives (examination of HSUVs of various treatment modalities in patients being treated for cancer). 2) The overview of the methods was not clear to the lay-reader, although the full document delineated some of the measures more clearly. 3) Inclusion of search strategies (specific terms) in the main body of the paper would be useful. 4) I'm interested in why no qualitative data was included especially as this study examines patient preferences.	Thanks for reviewing. When assessing preference/valuation of potential outcomes (e.g. experience of early vs advanced stage of disease, need for chemotherapy, as a marker of more treatment morbidity) using utility values from patients is informative to provide an accurate indication on the impact on quality life (i.e. utility/value) of the health state/outcome (considered a preference in this context). We have elaborated on this in the manuscript. HSUVs from experiencing the different outcomes are a major component of patient preferences as per the GRADE guidance for guideline developers. We have revised the abstract to try to provide better understanding of the concepts. We will create an accessible lay summary of the review, which will be cited in the manuscript. We have added a few key concepts about our search into the methods section. This review did not include qualitative data because we were trying to quantify preferences about the outcomes. If we were examining reasons for decision making, barriers etc we would have done so.
Anna Chiarelli, Samantha Fienberg, Jonathan Isenberg, Bronwen McCurdy,	Yes. • This was a well done systematic review • Objectives and methods were clear	Thank you. No changes requested.

Ayesha Salleh, Erin Svara, Rebecca Truscot, Meghan Walker, Ontario Health		
Melissa Coté, Québec INESSS (Institut national d'excellence en santé et en services sociaux)	Yes. The objective is clearly defined. The methodology is really well written. I appreciated the specification of some of the terms used (all/almost all" = 90%, a "large majority" = 75%, and a "majority" = 50%) which may sometimes seem obvious but which testify to the transparency and rigor behind your evaluation.	Thank you. No changes requested.
Jennie Dale, Dense Breasts Canada	No see comments in 4.	Please see responses in 4
MJ DeCoteau, Rethink Breast Cancer	Yes	Thank you. No changes requested.
Ariana Del Bianco, Canadian Cancer Society	No. Objective is clearly stated. Introduction section would benefit from plainer language or clearer explanation of complex concepts and study measures, especially for patients this will impact and others referencing this work. For example, providing an example of how to interpret a utility measure within the context of breast cancer screening.	Thank you. We have provided explanation in the abstract and introduction on how to interpret a utility value. We will also create an accessible lay summary of the review and cite this in the manuscript.
Paula Gordon, Dense Breasts Canada	No. Comments: I have been asked to review a draft done by the Alberta Research Centre for Health Evidence, based on 82 studies of the relative importance placed by patients aged ≥ 35 years on the potential outcomes (benefits vs risks) of breast-cancer screening. In the letter requesting my input, it said that "the findings will be considered as one form of patient input when the task force is balancing the effect estimates on benefits and harms…" It does NOT state what other forms of input will be used. That's a significant omission, given the limitations of the studies that were included. It would have been helpful for reviewers to be told more about the other forms of patient input, including their conclusion(s).	Thank you for your review. The guideline will certainly have this information, but we appreciate the request and have added details into this manuscript's abstract and objective, e.g. "Other forms of patient engagement are used during the development of recommendations, key messages, and knowledge dissemination tools (e.g., members of task force working group, public advisory network) (cited TF manual)

		and can help inform this topic on preferences and other considerations related to acceptability, resource use, and feasibility that contribute to recommendations.
Carol McClure, PEI Cancer Registry	No. Comments: I struggled with the disutility measurement. I would like a bit more description of it and how it is calculated (as shown in the supplemental table). I would love to see an example of how it calculated and what the difference in magnitude means. Most disutility measures in the manuscript were <0.10. Should I be interpreting a disutility of 0.8 as a huge difference in preferences? Whereas a disutility of 0.01 is relatively small and not preferentially meaningful?	Thanks very much. We have added information into the abstract and introduction on the range of data (0-1) for utilities and what can be considered important to the Canadian public (about 0.04 or higher).
Leah Palmer, Saskatchewan Cancer Agency	Yes	Thank you. No changes requested.
Sandy Sehdev, Canadian Breast Cancer Network	Yes. Very clear for a professional audience, less so for the general public	Thank you. We have added some clarification to help interpret some results, such as utilities, and will create an accessible lay summary.
Cheryl White, Dense Breasts Canada	No. I did not find this document to be focused on a single concise question. We are interested in attitudes and preferences of women, however some of the included studies did not seem to mention what women were told with respect to risks and benefits prior to responding to the surveys. Perhaps in some cases women were not informed of their risk of missed diagnosis or that the risk of overdiagnosis is mostly described as anxiety rather than a physical harm such as, for example, burns or bruises. In many studies an upper limit of number of secondary screenings required to save one life was not identified. Did the study authors consider that such a limit might be considerably higher than the magnitudes discussed in these studies? Is there a single study that states how many false positive screens would be too many in exchange for a life saved? It is possible the number would be quite high. Finally, this seems like a bit of circular argument. If we ask women if they would like to be screened and then choose to recommend or not recommend screening based on those opinions then what would be the purpose of recommendations?	Thank you for your review and comments. We have described the main data sources used to answer our question, one of which is the relative importance of the potential outcomes *as inferred* from attitudes/intentions (which can incorporate other considerations such as beliefs, cultural expectations). For these studies, we only included studies that provided participants with an estimate of the magnitude of effects for at least one benefit and one harm. We have clarified that these considerations, and our synthesis, was with respect to the outcomes rated by the working group as important or critical. Our study characteristics tables describe all of the data provided to women and we used this information

		(and the associated limitations about lacking data on outcomes such as overdiagnosis) during our synthesis. We used clinical input from the working group to decipher what an adequate description of overdiagnosis should look like, for example with its inclusion of cancer and not just “precancers”. Any description of how women would feel about overdiagnosis was not necessary for eligibility, and may have been interpreted as potentially biased as there are many potential consequences of being overdiagnosed (e.g., investigations, all treatments, labelling, stigma, psychological impact of being told one has breast cancer, financial). We agree that more data on elicited trade-offs between additional imaging and breast cancer mortality would have be useful. As described in the results, one study provided a very large range of numbers to select from (up to 10,000), but was otherwise considered at high risk of bias. Our low certainty about the findings for this comparison reflects this. We added a statement to indicate that an upper limit of the highest acceptable number of FPs (for preventing one life saved) was not evaluated. Other studies of women in their 40s suggest that the number may be variable and (for some) not that high; findings suggested that a majority (>50%) of women may decide against screening at about 5-600 recalls vs deaths averted (study information: 0.5 breast-

		cancer deaths prevented in 1000, 239-330 FPs and 2 to 10 overdiagnoses per 1000). This review is seeking to gain better appreciation of what magnitudes of effects are acceptable to women, as well as how much variability between people there may be. This could help with deciding about both the direction of recommendations as well as whether or to what degree informed decision making is recommended. Using a systematic review to understand this from an informed perspective is thought to be highly valuable.
Jamie-Lynne Bell, Department of Health and Wellness, PEI	Yes	Thank you. No changes requested.
Trina Buick, Canadian Association of Nurses in Oncology	Yes. Comprehensive overview of methods and is well structured. Provides a detailed explanation of the complex eligibility criteria and decisions made	Thank you. No changes requested.
Rene Wittmer (peer reviewer), Médecin de famille, Universite de Montreal	Yes. Very clear, nothing to add.	Thank you. No changes requested.
Question 2 Were the results clearly stated?		
Howard	Yes. I suggest highlighting in the abstract, when discussing overdiagnosis, that only 33% of women correctly identified/ defined overdiagnosis on acknowledge test. It's an important caveat and limitation to that evidence.	Thank you. We added a comment to this effect in the abstract.
Ibezi	Yes.	Thank you. No changes requested.

Boyd	Yes. Results were clearly stated/summarized with clear identification of exclusions and why, including how that process took place and who was involved.	Thank you. No changes requested.
YongHing	Yes	Thank you. No changes requested.
Carol	Yes	Thank you. No changes requested.
Brunet	No. ABSTRACT Without further explanation in the Method abstract section, it's not clear what disutilities numbers means in the FINDINGS abstract section for people not aware with the concept of health-state utility values. RESULTS (MAIN SECTION) Literature flow (p18) I don't understand if this document is an update why 28/82 included studies were included in the previous review? (and one of the previously included studies was excluded in this review…) More clarity is need. Table 2: add BCS to the acronym list under the table p23	Thank you for your comments. We have elaborated in the abstract and introduction how to interpret utility values. When we update reviews we include all studies previously reviewed that meet our eligibility criteria, to ensure our synthesis captures all relevant data for our research question. We have made this more explicit, and explained that the one excluded study was because of revised eligibility about age of participants in this update. Thank you for pointing out our omission of describing this abbreviation in our footnotes; we have corrected this.
Gallagher	Yes	Thank you. No changes requested.
TundeByass	Yes	Thank you. No changes requested.
Bryant	Yes. The results are exceptionally clearly stated. The addition of the column "What does the evidence say?" is a brilliant addition to the results, as it allows the reader to understand the outcome for each element that will be considered in the final interpretation of how women view the risks and benefits of screening.	Thank you. No changes requested.
Payne	Not selected: Had there been an elaboration of the term utility (and how to interpret it in the methods statement), then the results statement as is would be fine. Right now, something is missing. As I quickly scanned the tables, I am a bit concerned that in the second table of supplementary material, there are a couple of references to samples of individuals who were attending genetic testing centres. This is concerning given that individuals undergoing genetic screening may be at significantly higher breast cancer risk than an average risk population, and these people may consider information re breast cancer very differently – I don't think they should be included.	Thank you. We have elaborated in the abstract and introduction how to interpret utility values. One study of participants attending a genetic testing centre (n 33 of >8000 in the analysis) met our eligibility criteria, because it was determined that fewer than 20% were considered at high risk for breast cancer. Studies with people at moderately increased risk were eligible.

Wilson	No. See overall commentary below. 1) The authors noted that patient recruitment included cancer centers and outpatient centers – I would be interested if this excluded a portion of the populations, especially more rural residents. Are there any studies that address the preferences of rural patients as screening for them may involve the additional considerations of travel and extra cost. 2) Table 3 to me really summarized the initially identified objectives of the review. 3) It was not clear to me how disutility on treatment options affected one's opinions on screening – perhaps an area for further discussion? This seemed outside of the intended scope of this paper. 4) Table 2 initially looks at screening and diagnostics for false positive screens, but then proceeds to discuss treatment. I'm not sure how this relates to the initial stated objectives or if the scope needs to be broadened to reflect the data included. There is also some discussion of patients who were not detected by screening, which is again not entirely related to patient preferences on screening. 5) Table 4 lays out the evidence, but it seems to contradict the discussion outlined in the body of the text. Is there a way to visually or thematically display the weight of the evidence more effectively?	Thank you for your review and comments. 1. Our protocol did not specify geographical residence as a major potential confounder for our review which focus on the importance of outcomes to individuals rather than possible inequities in screening effects from accessibility. 2. When assessing preference/valuation of potential outcomes (e.g. experience of early vs advanced stage of disease, need for chemotherapy, as a marker of more treatment morbidity) using utility values from patients is informative to provide an accurate indication on the impact on quality life (ie disutility) of the health state/outcome (considered a preference in this context). We have elaborated on this in the manuscript. We have revised the abstract to try to provide better understanding of the concepts. We agree though that the data in Table 3 is very relevant and perhaps easiest to use for decision making. 3. Table 2 examines the value to people of all of the outcomes examined by the task force for this update. The task force was interested in the value to patients of reducing late stage disease or being able to avoid treatments, such as chemotherapy, that could cause more serious side effects. 4. The summary statements include in table 4 are our main

		conclusions, supported by the narrative and the tables in the supplement 3. It is our assessment that the summary statements capture what was reported across the relevant studies quite well (that is, are not contradictory to what the studies reported). Because we had no other comments about the information in this table we did not make any changes.
Chiarelli Fienberg, Isenberg, McCurdy, Salleh, Svara, Truscot, Walker	Yes. • Results are clearly stated in the tables, however may be difficult to understand for individuals without this expertise/not working in this area. • Suggest providing an example of how to interpret utility/disutility findings, to ensure the audience understands the results (e.g, Table 2. Summary of findings on health state utilities).	Thank you. We are making an accessible lay summary to cite in the manuscript. We have also added in several places more information on how to interpret utility values.
Coté	Yes. I appreciate the tabular presentation of the results, which gives a quick look at the overall picture. However, I would have liked to have the references written with the author name and the year of publication (table 2), as in appendix 2, to facilitate identification of the studies selected. I understand, however, that this would significantly lengthen the table.	Thank you for the feedback We appreciate that tables cannot include all information that may be useful to all readers, but need to ensure the tables are concise and follow journal formatting style. We hope that the supplemental files assist those that require more information.
Dale	No see comments in 4.	Please see responses in 4.
DeCoteau	No. While the objectives and methods of the review were clear, particularly to lay people, the findings should have a plain language version to facilitate clear communication and understanding among the general population. If the intent of this evidence review is to inform public-facing guidelines, then the evidence review should be accessible and framed in plain language that can be understood by the public.	Thank you we are preparing a lay summary to cite in the manuscript.
DelBianco	Yes	Thank you. No changes requested.
Gordon	No. Comments: The stated purpose of this systematic review as to examine the relative importance placed by patients aged ≥ 35 years on the potential outcomes of breast-cancer screening. The conclusion stated that “how patients value the potential outcomes from breast-cancer screening will be useful during decision-making for recommendations.”	Thank you. We have revisited our conclusion statements and made some minor edits. The data across all datasets (e.g. important disutility of a diagnosis of cancer and a FP, majority in 40s possibly not accepting

	Based on the results of most of the included studies, a large majority of women prioritize early detection of cancer through screening, are not concerned about recalls, and are willing to accept a large percentage of overdiagnosis. THAT should have been the conclusion.	screening if offering low net benefit) does not strongly support your proposed conclusions. We are confident we have described all results accurately and reached sound conclusions with careful consideration of the study limitations and how we used the findings of each study to answer our research questions.
Isenberg	Yes. • Results are clearly stated in the tables, however may be difficult to understand for individuals without this expertise/not working in this area. • Suggest providing an example of how to interpret utility/disutility findings, to ensure the audience understands the results (e.g, Table 2. Summary of findings on health state utilities).	Thank you we have added clarity about interpreting utility values.
McClure	Yes.	Thank you. No changes requested.
Palmer	Yes	Thank you. No changes requested.
Sehdev	Yes. Quality of evidence, risk of bias and magnitude of conclusions were clearly indicated.	Thank you. No changes requested.
White	No. This document does not do a good job of summarizing topics and including the key points in the conclusion. A summary should be added.	Thank you. We have revised our conclusions to be more comprehensive and have prepared a lay summary.
Bell	Yes	Thank you. No changes requested.
Buick	Yes. Presentation of tables was valuable.	Thank you. No changes requested.
Wittmer (peer reviewer)	Yes. Very clear, nothing to add.	Thank you. No changes requested.
Question 3 Are the conclusions in the review supported by the data that were reviewed?		
Howard	Yes. I agree with all the points you make in the discussion section. I was surprised by the low disutilities for several treatments. I found it interesting that given the same low net-benefit scenario, women in their 40s tended to decide against screening, while women in their 50s prefer screening. Your statement about “belief perseverance” may be relevant here. I think there may be an effect due societal norms or preexisting bias or expectation on the part of women in their 50s, knowing they should be screened, or perhaps their personal experience with friends who’ve been diagnosed with breast cancer, leading them to discount the benefits/ harms data that is presented. Patient preferences are very important in shared decision making for breast cancer screening, but it’s important to understand that these preferences can be affected by external factors including context and prior patient expectations.	Thank you. No changes requested.
Ibezi	Yes	Thank you. No changes requested.
Boyd	Yes	Thank you. No changes requested.

YongHing	Yes	Thank you. No changes requested.
Carol	Yes	Thank you. No changes requested.
Brunet	Yes. ABSTRACT One sentence could be added to Abstract Conclusions, after the first sentence in other to be more precis? ex: the evidence strongly suggests that the outcomes examined have importance to women of any age.	Thank you we agree and have revised.
Gallagher	Yes	Thank you. No changes requested.
TundeByass	Yes	Thank you. No changes requested.
Bryant	Yes. The conclusions relate well back to the huge volume of data that is presented, and there is a good explanation of the interpretation. It is especially helpful that the authors are quite transparent about the areas about which they are uncertain, and to some extent, to the degree of that uncertainty.	Thank you. No changes requested.
Payne	Yes. The results section (which is beautifully laid out in the accompanying tables in excruciating detail which is appreciated) is a bit overwhelming to review, but as best as I can tell, yes.	Thank you. No changes requested.
Wilson	Yes.	Thank you. No changes requested.
Chiarelli, Fienberg, Isenberg, McCurdy, Salleh, Svara, Truscot, Walker	Not selected. • Suggest adding implications of these results to screening programs in the conclusion	Thank you for the suggestion. We have revised to add “that provision of information on the likelihood of the outcomes may be necessary to enable informed decision making”.
Coté	Yes. The conclusions are relevant to the results of the literature review. Your synthesis of the results is very representative. I appreciate the fact that you have tried to explain certain unexpected results and that you have formulated hypotheses.	Thank you. No changes requested.
Dale	No see comments in 4.	Please see responses in 4
DeCoteau	Yes. The conclusions in the review are reflective of the data; however, as a patient group, our concern is two-fold: - There is a need to have lay-person friendly communications to share the conclusion and an overview of the data. - Does the data reflect and address considerations around a risk-based screening model that considers factors other than age?	Thank you. We are preparing an accessible lay summary. We have added a note in the limitations that there was some but limited data to indicate findings may be similar across risk groups within age groups. We did not access views on different screening approaches based on risk, though the review on evidence of screening effectiveness will do this and other information will be examined about this by the task force.

DelBianco	Yes	Thank you. No changes requested.
Gordon	No. Comments: There were no conclusions clearly stated. Moreover, the majority of the studies were rated "Low" by GRADE. So even though a large majority of women prioritize early detection of cancer through screening, are not concerned about recalls, and are willing to accept a large percentage of overdiagnosis, the risk is that the Task Force will ignore women's preferences.	Kindly also see our response to your comment in question 2. The task force finds patients preferences (and the level of certainty about the findings) quite important when deciding on the strength and direction of their recommendations.
McClure	Yes. Comments: Yes, but this goes back to my lack of understanding of the magnitude of the disutility measurement. What magnitude would make me want to absorb patient preference into my BCS guidelines?	Thank you. Hopefully our addition of information to help interpret the utility data is helpful for you and others.
Palmer	Yes	Thank you. No changes requested.
Sehdev	Yes.	Thank you. No changes requested.
White	No. Again, the conclusion is poorly written.	Thank you. We have revised our conclusions to be more comprehensive and have prepared a lay summary
Bell	Yes	Thank you. No changes requested.
Buick	Yes. Limitations were succinctly addressed. The challenges to the definition/conceptualization and variability of FP (ie viewed as a benefit and not a harm) could be expanded in the discussion (in terms of bias/women's understanding). While qualitative evidence wasn't examined – it could have supported and added additional insight to address some of the limitations raised. Conclusions supported that the outcomes examined have importance to women of any age and that personal choice is vital to enable informed decision-making.	Thank you. We have reviewed our discussion of the limitations and think we've captured the first two points sufficiently for the purposes of this review. We have expanded on our conclusions and incorporated your suggestions.
Wittmer (peer reviewer)	Yes. The results are nuanced, and the terminology aligns well with the data that were reviewed. The entire process is very transparent and easy to understand.	Thank you. No changes requested.
Question 4 Do you have any additional comments?		
Howard	No response	No changes requested.
Ibezi	I note the paucity of race-related data in the studies.	We agree and have added this as a limitation in the abstract and manuscript conclusions.
Boyd	Extensive research provided with clear objective and intentions.	Thank you. No changes requested.
YongHing	Task force recommendations should not be based on low certainty evidence.	Thank you. No changes requested.
Carol	It was an interesting read. I do not have anything substantial to contribute, but I'm providing my completed checklist. The only consideration I had was for future research – namely, how some populations' growing mistrust of health care, coupled with limited relationships with	Thank you for the additional comments. Your comments help support that the evidence from decision aids and other informational

	healthcare providers due to lack of access, impact the patient's ability to objectively make a decision on the benefits/harms of an intervention. Individuals' personal trust in healthcare was out of scope for this research, however not everyone makes objective decisions solely based on advice from their HCP, particularly if they don't trust them. It's hard to sever personal feelings about a health care intervention, so we should consider this when researching informed decisions and consent and keep it in mind for BC screening guidelines. I didn't include this in my comments because, as I said, it was stated as out of scope	materials asking women to decide whether or not to screen (ie data in Table 4) is indirect for the purposes of this review which tries to emulate rational decision-making (based on the effects on outcomes) which is only one aspect of many people's decision making. We have decided to not add this as a major research objective since it was not the focus of our review (we did not examine all the evidence on factors influencing screening decisions) but we think this aspect has been described in a sufficient manner in our text and Table 1.
Brunet	DRAFT-KQ3 Supplementary file 2: Table S2.7. Summary of Risk of Bias, The reference of the tool use to perform this assessment should be indicate in the bottom of the table? DRAFT-KQ3 Supplementary file 3: Tables S3.7. Summary of Rias of Bias Across Studies Same comment + correction	Thank you we have added reference to the tool in these supplements.
Gallagher	No.	Thank you. No changes requested.
TundeByass	It is clear for the most part that informed women value the potentials of breast screening above potential risks. Therefore, an informed decision making process based on personal circumstances and preferences is key in this review and subsequent recommendations.	Thank you. No changes requested.
Bryant	It is refreshing to see such a clear and quantitative approach to a complex issue. It will undoubtedly be a helpful lens through which to look once the evidence on effectiveness of screening itself is collated.	Thank you. No changes requested.
Payne	Blank	No changes requested.
Wilson	Additional comments include: 1) Is there any information given to women about the harms of biopsy (bleeding infection death versus the harms of a cancer found at a later stage?) and how this affects patient decision making 2) The paper notes: "There is uncertainty about whether these trade- offs would be acceptable in situations where the outcome was well understood and in	Thank you for the additional comments. 1. No, none of the studies examined this. 2. This review focused on informed preferences, as much as possible, so when data from knowledge tests (as described in the text and

	view of this we rated down for indirectness. The findings appear to apply across 40 to 70-year-olds." The authors discuss "rating down" for indirectness, but this is a little vague. To me this is the crux of the paper – what are patient's preferences for screening and acceptance of interventions to rule out false positives, versus treatment preferences for identified cases - and it should be explored further. Also I wonder why there was concern about the outcome not being well-understood – was there discussion or evidence of this in the literature? If so, could it be discussed further? 3) Was there any discussion about differences of values in younger populations? I would presume that preventing a cancer related death or having an early diagnosis may carry more weight in your forties than in your sixties. 4) There were some interesting studies that involved that use of a decision aid – was there any evidence of whether patients preferred the use of a standardized decision aid? Or what kinds of decision aids are the most useful (those delivered in person versus a self-administered one online?) 5) Pg 46 - The discussion addresses that people overestimate the benefits of screening, but in previous discussion and analysis patients are willing to tolerate a high level of false positives in order to identify a cancer diagnosis. I think the discussion around this could be strengthened by inclusion of the latter.	reported in detail in the supplementary files) indicated that outcomes such as overdiagnosis were not very well understood in some studies this reduced our certainty since the exposure of interest (information on the benefits and harms) was not well understood. We added a description to the results section to help justify why this was viewed as indirectness. 3. Much of our analysis focused on findings by age and our judgements about the net benefit of screening presented in the studies took into account the larger beneficial effects in those (older) people with higher baseline rates of cancer. 4. Our review did not examine the effectiveness (e.g. use for decision making) or preferences about decision aids. 5. Our discussion about the overestimation of benefits from screening was to justify why we only used studies that provided participants with descriptions and magnitudes of effects, and for why we rated studies at higher risk of bias if they only provided relative effects. The findings of a relatively large number of acceptable false positives is a valid finding of this review and considered a different topic.
Chiarelli, Fienberg, Isenberg,	• Many people using this systematic review may not have expertise in this specific area. Additional information on how to use the results and interpret these findings would be beneficial.	Thank you, we have added some clarification to help interpret the data

McCurdy, Salleh, Svara, Truscot, Walker		and are preparing an accessible lay summary.
Coté	General comments: The document is very well written. Your assessment is very thorough. Congratulations on your excellent work. Although I am not familiar with HSUV-type evaluations, the methodology allowed me to fully understand and interpret the results. I really appreciated that you added this type of study to your evaluation compared to your 2018 publication. Breast cancer screening: The selected studies come from all over the world: Netherlands, USA, South Korea, Finland, the United Kingdom, Australia, Japan, Spain, Norway, Italy, Croatia, Thailand, France, Greece, England, Wales and Germany. Considering that screening is not carried out in the same way everywhere, certain particularities could have an impact on women's perspectives. Here are just a few that came to my mind: - Biennial vs. annual screening; - Independent double reading vs. single reading (impact on response time and recall rate); - Costs of screening incurred by the healthcare system vs. women screened; - Accessibility of screening. Without adding sub-analyses to your evaluation, I believe that mentioning these particularities would make it possible to be transparent about the risk of bias in the transferability of results to our Canadian context. Communication: Although you mentioned it, it would have been relevant to know women's main sources of information about breast cancer screening, in order to better reach the target audience with the new recommendations.	Thank you for your additional comments. While it is likely that studies on the effects of screening (actual recall rates, mortality reductions) may be influenced by many of the factors you mention, we think it is less relevant to the data we examined which is focused on responses to the relative magnitudes of the outcomes from screening as presented to women for which studies were quite consistent because of using reported effects based on systematic reviews. We did take into account during our analysis the potential bias/limited applicability from studies conducted in countries where screening is not routine (e.g. Hong Kong). Our review did not include data on or examine information needs or sources of information for women. Hopefully the dissemination tools created with this task force recommendation will reach the target audience.
Dale	As a patient advocate and cofounder of an organization advocating for optimal breast cancer screening, I appreciate the chance to review the Patient Preferences Review Update. In 2018, I was a reviewer of the overall draft guideline. Once again, it is clear from the evidence review that women place higher weight on the benefits as compared to the harms of breast cancer screening. It is clear that they are willing to experience “false positives and “overdiagnosis” to reduce the chance of dying from breast cancer. And yet, similar to 2018, it is inferred that women did not	Thank you for these additional comments. We did not intend to be patronizing to any extent and many of the studies we reviewed that examined/tested knowledge made clear statements to this effect, that is, the difficulty to

	understand what they were being asked. The conclusion is patronizing. As well, the review itself is seen to have a number of issues. 1. The terminology used to assess values and preferences is problematic. The term false positives is pejorative. The review mentions that the Task Force is recommending the use of another term, but false positives is still the term used in the studies. As well, the term overdiagnosis is misleading. Screening does not result in overdiagnosis- only pathology does. The Task Force has continued to overestimate the rate of overdiagnosis due to its use of the compromised CNBSS. 2. In addition to the terminology issue, the time framing is an issue. The 10-year window used to assess benefit is too short a time frame to be used, particularly for a woman in her 40s or 50s. This short time frame may have resulted in some benefits of screening as being presented as very low. 3. Most of the studies used in the review were small. The confidence in the was rated as low. We have seen in previous reviews that some evidence is rated very low. It is not clear what mechanisms are in place to ensure different evidence reviews rate the evidence quality in a consistent manner. 4. The continued focus on All Cause Mortality (ACM) in the review is misleading. ACM is not a useful measure of the impact of breast cancer screening, because deaths from breast cancer represent a relatively small fraction of ACM. Attempts to measure ACM will be impaired by lack of statistical power as well as the challenges of follow-up. 5. The evidence surrounding disutility is questionable. The amount of time disutility is experienced also needs to be taken into account. Overall, the studies showed that most women are accepting of overdiagnosis and call backs. They are willing to trade a decrease in mortality against an increase in so called harms. These are the same findings which were ignored in 2018. It is hoped that in 2024 women's preferences will not continue to be dismissed.	understand the concept of overdiagnosis. 1. We have indicated that false positive is not the best term (and the recommendations will avoid this term) but decided to keep this term in our review since the research widely uses this term. We also carefully examined how studies described this outcome and have added a discussion about this in the manuscript. By definition, in this context, overdiagnosis can only happen within the screened people. 2. We have reported the timeframe used by the studies and interpreted the data in light of this. We are not judging whether the effects are accurate/valid in the studies but rather differentiating between results based on these variations. 3. We have carefully followed internationally accepted guidance when assessing the certainty of the findings, and have justified our rationale for each rating. Though not all reviewers would come to the exact same conclusions, as per GRADE guidance we have been transparent with our assessments and provide all the data to support these. 4. We realize effects on all-cause mortality are hard to establish but this does not mean that it is not important to consider. We did not rate down (e.g. for risk of bias) studies that did not provide women with estimates of the effects on all-cause mortality because we realize

		how hard it is to present data while also describing the uncertainty around it. 5. We think we have made it quite clear that duration of experience with an outcome will impact its disutility, and have focused a good portion of our discussion on this. In 2018, the task force used findings about patient preferences, especially around the consistent evidence that there will be some variability in values among patients, to a large extent when developing their recommendations. The findings of both reviews undertaken were clear that not all women will think the benefits outweigh the harms.
DeCoteau	The discussion around breast cancer screening continues to focus primarily on age as the risk factor to determine when screening should commence. Considerations should be given to the mounting data that supports a more sophisticated risk-stratification approach to breast screening.	Thank you for the comments. We added a comment in the discussion about the limited data on whether findings would differ based on risk. The review on effectiveness and other information will be used to inform the task force deliberations about differing risk.
DelBianco	An additional limitation may be the lack of HSUV studies from Canada.	We agree and added this.
Gordon	NOT ENOUGH ROOM, SO REPEATED BELOW, OUTSIDE THIS TABLE 1. In spite of the fact that the vast majority of the studies concluded that women put a higher value of the benefits than the harms of screening, all but a few rated “low” or “very low” by GRADE. This opens the door for the Task Force to ignore women’s true preferences. 2. Most RCTs used only breast cancer mortality as a benefit of screening, ignoring the significant benefit of the opportunity to have successful treatment with less aggressive surgery (lumpectomy vs mastectomy, sentinel node biopsy vs axillary dissection) and the opportunity to avoid chemotherapy. Had those additional	Thank you taking the time to make these additional comments which we appreciate. We have numbered your comments into broad categories and have responded to each. 1. During our GRADE assessments we carefully interpreted the findings and are confident in our assessments. As per the study findings and our summary statements, it is not clear (and

	benefits been included, it would undoubtedly have increased women's willingness to accept higher frequency of recalls ("false positives") and overdiagnosis. The studies that did include stage distribution (reduced advanced disease) and treatment burden (reduced mastectomy) vs FP biopsies were rated low on GRADE. The Valentine study 2022 (their ref 14), is emblematic of the bias in many of the studies. It was a concerted effort to find ways to persuade patients to decline screening. They seem unable to accept that some choose to be screened, even when there is no data showing mortality reduction. They speak about improving "the quality of patient decisions," but it's clear that they define improvement as choosing not to be screened. They describe their discussion of the potential harms of overtreatment, but not the harms of undertreatment. They included patients aged 40-70, but don't indicate whether they tailored the information provided based on age, or whether they told participants about the negligible incidence of overdetection in younger people. In their experiment, they told patients that "the intended purpose of cancer screening, which is to detect cancer early with the goal of preventing cancer death," and that mammography saved 1 life per 1000 people screened, and "harms consisted of the rates of false positives and overdetection per 1000 people screened over a decade for women." They didn't state where they sourced the numbers provided to the participants. All the examples they described used mortality reduction as the only benefit. As one example, they referred to a study that claimed that "Overuse of screening tests (e.g., screening when there is no evidence that the test will benefit patient outcomes) have been reported for… breast cancer.17 It looked at mammography rates in women with severe cognitive impairment, noting that "certain subgroups with cognitive impairment are often screened despite lack of probable benefit." They concluded that guidelines should explicitly recommend against screening these women." In this group, the important benefit of early detection that they ignored was the opportunity to do minimal therapy to improve quality of life. Therapy in these patients can be limited to excision under local anesthetic, when cancer is detected early. If cancers are not detected until clinically apparent, it's often because they have eroded through the skin, and become a nursing challenge: they need constant dressing changes because they ooze; they have unpleasant odor, and are vulnerable to infection.	definitely not with high confidence) that across all ages and scenarios presented that "a vast majority" value the benefits as greater than the harms. In several cases the evidence suggests there will be at least some variability among individuals, even when the net benefit is portrayed as quite high. Further, as you rightly point out there are limitations across studies, such as what an how data was presented to women (all detailed in our tables) and how the potential outcomes were described conceptually and in respect of the effects from screening. Our certainty is reflective of our confidence in our summary statements, for example in how confident we were about the proportions of women that would have the same views. This does not suggest that the task force will ignore the results or consider preferences as less important. Indeed, the task force (as with any guideline panel) has to reflect on much evidence that is of low certainty and does not use this in isolation to support recommendations for or against screening. We have clarified in this manuscript that the findings of the review are one source of information on patient preferences. 2. We agree that the findings (especially for the indirect evidence used for findings in Table 4) relate to what information is provided in

	Mehta KM, Fung KZ, Kistler CE, Chang A, Walter LC. Impact of cognitive impairment on screening mammography use in older US women. Am J Public Health. 2010;100(10):1917–23. Finally, in “Intervention 4: Narrative,” they stated: Research has found that the most impactful narratives on screening intentions are those describing physical harms from screening, rather than emotional harms or overdetection.37 Since the primary goal of the narrative intervention was to decrease interest in the unbeneficial test, we wanted to present a narrative that described a clearly negative (albeit uncommon) experience while also remaining realistic. Thus, we crafted our narrative to maximally decrease interest for the unbeneficial test by including information about experiencing physical harm, as well as emotional and financial harms, in the context of a false-positive screening test result. The full narrative can be viewed in the supplemental materials. They did not do the same for the harms of underdiagnosis, ie the avoidable pain and suffering when cancer is not detected as early as possible. Valentine KD, Wegier P, Shaffer VA, Scherer LD. The Impact of 4 Risk Communication Interventions on Cancer Screening Preferences and Knowledge. Medical Decision Making. 2022;42(3):387-97. - The same applies to any research that passes judgement as to whether screening is justified, and any decision aid that uses that research: ALL the benefits must be included, and by limiting the review to RCTs, that shows how this whole review was biased. - The outcome of studies that used “decision aids” very much depends on the information given in those decision aids. For example: was breast cancer mortality reduction based on randomized trials (RCTs) or observational studies. Lay women may not realize that data from RCTs underestimates mortality reduction because of contamination and non-compliance. An individual woman in the process of deciding whether she should participate in screening would be better served by observational data (ie potentially 50% breast cancer mortality reduction) rather than population-based, intention-to-treat data (15-20% breast cancer mortality reduction). In particular, since this review is for Canadian women, the decision aids should ideally be based on Canadian data, including the Coldman 2014 “Pan-Canadian study,” and information on recall rates from the Canadian Partnership Against Cancer.	the decision aids or other educational/information materials. We agree that there are limitations from studies not reporting on all important outcomes or when descriptions of the outcomes are potentially biased. We had commented on the fact that few of the studies providing information about the effects of screening presented information on the possibility of lower treatment morbidity and avoiding some treatment such as chemotherapy. It is difficult to judge these studies too harsh for bias when there is limited high quality research (from comparing screened versus not screened populations rather than case series of patients with cancer) supporting these effects. We have added a comment on this limitation to the discussion but do not feel it changes our overall findings. We have interpreted the findings based on assessments of these limitations. For example, poor descriptions of overdiagnosis or only using relative effects for benefits on breast-cancer mortality deemed ratings of high risk of bias for studies, and the magnitudes of effects across the outcomes has considered when we made judgements about the net benefit of the data presented. One example is that for the Valentine study as you mention, the data from “intervention 4” (viewed as biased by only focusing on rare harms from a biopsy) was not relied upon at all

	3. Comments on specific Sections All-cause versus BC Mortality - Deaths from breast cancer represent a small percentage of all deaths. So even trials that show significant reduction in breast cancer mortality are unlikely to show reduction in all-cause mortality. - so outcomes from any decision aid that relies on all-cause mortality should be excluded, since it is mathematically almost impossible for all-cause mortality to be positively affected by screening, and it is misleading to tell women that screening doesn't impact all-cause mortality without a full explanation of the math - Tabar et al demonstrated reduction in all-cause mortality among the women diagnosed with breast cancer. They used data from the Swedish Two-County Trial of mammographic screening for breast cancer, in which 77 080 women were randomised to an invitation to screening and 55 985 to no invitation. There was a significant 31% reduction in breast cancer mortality in the invited group (RR 0.69, 95% confidence interval (CI) 0.58-0.80; p<0.001), and a significant 19% reduction in deaths from all causes was observed among breast cancer cases in the group invited to screening (RR 0.81, 95% CI 0.72-0.90; p<0.001). Tabar L, Duffy SW, Yen MF, Warwick J, Vitak B, Chen HH, Smith RA. All-cause mortality among breast cancer patients in a screening trial: support for breast cancer mortality as an end point. J Med Screen. 2002;9(4):159-62. doi: 10.1136/jms.9.4.159. PMID: 12518005. BC Mortality versus Overdiagnosis - conclusion that across all ages, at least a majority (>50%) and possibly a large majority (>75%) of patients probably accept up to 6 cases of overdiagnoses to save one death from breast cancer, and the rating by GRADE is moderate. Clearly, women put greater importance on early detection, than the small real risk of overdiagnosis (Puliti). Given that the Task Force overestimates the incidence of overdiagnosis based on the flawed CNBSS trials, it's not clear how they will use this information. Puliti D, Duffy SW, Miccinesi G, de Koning H, Lynge E, Zappa M, Paci E; EUROSCREEN Working Group. Overdiagnosis in mammographic screening for breast cancer in Europe: a literature review. J Med Screen. 2012;19 Suppl 1:42-56. doi: 10.1258/jms.2012.012082. PMID: 22972810. BC Mortality versus False Positives	when making our summary statements/conclusions. Of note, our assessments focused on the outcomes rated by the task force's working group to be important/critical for decision making, and undertreatment from missing a cancer was included whereas all-cause mortality (even if effects are uncertain) was. Our review did not try to evaluate whether the effects presented were accurate/valid, but rather what the views were in light of the differences in the data presented. Several studies using data from observational or modelling studies indicated higher net benefit from screening and the results were compared with other studies using trial data (e.g. showing a lower net benefit) for their materials, regardless of whether we felt one set of data was more valid than the other. This is the main basis for synthesizing these findings. As indicated in our review (and GRADE guidance), the findings on the disutilities of the outcomes and from direct tradeoffs between outcomes are considered more relevant to this research questions and was weighed accordingly (for example by us rating down all evidence on attitudes/intentions from decision aids/educational materials as indirect because intentions to screen are based on a large number of factors apart from the numbers/outcomes).

	- women 40-49 prioritize reduction in mortality, over false positives. This is less so in women 50-59, but all studies are rated as low, so this gives the TF the opportunity to ignore women's preferences. BC Mortality versus FP biopsies - For patients 40 or older, a large majority of patients may accept that between 10-15 people experience a FP biopsy to prevent one BC death over many years, but the solitary study is rated as low, so this gives the TF the opportunity to ignore women's preferences. Stage Distribution (reduced advanced disease) versus FPs - For patients 40 or older, a large majority of patients may accept that at least 25 people experience a FP to prevent one advanced stage cancer, but the studies are rated as low, so this gives the TF the opportunity to ignore women's preferences. Stage distribution (reduced advanced disease) versus FP biopsies - For patients 40 or older, a large majority of patients may accept that at least 4 people experience a FP biopsy to prevent one advanced stage cancer, but the studies are rated as low, so this gives the TF the opportunity to ignore women's preferences. Treatment burden (reduced mastectomy) versus FP biopsies - For patients 40 or older, avoiding mastectomy may be much more important than experiencing a FP for a majority of patients, but the solitary study is rated as low, so this gives the TF the opportunity to ignore women's preferences. Table S3.2. Making Inferences from Attitudes, Intentions, And Behaviors, by age and judgement of net benefit presented - For women 40-49: as expected, attitudes on benefits:harms varied based on whether the net-benefit scenarios were presented as high, moderate or low - for 50 yr old women: a large majority of patients 50 years old probably weigh the benefits as greater than the harms from screening, whether presented with a high- or moderate net-benefit scenario. - for women 50-69, a large majority of 50 to 69-year-old patients probably weigh the benefits as greater than the harms from screening in both high-and low net-benefit scenarios. - for women 70+, a large majority of patients 70-71 years of age who have recently screened probably think the benefits outweigh the harms for continuing to screen in a moderate-to-low net benefit scenario, and this was based on one study rated as moderate on GRADE.	3. Though we agree that all-cause mortality (within a specific timeframe) may be difficult to demonstrate (mainly due to imprecision) this does not necessarily prevent it from being an outcome that is important to assess even if to mainly comment on this uncertainty. Recommendations for screening for any age/population group is a clear indication that this outcome is not the sole factor relied upon for the overall direction of recommendations. We are confident in our assessments of the findings for the data in table 3 on trade-offs though cannot comment on how the task force will judge the effects from their review on screening effectiveness or on how much findings from this review contribute to their decision making. We recognize that the few studies that mentioned all-cause mortality did not clearly qualify this with a statement about the uncertainty. Regardless, we do not think this seriously impacts the main findings, for example that "for patients 50 or older, at least a large majority (>75%) of patients may think that reducing breast-cancer mortality is beneficial even if there is no impact on all-cause mortality." We carefully examined the descriptions studies provided about the outcomes, and rated down when this was of serious concern, as it was for the Schwartz study. With working group input, our most serious concern for

	- in 3 studies of women 75-early 80s, who have recently screened, a majority but possibly not a large majority may weigh the benefits as greater than the harms for continuing to screen. It is unclear what impact life expectancy has on this preference, but these were rated low on GRADE. Table S3.3. Direct Preference Data from Mixed Ages The first 2 studies cited under "Weighing all-cause and BC mortality" Reder 2017 (Germany) and Davey 2005 (Australia) both included all cause mortality in their decision aid. - Deaths from breast cancer represent a small percentage of all deaths. So even trials that show significant reduction in breast cancer mortality are unlikely to show reduction in all-cause mortality. - so outcomes from any decision aid that relies on all-cause mortality should be excluded, since it is mathematically almost impossible for all-cause mortality to be positively affected by screening, and it is misleading to tell women that screening doesn't impact all-cause mortality without a full explanation of the math - Tabar et al demonstrated reduction in all-cause mortality among the women diagnosed with breast cancer. They used data from the Swedish Two-County Trial of mammographic screening for breast cancer, in which 77 080 women were randomised to an invitation to screening and 55 985 to no invitation. There was a significant 31% reduction in breast cancer mortality in the invited group (RR 0.69, 95% confidence interval (CI) 0.58-0.80; p<0.001), and a significant 19% reduction in deaths from all causes was observed among breast cancer cases in the group invited to screening (RR 0.81, 95% CI 0.72-0.90; p<0.001). Tabar L, Duffy SW, Yen MF, Warwick J, Vitak B, Chen HH, Smith RA. All-cause mortality among breast cancer patients in a screening trial: support for breast cancer mortality as an end point. J Med Screen. 2002;9(4):159-62. doi: 10.1136/jms.9.4.159. PMID: 12518005. Stiggelbout 2020 (The Netherlands and Australia) tested women's understanding of the terms "false positive" and "overdiagnoses," before asking what their tolerance was for overdiagnosis. But the information given to the women was incorrect: "Cancers like this may grow very slowly or just stay the same. Without screening, they would never be noticed or cause any trouble...doctors cannot be sure which cancers will be harmless. Therefore, treatment is recommended...". They provided "explicit information on the nature of invasive treatments… (e.g., in mastectomy scenario 37% of women who correctly picked the overdetection description would	descriptions of overdiagnosis was when the diagnosis was not of a cancer but only a precancerous lesion or DCIS. We have added discussion points about the poor descriptions of false positives across several studies (most seriously when stating they are all suspicious for cancer). Because we rated our certainty in the findings (often rating down for these issues) and used broad categorizations about the findings (e.g. majority vs large majority) we are quite confident in the conclusions. As mentioned above, having to use low certainty evidence does not suggest that the task force will ignore findings or the importance of patient preferences.

	always screen versus 52% of women who (incorrectly) picked the FP; 49% vs 58% for lumpectomy." The Schwartz 2000, USA is particularly troublesome. The wording in their Format & Definitions section is hugely misleading: in print survey questions; FP: "in a woman who gets mammogram annually for the next 10 years, one of her mammograms will look like she has BC even though she does not." In fact, most recalls are for findings that are not highly suspicious. Some reporting systems allow the reporting radiologists to indicate the level of suspicion. But most tell women that most recalls are false alarms, and only a small fraction of women who are recalled are diagnosed with cancer. Re: overdiagnosis, "We would like to ask your opinion about ductal carcinoma in situ or DCIS, a breast abnormality which can only be picked up by mammograms. Specialists are confused about DCIS because some-times it becomes invasive and sometimes it doesn't. If DCIS does not become invasive, it will not affect how long a person will live even without treatment. Doctors don't know which DCIS will become invasive. Nowadays, almost everyone with DCIS gets treated. Many people receive surgery, chemotherapy, or radiation who would never have gotten sick. For these people, treatment provides no physical benefit."; no trade -offs for overdiagnosis or estimates provided Under Relative importance for decision making, they state "Overdiagnosis: important for 60% (71% in 18-39 yr); not included in primary synthesis since no numerics provided or trade-offs elicited)" Why were women aged 18 included? Why were 25% of the women younger than 40? Overdiagnosis is vanishingly low in younger women. Was overdiagnosis explained just in the context of DCIS? Was it explained in terms of competing causes of death, like other cancers, heart disease, and other illnesses uncommon in younger women, but applicable to the elderly? In Lewis, 2003, USA, they framed the communication piece in a manner that would increase anxiety among the participants. Women were surveyed about their beliefs about the importance of mortality reduction vs false positives. Then they were told in videos, "an abnormal mammogram when there is nothing actually wrong, but the result may require more tests or a biopsy to find out that there was no cancer; more than one third of women with a false positive continue to worry about having breast cancer." There was no need to include the last phrase. Women were surveyed again after watching the videos and there was no change in their beliefs.	

McClure	No, thank you for allowing me to review this manuscript. Thank you for working on the updated guidelines.	Thank you. No changes requested.
Palmer	Blank	No changes requested.
Sehdev	Outstanding and comprehensive review	Thank you. No changes requested.
White	In 2019 I was diagnosed with locally advanced breast cancer with lymph node involvement. It was 2 weeks after my 43rd birthday, a few years earlier I asked my family doctor for a mammogram, and she refused citing the task force. I asked again but she said no. I understand this is an academic process, that for many involved in this process it is routine, but to me this is personal. I could have received an earlier diagnosis, I could have saved myself aggressive chemotherapy that has aged me and 29 painful fractions of radiation. The women who are included in these studies, they are not being told that this does happen in younger women. They don't know that mammograms are mostly just uncomfortable and not painful and they certainly don't know that a 'false positive' most often means returning for a quick and painless ultrasound or MRI to rule out cancer. Even if a needle biopsy is required, it is a quick outpatient procedure and not more painful than some dental visits. I speak to friends and family and share my story and many end up struggling to access this care at 40 because the task force recommendations are held up as gold standard evidence by family doctors who don't have time to read primary literature and who don't realize that saving lives is being measured against transient anxiety. Thank you for reading this statement and I hope you will include my words in the final version of the document.	Thank you very much for sharing your thoughts and experience. It is intended by the task force to make their recommendations more explicit/understandable as there was never the intent to prevent anyone in their 40s from being screened as long as they were well informed about the benefits and risks and chose to screen. With the updated recommendations the task force is preparing to undertake a fulsome evaluation through stakeholder and public comments and active usability testing with patients/public, to ensure the intent of their recommendations and related messaging is well understood. Many of the included studies gave data on the expected rates of breast cancer for women in their age group. We have commented on the oftentimes insufficient description of false positives and agree this could definitely be improved upon in information provided to women when making decisions. Nevertheless, findings suggest quite a large number of false positives may be acceptable. We had commented on the fact that few of the studies providing information about the effects of screening presented information on the possibility of lower treatment morbidity and avoiding some treatment such as

		chemotherapy. We have added a comment on this limitation to the discussion. While we are not incorporating views from stakeholders directly into the manuscript your message from a patient perspective will be shared with task force
Bell	No	Thank you. No changes requested.
Buick	Blank	No changes requested.
Wittmer (peer reviewer)	One part that was not clear for me was the inclusion of articles submitted by stakeholders (literature flow). Given the fact that stakeholders may wish to know transparently what evidence was included versus not, I would find it helpful to see why stakeholder submissions were not retained (evidence already found through database searches ?), this does not seem to be commented in the paragraph on the literature flow.	Thanks for this request. We have added a comment on this to the manuscript, “All of the eligible studies submitted by stakeholders were also found in our searches.”