

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Assessment of the effects of decision aids about breast cancer screening: a systematic review and meta-analysis
AUTHORS	Martinez-Alonso, Montserrat; Carles-Lavila, Misericòrdia; Pérez-Lacasta, María José; Pons-Rodriguez, Anna; Garcia, Montse; Rue, Montserrat

VERSION 1 – REVIEW

REVIEWER	Jolyn Hersch The University of Sydney, Australia
REVIEW RETURNED	04-Apr-2017

GENERAL COMMENTS	<p>Thank you for the opportunity to review this review! This is a topical subject and it's a good time to be systematically bringing together the small but growing body of research about breast cancer screening decision aids. I think this paper will ultimately find a good home at BMJ Open, but I do think a little more work is needed before the manuscript is ready for publication. Please see below for detailed comments.</p> <p>1. To me, the title is a bit misleading because it could be misunderstood as 'the impact that DAs have on screening (behaviour)' whereas it's actually about the impact that screening DAs have on various outcomes (not including actual screening behaviour). I would suggest "Assessment of the effects of decision aids about breast cancer screening".</p> <p>2. Abstract:</p> <p>a. Abstract should include more methodological information eg about eligibility criteria (for selecting studies to review) and method used to appraise study quality.</p> <p>b. Outcomes – change "decisional conflict" to "decisional conflict and/or confidence".</p> <p>c. Results – can the authors please include numbers for the results reported? Why does it say "Confidence in the decision was lower in the intervention group" when the meta-analysis shown in Figure 3 did not find a significant effect?</p> <p>3. Both the title and abstract should mention that meta-analysis was done as part of this study.</p> <p>4. The Strengths and Limitations section should relate only to the methods (of the review and the included studies). This section is not meant for restating the results of the review.</p> <p>5. P11 line 46 re Eden study sampling bias – I find the meaning unclear here.</p>
-------------------------	--

	<p>6. P12 line 7 is helpful in characterising Tables 3 and 4 as dichotomous vs continuous variables – please include these terms in the relevant table and figure headings.</p> <p>7. I think it is important to mention somewhere – at the very least in Table A2.2 but possibly also somewhere in the main manuscript – that the Mathieu study involved young women deciding whether to (a) start screening in their 40s (ie before the recommended age of 50) or (b) wait until they are 50. So it's a bit different from other studies targeting women who are approaching 50 and deciding whether to screen as per their national program. This helps explain, for example, why in Table 3 the proportions intending to screen are so much lower for the Mathieu study than for the other studies.</p> <p>8. P14 line 18 (Table 4) the superscript “c” appears twice in the knowledge section (mistakenly I think?) as well as in the decisional conflict section (correctly I think).</p> <p>9. Table 4 footnotes “e” and “f” – why “change in” confidence scale? Aren't these just the raw scores?</p> <p>10. P15 line 30 – I do not understand this sentence at all.</p> <p>11. P17 line 22 – this sentence is not altogether accurate; Hersch et al measured some basic knowledge at baseline using a subset of items and showed similarity between groups.</p> <p>12. P17 line 35 – I think it's probably worth mentioning that the results of the follow-up for final screening participation are not in the included paper (and are not yet published).</p> <p>13. Did the authors use a quantitative measure of consistency (eg I squared) in the meta-analyses? Could this be added to the forest plots?</p> <p>14. Tables A2.4 and A2.5 Selective reporting (Reporting bias) appears to have been judged according to the response rate across groups, but shouldn't this be judged with reference to the trial protocol?</p> <p>15. Discussion – unfortunately for the authors, another similar review was published just a few weeks ago (https://www.ncbi.nlm.nih.gov/pubmed/?term=28289963), so the authors will now need to add a short discussion of the other study. The authors also need to discuss any potential limitations of the review itself (not just of the included studies), and comparing this review to the published review may help identify both strengths and weaknesses of this one.</p>
--	--

REVIEWER	Irene Wong University of Hong Kong, Hong Kong SAR
REVIEW RETURNED	10-Apr-2017

GENERAL COMMENTS	<p>Major comments:</p> <p>1) p.5, it is mentioned that there was a systematic review in regards to decision aids for people facing health treatment or screening decisions conducted by the Cochrane Database in 2014. Any research gap and specific motivation for authors to conduct this systematic review on breast cancer screening at present. Please justify.</p> <p>2) p.9, it is mentioned that only 14 articles were selected for evaluation of the inclusion/exclusion criteria out of the 607 unique citations the authors identified. Only 4 studies were included at last. The criteria are not clearly stated in the article. The article says that details of the selection process can be accessed at Table S2.1, but it's not included in the print copy that I have.</p>
-------------------------	---

	3) In Methods, please add a few lines about the inclusion/exclusion process, and in particular discuss the implications of the fact that the authors included so few studies out of what the authors found. Does this mean the quality of DA studies is usually low? Have the studies overlooked some important dimensions?
--	---

REVIEWER	Karen B. Eden Oregon Health & Science University, USA
REVIEW RETURNED	11-Apr-2017

GENERAL COMMENTS	<p>Overall:</p> <p>This paper addresses an important topic of understanding the effect of breast cancer screening decision aids on screening decisions. The paper has several places that need to need to be clarified to strengthen the paper. The choice to combine RCTs with a before-after study makes the validity of the findings hard to assess.</p> <p>DA is defined as decision aids but DA reads singularly in places and plural in other places. I would define DA as decision aid and use “DAs” when plural so that the reader understands it means decision aids. I would suggest having an editor carefully review manuscript as several places were hard to follow or contained errors in word tense, plural/singular word use, word choice etc.</p> <p>Abstract:</p> <p>Was a protocol written and registered? If so, please state where this was registered.</p> <p>Objective: this really is a review of studies that included women age 50 and under so this should be stated in the objective as the reader might think this is all age groups.</p> <p>Setting: Because this is a systematic review, list the various settings in the included studies.</p> <p>Primary and secondary outcomes: clarify “participation intention in breast cancer screening”. Is this intention to be screened? or is this actual screening? Using both “participation” and “intention” makes it unclear what was measured.</p> <p>Results: “Taking” should be “making”. I am not certain you can draw a conclusion about “confidence” since Eden had an increase in decision self-efficacy (several statements rated on confidence scale) and the two RCTs showed drop in decision confidence?</p> <p>The conclusion is hard to follow both in the abstract and in the main text.</p> <p>Strengths and limitations of the study:</p> <p>1. The statement “This is the first systematic review focused in the</p>
-------------------------	--

	<p>impact of DA on breast cancer screening.” There is a new systematic review by Ilya Ivlev that focused on intention so I would delete this statement.</p> <p>Ivlev I, Hickman EN, McDonagh M, Eden KB. Use of patient decision aids increased younger women's reluctance to begin screening mammography: a systematic review and meta-analysis. J Gen Intern Med, 2017. doi:10.1007/s11606-017-4027-9</p> <p>Introduction:</p> <p>There are several new papers by Heidi Nelson that came out late 2016 and early 2017 that updated the breast cancer evidence that would be important to replace the older studies cited in the manuscript.</p> <ul style="list-style-type: none"> • Siu A, U.S. Preventive Services Task Force. Screening for breast cancer: USPSTF recommendation statement. Ann Intern Med. 2016;164:279-296. • Nelson HD, Weerasinghe R, Wang L, Grunkemeier G. Mammography screening in a large health system following the U.S. Preventive Services Task Force Recommendations and the Affordable Care Act. PLoS ONE 2015;10(6):e0131903. • Nelson HD, Fu R, Cantor A, Pappas M, Daegas M, Humphrey L. Effectiveness of breast cancer screening: systematic review and meta-analysis to update the 2009 USPSTF recommendation. Ann Intern Med. 2016;164:244-255. • Nelson HD, Pappas M, Cantor A, Griffin J, Daegas M, Humphrey L. Harms of breast cancer screening: systematic review to update the 2009 USPSTF recommendation. Ann Intern Med. 2016;164:256-267. <p>Methods:</p> <ol style="list-style-type: none"> 1. If the protocol was written and published, please state this in the paper and provide a link. 2. Language: “We included articles reported in any language.” – Were non- English studies included? 3. “The search strategy was performed in MEDLINE and SCOPUS and adapted and replicated in EMBASE ...” – Scopus contains 100% EMBASE so this doesn't really add new information. 4. The methods should be read carefully for accuracy e.g., outcomes measures. Some of the studies (e.g., Eden) measured decision self-efficacy which measures confidence in several ways but the scale itself measures decision self-efficacy. 5. “All the studies satisfying the inclusion and exclusion criteria referred to design, participants and interventions were included in this review.” – this is hard to follow as it suggests that studies that
--	--

	<p>meet the exclusion criteria could be included. Please restate this.</p> <p>6. Used “Cochrane Collaboration’s tool for assessing risk of bias” is for RCTs; however, it appears to have been applied to the before-after study (Table A2.3)? See Ivlev 2017 for an approach to evaluate this type of study.</p> <p>7. What model for the meta-analysis was used (random/fixed)? Please provide these details.</p> <p>8. The results appear to pool overall effect across RCTs and a before-after study. Typically, we do not want to mix these studies with different quality of evidence in one analysis.</p> <p>Results:</p> <p>1. Was the Gummersbach leaflet actually a DA? The description of the intervention could be expanded on to clarify what it contained. Two leaflets were compared so it might have been hard to find an effect?</p> <p>2. Ivlev, JGIM paper has different values for undecided women.</p> <p>Martínez-Alonso before 55/75, after 15/75. Ivlev before 11/75, after 9/75. Can you verify these results? Can you check the Ivlev 2017 paper for including other studies and update your search?</p> <p>3. Did you assess strength/quality of evidence?</p> <p>4. Results, page 12, line 17, “taking” should be “making”. I am not certain you can draw a conclusion about “confidence” since Eden had an increase in decision self-efficacy (several statements rated on confidence scale) and the two RCTs showed drop in decision confidence? It might be important to discuss the different ways confidence was measured in these studies to explain the contrasting finding.</p> <p>5. Figures 2, 3: the overall effect was pooled across RCTs and before-after studies (I wouldn’t put these two study designs together).</p> <p>6. Table 3: Informed choice wasn’t measured in Eden (subnote ‘b’) so the footnote is confusing. Preparation for decision making is a different scale (than informed choice) and measured only after using an aid. I would delete this note.</p> <p>Discussion:</p> <p>1. page 17, line 49. “All studies evaluated brochures from the women’s perspective. . . The Eden study was an interactive website not a brochure. It might be helpful to have an appendix with more details on each of the interventions as all were not brochures.</p>
--	--

REVIEWER	Karsten Juhl Jørgensen The Nordic Cochrane Centre, Rigshospitalet Dept. 7811, Copenhagen, Denmark.
REVIEW RETURNED	18-Apr-2017

GENERAL COMMENTS	<p>This systematic review is important, well performed, and well-written. I have a few suggestions for improvements:</p> <p>1: A subgroup analysis of women who had already been offered screening versus women too young to have been offered screening may be relevant, if possible. An interesting finding in the study by Hersch and colleagues was that women who had previously made a decision were likely to hold on to that decision and defend it, rather than change opinion (in either direction) with increased knowledge. This is likely a general human mechanism. It is relevant as the impact of DAs may be highly dependent on prior decisions.</p> <p>2: It is a bit unclear exactly how you have quality-assessed the one observational study that you include. I would like to draw your attention to the ROBINS-I tool for this task, which in my opinion is the best one yet (we have recently used it successfully in a Cochrane review submitted a few weeks ago). See: http://www.bmj.com/content/355/bmj.i4919.</p> <p>3: I have seen so many observational studies of “how to reduce barriers to participation in breast screening” amongst various minority groups of women that I wouldn’t be the least surprised to see such a study performed in “low-income transgender persons of sub-Saharan descent living in the Outer Hebrides”. Although I am sure he/she is a wonderful person (although there is very likely less than one) and well worth our best health care efforts, I am sick and tired of such studies and I find them unethical. It could be interesting to see how many of your discarded studies (> 600) that belonged to this “reducing barriers/increasing uptake” group and contrast it to how many (4) studies that you found which focussed on increasing informed consent and knowledge of the intervention offered.</p> <p>4: I think a major limitation of lumping various studies of the effects of DAs is that their effects are likely highly context, varying for example by age group included, country (setting) and over time. This should be mentioned as a limitation in the Discussion.</p> <p>Minor points</p> <p>1. In the Objectives of your abstract, it seems that you planned to include “observational studies”, but later on it appears that you included only before-after studies. This needs clarification.</p> <p>2. It is not surprising that confidence in decisions was lower in the intervention group. Women go from being told that “You should have a mammogram or you need more than your breasts examined” (this is not a joke but an actual campaign message from the ACS) to being told about there being a balance between benefits and harms, some of which are serious. No wonder people start having doubts.</p>
-------------------------	---

	<p>This is worth mentioning in your Discussion.</p> <p>3. Your method of choosing whether to pool results is not precise. It should depend on whether the studies (methods, participants, setting etc.) are judged by you to be comparable. It should NOT include extracted “reported outcomes”. This should also determine your choice of either fixed effect or random effects meta-analysis. Please refer to the Cochrane reviewer’s handbook, freely available on line.</p> <p>4. Regarding your bullet points on strengths and limitations: “Decision aids do not affect decision conflict” appears to conflict with your findings (see above). Please clarify.</p> <p>5. The two last bullet points can be combined to one for clarity.</p>
--	---

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1 Jolyn Hersch

1. To me, the title is a bit misleading because it could be misunderstood as ‘the impact that DAs have on screening (behaviour)’ whereas it’s actually about the impact that screening DAs have on various outcomes (not including actual screening behaviour). I would suggest “Assessment of the effects of decision aids about breast cancer screening”.

Answer: We thank the Reviewer for her comments. We have changed the title according to her suggestion.

2. Abstract:

a. Abstract should include more methodological information eg about eligibility criteria (for selecting studies to review) and method used to appraise study quality.

Answer: Done.

b. Outcomes – change “decisional conflict” to “decisional conflict and/or confidence”.

Answer: Done.

c. Results – can the authors please include numbers for the results reported?

Answer: We have included numbers for the informed choice and adequate knowledge outcomes.

Why does it say “Confidence in the decision was lower in the intervention group” when the meta-analysis shown in Figure 3 did not find a significant effect?

Answer: We agree with the Reviewer that this sentence is not correct, given the results. We have changed the sentence to “We observed heterogeneity among the studies in confidence in the decision”.

3. Both the title and abstract should mention that meta-analysis was done as part of this study.

Answer: Done.

4. The Strengths and Limitations section should relate only to the methods (of the review and the included studies). This section is not meant for restating the results of the review.

Answer: We have modified the Strengths and Limitations section. The new version does not include the results of the review.

5. P11 line 46 re Eden study sampling bias – I find the meaning unclear here.

Answer: We have changed the sentence. As suggested by professor Eden, given that one of the included studies was not a randomized controlled trial, we changed the tool for assessing risk of bias (see Table A2.3 in the supplementary appendix 2). Now we mention that the sample representativeness is limited.

6. P12 line 7 is helpful in characterising Tables 3 and 4 as dichotomous vs continuous variables – please include these terms in the relevant table and figure headings.

Answer: The terms have been included in the table and figure headings.

7. I think it is important to mention somewhere – at the very least in Table A2.2 but possibly also somewhere in the main manuscript – that the Mathieu study involved young women deciding whether to (a) start screening in their 40s (ie before the recommended age of 50) or (b) wait until they are 50. So it's a bit different from other studies targeting women who are approaching 50 and deciding whether to screen as per their national program. This helps explain, for example, why in Table 3 the proportions intending to screen are so much lower for the Mathieu study than for the other studies.

Answer: We have added a sentence in the Results subsection 'Study characteristics' and 'Choice of BC screening' that remind the reader of the difference between women in the Mathieu study and the other studies. Also, we have added this information in Table A2.2.

8. P14 line 18 (Table 4) the superscript “c” appears twice in the knowledge section (mistakenly I think?) as well as in the decisional conflict section (correctly I think).

Answer: Corrected.

9. Table 4 footnotes “e” and “f” – why “change in” confidence scale? Aren't these just the raw scores?

Answer: Corrected.

10. P15 line 30 – I do not understand this sentence at all.

Answer: The sentence was not correct. We have changed it to "Only Hersch noted a statistically significant decrease in the intention to be screened and Gummersbach a nearly significant decrease. The lower proportions intending to screen in the Mathieu study with respect to the other studies can be attributed to the fact that women were younger than 50, the recommended age for starting screening in Australia."

11. P17 line 22 – this sentence is not altogether accurate; Hersch et al measured some basic knowledge at baseline using a subset of items and showed similarity between groups.

Answer: Corrected.

12. P17 line 35 – I think it's probably worth mentioning that the results of the follow-up for final screening participation are not in the included paper (and are not yet published).

Answer: Mentioned.

13. Did the authors use a quantitative measure of consistency (eg I squared) in the meta-analyses? Could this be added to the forest plots?

Answer: Yes, we used I² to quantify consistency and the Q test to assess heterogeneity. These measures have been mentioned in the Methods section and the results have been included as footnotes of the forest plots that present the meta-analyses.

14. Tables A2.4 and A2.5 Selective reporting (Reporting bias) appears to have been judged according

to the response rate across groups, but shouldn't this be judged with reference to the trial protocol?
Answer: The reviewer is right and selective reporting was not properly assessed. Both tables have been corrected.

15. Discussion – unfortunately for the authors, another similar review was published just a few weeks ago (<https://www.ncbi.nlm.nih.gov/pubmed/?term=28289963>), so the authors will now need to add a short discussion of the other study. The authors also need to discuss any potential limitations of the review itself (not just of the included studies), and comparing this review to the published review may help identify both strengths and weaknesses of this one.

Answer: We thank the Reviewer for making us aware of the Ivlev et al. systematic review. This focuses on assessing the impact of DAs on intention to participate in screening. Our review includes the DAs impact on knowledge, informed choice, decisional conflict and/or confidence as well as screening intention. We have mentioned the Ivlev et al. review and commented on their results in the Discussion section, at the end of the first paragraph.

Reviewer: 2 Irene Wong

Major comments:

1) p.5, it is mentioned that there was a systematic review in regards to decision aids for people facing health treatment or screening decisions conducted by the Cochrane Database in 2014. Any research gap and specific motivation for authors to conduct this systematic review on breast cancer screening at present. Please justify.

Answer: We thank the Reviewer for their comments. As we mentioned in our response to the Editor, the recently published Cochrane review does not include any studies on breast cancer screening other than the ones included in the previous review, published in 2014. The Cochrane systematic review has a wider scope than ours, which is focused on DAs for breast cancer screening. In addition, due to the closing date of the Cochrane review (April 2015), three of our included studies (Hersch, Eden, and Gummersbach) were not included in it. Therefore, our study adds relevant information to the evidence at present.

2) p.9, it is mentioned that only 14 articles were selected for evaluation of the inclusion/exclusion criteria out of the 607 unique citations the authors identified. Only 4 studies were included at last. The criteria are not clearly stated in the article. The article says that details of the selection process can be accessed at Table S2.1, but it's not included in the print copy that I have.

Answer: We made a mistake and wrote Table S2.1 instead of Table A2.1. We have corrected it. Inclusion and exclusion criteria have been clarified in the manuscript for the studies, as well as for participants and interventions.

3) In Methods, please add a few lines about the inclusion/exclusion process, and in particular discuss the implications of the fact that the authors included so few studies out of what the authors found. Does this mean the quality of DA studies is usually low? Have the studies overlooked some important dimensions?

Answer: As suggested by the Reviewer, we have added more detail about the inclusion/exclusion process and also a paragraph in the Discussion about having included so few studies. Briefly, until recent years, DAs have not been commonly used in studies of breast cancer screening. In the past, many studies were designed to increase women's participation in screening and did not use DAs.

Reviewer: 3 Karen B. Eden

Overall:

This paper addresses an important topic of understanding the effect of breast cancer screening decision aids on screening decisions. The paper has several places that need to need to be clarified to strengthen the paper. The choice to combine RCTs with a before-after study makes the validity of

the findings hard to assess.

DA is defined as decision aids but DA reads singularly in places and plural in other places. I would define DA as decision aid and use “DAs” when plural so that the reader understands it means decision aids. I would suggest having an editor carefully review manuscript as several places were hard to follow or contained errors in word tense, plural/singular word use, word choice etc.

Answer: We thank the Reviewer for their comments. We have changed DA to DAs when plural. The manuscript has been reviewed for correct English.

Abstract:

Was a protocol written and registered? If so, please state where this was registered.

Answer: The protocol was not registered.

Objective: this really is a review of studies that included women age 50 and under so this should be stated in the objective as the reader might think this is all age groups.

Answer: The age specification has been added.

Setting: Because this is a systematic review, list the various settings in the included studies.

Answer: We have added the setting in the studies' description tables in supplementary appendix 2.

Primary and secondary outcomes: clarify “participation intention in breast cancer screening”. Is this intention to be screened? or is this actual screening? Using both “participation” and “intention” makes it unclear what was measured.

Answer: Changed to "intention to be screened".

Results: “Taking” should be “making”.

Answer: Corrected.

I am not certain you can draw a conclusion about “confidence” since Eden had an increase in decision self-efficacy (several statements rated on confidence scale) and the two RCTs showed drop in decision confidence?

Answer: The Reviewer is right. We have modified the conclusion to indicate that there was heterogeneity in the studies' results on decision confidence.

The conclusion is hard to follow both in the abstract and in the main text.

Answer: It has been rewritten for clarity.

Strengths and limitations of the study:

1. The statement “This is the first systematic review focused in the impact of DA on breast cancer screening.” There is a new systematic review by Ilya Ivlev that focused on intention so I would delete this statement.

Ivlev I, Hickman EN, McDonagh M, Eden KB. Use of patient decision aids increased younger women's reluctance to begin screening mammography: a systematic review and meta-analysis. *J Gen Intern Med*, 2017. doi:10.1007/s11606-017-4027-9

Answer: We have added the Ivlev et al. citation and commented on the similarity between their results and ours at the end of the first paragraph of the Discussion. We have added the outcome measures that we have included in our work to the statement that the Reviewer has commented on. Our revision can be considered the first that includes other outcomes besides intention to be screened.

Introduction:

There are several new papers by Heidi Nelson that came out late 2016 and early 2017 that updated the breast cancer evidence that would be important to replace the older studies cited in the manuscript.

- Siu A, U.S. Preventive Services Task Force. Screening for breast cancer: USPSTF recommendation statement. *Ann Intern Med.* 2016;164:279-296.
- Nelson HD, Weerasinghe R, Wang L, Grunkemeier G. Mammography screening in a large health system following the U.S. Preventive Services Task Force Recommendations and the Affordable Care Act. *PLoS ONE* 2015;10(6):e0131903.
- Nelson HD, Fu R, Cantor A, Pappas M, Daegas M, Humphrey L. Effectiveness of breast cancer screening: systematic review and meta-analysis to update the 2009 USPSTF recommendation. *Ann Intern Med.* 2016;164:244-255.
- Nelson HD, Pappas M, Cantor A, Griffin J, Daegas M, Humphrey L. Harms of breast cancer screening: systematic review to update the 2009 USPSTF recommendation. *Ann Intern Med.* 2016;164:256-267.

Answer: We have added three of these references to the Introduction section.

Methods:

1. If the protocol was written and published, please state this in the paper and provide a link.

Answer: The protocol of the review was not published.

2. Language: “We included articles reported in any language.” – Were non- English studies included?

Answer: We did not exclude any article due to the language. However, we did not find any non-English article that was both identified using the search strategy detailed in the supplementary material and which met the inclusion/exclusion criteria.

3. “The search strategy was performed in MEDLINE and SCOPUS and adapted and replicated in EMBASE ...” – Scopus contains 100% EMBASE so this doesn’t really add new information.

Answer: Scopus includes most, but not all, of EMBASE content and index terms. Scopus searches focus on abstracts and citations, while a search in EMBASE provides additional insights as a result of the structured full-text indexing of content. Since Scopus does not use Emtree to facilitate synonym mapping and hierarchical searches, it may retrieve significantly fewer results than EMBASE.

In addition, EMBASE subheadings are not available on Scopus, so searches cannot be focused in the same way (for more details or examples of differences, see www.elsevier.com/solutions/embase-biomedical-research/learn-and-support).

4. The methods should be read carefully for accuracy e.g., outcomes measures. Some of the studies (e.g., Eden) measured decision self-efficacy which measures confidence in several ways but the scale itself measures decision self-efficacy.

Answer: We considered decision self-efficacy, the measure selected by Eden et al. (2015), to quantify women’s confidence in decision making about screening, since the authors themselves conclude in similar terms. We agree that this adds heterogeneity to the outcome measurements, and we express this in our review.

5. “All the studies satisfying the inclusion and exclusion criteria referred to design, participants and interventions were included in this review.” – this is hard to follow as it suggests that studies that meet the exclusion criteria could be included. Please restate this.

Answer: We thank the Reviewer for noticing this. We have restated the sentence.

6. Used “Cochrane Collaboration’s tool for assessing risk of bias” is for RCTs; however, it appears to have been applied to the before-after study (Table A2.3)? See Ivlev 2017 for an approach to evaluate this type of study.

Answer: We appreciate the Reviewer’s suggestion of using a more adequate risk of bias assessment tool for the before-after study. As in Ivlev 2017, now we have used the National Institutes of Health Quality Assessment Tool for before–after studies with no control group. We mention this in the

Methods section.

7. What model for the meta-analysis was used (random/fixed)? Please provide these details.

Answer: We used a random-effects model, numerically estimated by restricted maximum-likelihood (REML) with inverse-variance weights, since a simulation study selected REML as the most suitable estimation method (Viechtbauer et al 2002). Measures of model heterogeneity and consistency have been mentioned in the Methods section and the results have been included in the footnotes for the meta-analysis figures.

8. The results appear to pool overall effect across RCTs and a before-after study. Typically, we do not want to mix these studies with different quality of evidence in one analysis.

Answer: We have decided to keep the analysis of all the studies together in the main text and to present the results of the RCT analysis in the supplementary file, appendix 3. The RCT meta-analysis does not include decision conflict because only the Hersch study assessed it. Other outcomes do not differ from the overall analysis because they were assessed in RCTs exclusively. The RCTs meta-analysis shows that decision confidence significantly decreased in the intervention group.

Results:

1. Was the Gummersbach leaflet actually a DA? The description of the intervention could be expanded on to clarify what it contained. Two leaflets were compared so it might have been hard to find an effect?

Answer: We have added information on the Gummersbach leaflet in Table 1. Although the leaflet was not created in accordance with published criteria for evidence-based patient information (as stated by the author), we have included the study in the review because the intervention leaflet contains much more information relevant to decision-making (i.e. overdiagnosis) than the control leaflet.

2. Ivlev, JGIM paper has different values for undecided women.

Martínez-Alonso before 55/75, after 15/75. Ivlev before 11/75, after 9/75. Can you verify these results?

Answer: The values 55.0 and 15.0 in our previous Table 3 referred to the before-after means of the uncertainty scale in the Eden's study. Due to an error, we did not mention it. Now, we have excluded these data from Table 3, for clarity, given that it presents the results of dichotomous outcomes. In the present version we mention the uncertainty score results in the text.

Can you check the Ivlev 2017 paper for including other studies and update your search?

Answer: As suggested, we checked all the references identified by the Ivlev et al. systematic review. Although it was not clearly stated in our Methods section previously, we excluded pilot studies, as well as studies addressed to older women.

3. Did you assess strength/quality of evidence?

Answer: We assessed the risk of bias in all the studies included in the review, as well as the consistency and heterogeneity of the meta-analysis. These results are included in the supplementary appendix 2 and in the forest plots footnotes, respectively.

4. Results, page 12, line 17, "taking" should be "making".

Answer: Corrected.

I am not certain you can draw a conclusion about "confidence" since Eden had an increase in decision self-efficacy (several statements rated on confidence scale) and the two RCTs showed drop in decision confidence? It might be important to discuss the different ways confidence was measured in these studies to explain the contrasting finding.

Answer: The Reviewer is correct about the lack of clarity in our conclusion. We have re-stated the

sentences in the Conclusion section. We also have commented on the potential reason for the heterogeneity in the decisional confidence results. See subsections 'Study characteristics' and 'Choice of BC screening' in the Results section.

5. Figures 2, 3: the overall effect was pooled across RCTs and before-after studies (I wouldn't put these two study designs together).

Answer: As mentioned above, an additional analysis of the RCTs has been performed and its results have been added to the supplementary appendix 3.

6. Table 3: Informed choice wasn't measured in Eden (subnote 'b') so the footnote is confusing. Preparation for decision making is a different scale (than informed choice) and measured only after using an aid. I would delete this note.

Answer: We agree with the Reviewer and have deleted the subnote "b" from the Table 2 footnotes.

Discussion:

1. page 17, line 49. "All studies evaluated brochures from the women's perspective. . . The Eden study was an interactive website not a brochure. It might be helpful to have an appendix with more details on each of the interventions as all were not brochures.

Answer: We have changed "brochures" to "DAs" in the Discussion sentence. We have added information on the DAs in Table 1.

Reviewer: 4 Reviewer Name: Karsten Juhl Jørgensen

This systematic review is important, well performed, and well-written. I have a few suggestions for improvements:

1: A subgroup analysis of women who had already been offered screening versus women too young to have been offered screening may be relevant, if possible. An interesting finding in the study by Hersch and colleagues was that women who had previously made a decision were likely to hold on to that decision and defend it, rather than change opinion (in either direction) with increased knowledge. This is likely a general human mechanism. It is relevant as the impact of DAs may be highly dependent on prior decisions.

Answer: We thank the Reviewer for his comments. We agree with him on the interest of his proposal, but it is not possible to perform the subgroup analysis from the published information. In fact, even the information on the number of participants with a previous mammography is not reported for two of the four studies, as detailed in Table 2.

2: It is a bit unclear exactly how you have quality-assessed the one observational study that you include. I would like to draw your attention to the ROBINS-I tool for this task, which in my opinion is the best one yet (we have recently used it successfully in a Cochrane review submitted a few weeks ago). See: <http://www.bmj.com/content/355/bmj.i4919>.

Answer: ROBINS-I ("Risk Of Bias In Non-randomised Studies of Interventions") was published in 2016 as a tool for evaluating risk of bias in non-randomised controlled studies. The Eden 2015 study, the unique observational study included in our review, is a before-after intervention study. We followed professor Eden's suggestion of using the National Institutes of Health Quality Assessment Tool for before-after studies with no control group.

3: I have seen so many observational studies of "how to reduce barriers to participation in breast screening" amongst various minority groups of women that I wouldn't be the least surprised to see such a study performed in "low-income transgender persons of sub-Saharan descent living in the Outer Hebrides". Although I am sure he/she is a wonderful person (although there is very likely less than one) and well worth our best health care efforts, I am sick and tired of such studies and I find them unethical. It could be interesting to see how many of your discarded studies (> 600) that belonged to

this “reducing barriers/increasing uptake” group and contrast it to how many (4) studies that you found which focussed on increasing informed consent and knowledge of the intervention offered.

Answer: We agree with the Reviewer on the interest of this result. We reviewed all the identified 607 references and found a total of 17 original articles exclusively designed to increase participation.

4: I think a major limitation of lumping various studies of the effects of DAs is that their effects are likely highly context, varying for example by age group included, country (setting) and over time. This should be mentioned as a limitation in the Discussion.

Answer: We included women around the recommended starting age of population screening. The external validity due to the reduced number of countries in the included studies is a limitation that we have commented on in the Discussion section.

Minor points

1. In the Objectives of your abstract, it seems that you planned to include “observational studies”, but later on it appears that you included only before-after studies. This needs clarification.

Answer: We did not plan to include only before-after studies, but a before-after study was the only one that fulfilled the inclusion criteria (reporting results of using DAs in women facing breast cancer screening).

2. It is not surprising that confidence in decisions was lower in the intervention group. Women go from being told that “You should have a mammogram or you need more than your breasts examined” (this is not a joke but an actual campaign message from the ACS) to being told about there being a balance between benefits and harms, some of which are serious. No wonder people start having doubts. This is worth mentioning in your Discussion.

Answer: We agree with the Reviewer. We have added the following sentence in the Discussion section, when commenting the decision confidence results: " This result can be explained by the impact of the information on adverse events of screening."

3. Your method of choosing whether to pool results is not precise. It should depend on whether the studies (methods, participants, setting etc.) are judged by you to be comparable. It should NOT include extracted “reported outcomes”. This should also determine your choice of either fixed effect or random effects meta-analysis. Please refer to the Cochrane reviewer’s handbook, freely available on line.

Answer: We used a random effects meta-analysis due to the high heterogeneity in the intervention effects of the four studies included in this review. We pooled results after re-scaling those scores with different ranges and re-calculated some estimates to increase the comparability of results.

4. Regarding your bullet points on strengths and limitations: “Decision aids do not affect decision conflict” appears to conflict with your findings (see above). Please clarify.

Answer: We have eliminated this sentence as suggested by the Reviewers.

5. The two last bullet points can be combined to one for clarity.

Answer: We have combined both sentences and provided more information on the external validity (i.e. countries where the DAs were developed).

VERSION 2 – REVIEW

REVIEWER	Jolyn Hersch The University of Sydney, Australia
REVIEW RETURNED	19-Jun-2017

GENERAL COMMENTS	<p>The authors have addressed my previous comments appropriately. There are 2 parts of the manuscript that have been revised where I think small changes are still needed – please see below. Also, I believe some of the other reviewers have greater expertise than I do about meta-analysis, so I would defer to their evaluations in terms of the methodological details of the analysis.</p> <p>1. In the Introduction, the paragraph referring to Stacey et al now says “The authors concluded that, compared with usual care, people exposed to DAs feel more knowledgeable” whereas the original manuscript said DAs improve knowledge. I think it’s important to state that DAs have been shown to improve knowledge (objectively measured), not only users’ subjective perceptions of how knowledgeable or informed they feel.</p> <p>2. Table A3.2 is labelled “Risk differences for the dichotomous outcome informed choice” but I think the outcome presented is actually screening intentions, not informed choice.</p>
-------------------------	---

REVIEWER	Irene Wong University of Hong Kong, Hong Kong SAR
REVIEW RETURNED	29-Jun-2017

GENERAL COMMENTS	<p>In this revision, the authors addressed my earlier concerns regarding specific motivation for authors to conduct this systematic review on breast cancer screening at the moment, as well as the inclusion/exclusion criteria.</p>
-------------------------	---

REVIEWER	Karsten Juhl Jørgensen The Nordic Cochrane Centre, Rigshospitalet, Dept. 7811, Blegdamsvej 9, DK-2100 Copenhagen
REVIEW RETURNED	12-Jun-2017

GENERAL COMMENTS	<p>The authors have adequately addressed my comments. One suggestion that the authors may use at their discretion is to include in their Discussion (as it is a post hoc finding) the information presented in an answer to me. They find that 17 papers included in their search were described interventions to increase uptake compared to 4 studies designed to increase knowledge about the important benefits and harms of the intervention. This is remarkable given that the search was not designed to identify the studies with "increased uptake" as a goal - there is many, many more out there. This question is central to the important debate about medical ethics in relation to screening interventions - basically the old fashioned paternalistic attitude versus citizen involvement.</p>
-------------------------	--

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1 Reviewer Name: Jolyn Hersch

The authors have addressed my previous comments appropriately. There are 2 parts of the manuscript that have been revised where I think small changes are still needed – please see below. Also, I believe some of the other reviewers have greater expertise than I do about meta-analysis, so I would defer to their evaluations in terms of the methodological details of the analysis.

1. In the Introduction, the paragraph referring to Stacey et al now says “The authors concluded that, compared with usual care, people exposed to DAs feel more knowledgeable” whereas the original manuscript said DAs improve knowledge. I think it’s important to state that DAs have been shown to improve knowledge (objectively measured), not only users’ subjective perceptions of how knowledgeable or informed they feel.

Answer: We agree with the Reviewer. We have changed the sentence according to the Reviewer suggestion. Now it says: “The authors concluded that, compared with usual care, DAs increase participants’ knowledge, objectively measured. In addition, people exposed to DAs feel more knowledgeable,...”.

2. Table A3.2 is labelled “Risk differences for the dichotomous outcome informed choice” but I think the outcome presented is actually screening intentions, not informed choice.

Answer: The Reviewer is right. Thanks for noticing the mistake. We have corrected the title.

Reviewer: 2 Irene Wong

In this revision, the authors addressed my earlier concerns regarding specific motivation for authors to conduct this systematic review on breast cancer screening at the moment, as well as the inclusion/exclusion criteria.

Reviewer: 4 Reviewer Name: Karsten Juhl Jørgensen

The authors have adequately addressed my comments. One suggestion that the authors may use at their discretion is to include in their Discussion (as it is a post hoc finding) the information presented in an answer to me. They find that 17 papers included in their search were described interventions to increase uptake compared to 4 studies designed to increase knowledge about the important benefits and harms of the intervention. This is remarkable given that the search was not designed to identify the studies with "increased uptake" as a goal - there is many, many more out there. This question is central to the important debate about medical ethics in relation to screening interventions - basically the old fashioned paternalistic attitude versus citizen involvement.

Answer: We agree with the Reviewer and have adapted and included his comments to the third paragraph of the Unanswered questions and future research section of the Discussion. Specifically, we have added: “In our search we found 17 papers that described interventions to increase uptake compared to four studies designed to increase knowledge about the benefits and harms of the intervention. Given that the search was not designed to identify studies with "increased uptake", this finding adds information to the important debate about medical ethics in relation to screening interventions - basically the old fashioned paternalistic attitude versus citizen involvement and shared decision-making.”