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)	Overdetection in breast cancer screening: development and preliminary evaluation of a decision aid
AUTHORS	Hersch, Jolyn; Jansen, Jesse; Barratt, Alexandra; Irwig, Les; Houssami, Nehmat; Jacklyn, Gemma; Thornton, Hazel; Dhillon, Haryana; McCaffery, Kirsten

VERSION 1 - REVIEW

REVIEWER	Jo Waller University College London UK
REVIEW RETURNED	17-Jul-2014

GENERAL COMMENTS	This is an interesting and very well-written paper describing the development of a decision aid designed to inform women about the harms and benefits of breast screening, particularly with respect to overdetection. The studies are very clearly described. I only had a couple of very minor queries about the paper, one of which has already been addressed by the authors in the limitations section. The sample in Study 1 was rather highly educated, and also included participants recruited from 'contacts' of the researchers. I'm not sure if this method of recruitment could be described/explained in a little more detail - but the authors do acknowledge in the Discussion that this convenience sampling was less than ideal. The sample in Study 2 includes women from a broader range of educational backgrounds. I was interested in what type of icon was used in the icon arrays, in light of recent suggestions that the shape of the icon affects responses (Zikmund-Fisher et al, 2014).

REVIEWER	Karsten Juhl Jørgensen The Nordic Cochrane Centre, Copenhagen, Denmark
REVIEW RETURNED	30-Jul-2014

GENERAL COMMENTS	This is, to my knowledge, the most comprehensive attempt to develop an evidence-based decision aid for breast screening that has been undertaken. It a difficult task, as the subject is complex and many issues counterintuitive, as these initial finding also suggest. This project is well underway and my comments will be limited to suggestions for improvements to the manuscript. 1: The abstract needs a bit more work. Most importantly, it needs to more clearly convey that this is a sub-study and a part of a larger project. The objective here was to iteratively develop and pilot-test a

	decision aid for later use in a more comprehensive, randomized trial of its feasibility. This fundamental objective only becomes clear to the reader when looking closer at the manuscript proper. The final section in the Introduction reads “The goal was to produce… etc.” I suggest something similar appears in the abstract. The abstract also needs to clearly convey that this initial project included two stages with 15 and 34 people in each stage, respectively, rather than simply stating that 49 people were included. This is important, especially as only the 34 people included in the second stage were randomised to a control intervention (and the abstract currently does not convey that any randomization occurred at all). 2: The use of the term “overdetection” rather than “overdiagnosis” in the title and text is in contrast with most existing literature on the subject, which tend to use “overdiagnosis”. This includes earlier papers in BMJ by this group (reference 15 in this manuscript). The motivation for this is mentioned in one line on page 5 (focus groups confused “overdiagnosis” with “misdiagnosis”), but I am not quite content with this (I miss some motivation why “overdetection” would not cause the same confusion and doubt that a then.informed focus group would be able to judge this) and believe it merits further discussion. I suggest to se “overdiagnosis” in the manuscript and “overdetection” in the pamphlet, if the latter term should be used at all. One can use the accepted scientific terminology in a paper and supplant this with a (perhaps, perhaps not) more understandable lay-term in information material without problems. 3: One major limitation of this study is not mentioned in the manuscript but needs to be. It is quite possible that the upcoming randomised study will substantially over-estimate the impact of an evidence-based decision-aid on informed decision-making in a real-life setting. This is because, contrary to in a trial setting where participant are informed that their knowledge and understanding will be tested, many invitees will very likely not read the information at all but simply assume that a programme offered by a public authority is a good idea. Reading the material is the sine qua non of decision aids and I would think that some kind of test of what percentage of these information leaflets will end up being read by the general public, and what percentage will end up unopened in the ‘bin, would be mandatory to evaluate its worth – is this planned? 4: As stage 2 of this sub-study involves randomisation, I think a more thorough description of the randomization procedure is warranted (it is now one line in the last section on page 8). As the manuscript stands, this study would not rate as very high quality if included in a Cochrane-review and subjected to the Cochrane Risk of Bias tool – most items would be rated as “unclear”. This is likely entirely unjustified. I suggest you use the tool and other reporting guidelines as inspiration for items to mention about the randomisation process in your Methods section. 5: I support your use of absolute numbers rather than relative risks, but suggest you expand a bit on this choice and use the research by Gerd Gigerentzer to motivate it. 6: The Results section may become even more clear if more distinctively divided into results for Stage 1 and Stage 2, as these are two quite different studies. 7: Regarding the results for stage 2, the percentages are now a bit confusing. I know the underlying numbers are presented in Table 5, but suggest “26 of 34 (76%) found the booklet to be about the right length” etc. Also, some of the results could do with a simple p-value, e.g. to test if there was a significant difference between the number who found the leaflet slanted either for or against screening. 8: One information item that I missed in the decision-aid was the

	issue of whether breast screening reduces the need for invasive surgery. I am sure the authors know all the evidence and arguments on this (for example, our Cochrane-review and a BMJ-study of numbers from Norway indicate the breast screening does not reduce the need for invasive surgery, perhaps on the contrary). Unfortunately and counter to the evidence, many women are told and believe that they reduce their risk of losing a breast by going to screening. This is an important topic and the authors need to discuss why this was not included in their information material. One could also discuss why it is not mentioned that breast screening does not measurably lower the risk of dying from all causes combined, which would likely also come as a surprise to many women given that they are continuously told about the importance of their participation in pro-screening campaigns. The leaflet was likely long enough as it is, but please go over these important topics. Should future leaflets be expanded as understanding of overdiagnosis would increase? Should such information be explained in more detail in apps or on the internet and these referred to in the leaflet?

VERSION 1 – AUTHOR RESPONSE

REVIEWER #1

1.1: The sample in Study 1 was rather highly educated, and also included participants recruited from 'contacts' of the researchers. I'm not sure if this method of recruitment could be described / explained in a little more detail - the authors do acknowledge this convenience sampling was less than ideal.

RESPONSE: To further elucidate this method of recruitment, we have expanded our explanation in the 'Stage 1 interviews' section as follows (p.5):
"Six women were recruited by convenience sampling among our contacts; they were not familiar with the study but were friends, relatives or partners of the project team or of colleagues."

1.2: I was interested in what type of icon was used in the icon arrays, in light of recent suggestions that the shape of the icon affects responses (Zikmund-Fisher et al, 2014).

RESPONSE: To address this query, we have slightly expanded our description in the 'Key design features' section as follows (p.4):
"The expected frequency of each outcome is illustrated by an icon array – a visual graphic display representing numerator and denominator together via differently coloured filled circles arranged in a matrix."

REVIEWER #2

2.1: The abstract needs a bit more work. Most importantly, it needs to more clearly convey that this is a sub-study and part of a larger project. The objective here was to iteratively develop and pilot-test a decision aid for later use in a more comprehensive randomized trial of its feasibility. This fundamental objective only becomes clear when looking closer at the manuscript proper. The final section in the Introduction reads "The goal was to produce… " I suggest something similar appears in the abstract. The abstract also needs to clearly convey that this initial project included 2 stages with 15 and 34 people in each stage, rather than simply stating that 49 people were included. This is important, especially as only the 34 people in the second stage were randomised to a control intervention (and the abstract currently does not convey that any randomization occurred at all).

RESPONSE: To address these points, we have revised the first part of the abstract as follows:
Objective: To develop, pilot, and refine a decision aid (ahead of randomised trial evaluation) for women around age 50 who are facing their initial decision about whether to undergo mammography screening

Design: Two-stage mixed-method pilot study including qualitative interviews (n=15) and randomised comparison using quantitative survey (n=34)
Setting: New South Wales, Australia
Participants: Women aged 43-59 years with no personal history of breast cancer
Interventions: The decision aid provides evidence-based information about important outcomes of mammography screening over 20 years (breast cancer mortality reduction, overdetection, and false positives) compared with no screening. The information is presented in a short booklet for women, combining text and visual formats. A control version produced for the purposes of comparison omits the overdetection-related content.

2.2: Use of the term “overdetection” rather than “overdiagnosis” in the title and text is in contrast with most existing literature on the subject, which tends to use “overdiagnosis”. This includes earlier papers by this group (ref. 15). The motivation for this is mentioned in one line on p. 5 (focus groups confused “overdiagnosis” with “misdiagnosis”), but I am not quite content with this (I miss some motivation why “overdetection” would not cause the same confusion and doubt a then-informed focus group would be able to judge this) and believe it merits further discussion. I suggest to use “overdiagnosis” in the manuscript and “overdetection” in the pamphlet, if the latter term should be used at all. One can use the accepted scientific terminology in a paper and supplant this with a (perhaps, perhaps not) more understandable lay-term in information material without problems.

RESPONSE: Overdiagnosis can arise from multiple sources (Moynihan, Doust, & Henry, BMJ 2012). This is a relatively new field, and the terms are still being defined. However, the term ‘overdetection’ is increasingly accepted in the specific context of screening to distinguish it from overdiagnosis that occurs via other mechanisms, such as broadening disease definitions. Our recently published RCT protocol uses ‘overdetection’, and we consider it important for the current paper to be consistent with the protocol paper. We agree, though, that this issue merits further work in future.

2.3: One major limitation of this study is not mentioned in the manuscript but needs to be. It is quite possible that the upcoming randomised study will substantially over-estimate the impact of an evidence-based decision-aid on informed decision-making in a real-life setting. This is because, contrary to in a trial setting where participants are informed that their knowledge/understanding will be tested, many invitees will very likely not read the information at all but simply assume that a programme offered by a public authority is a good idea. Reading the material is the sine qua non of decision aids and I would think that some kind of test of what percentage of these information leaflets will end up being read by the general public, and what percentage will end up unopened in the bin, would be mandatory to evaluate its worth – is this planned?

RESPONSE: We thank the reviewer for raising these important points, but we do not feel the current paper on the decision aid development is the right place to discuss these aspects of the trial design. They will be discussed in future papers reporting results from the trial, which of course will also give details of the recruitment and follow-up rates to help contextualise the findings. The trial has been designed as an efficacy study aiming to test whether we can adequately communicate about overdetection in this way, under the best possible circumstances. It is not an attempt to measure the true impact of implementing a decision aid within the national program. We agree it is likely that many women would not read all the materials sent to them by a screening program and therefore the effect may be less in the ‘real world’ – this could be tested in a pragmatic study which may follow this trial, depending on the results.

2.4: As stage 2 of this sub-study involves randomisation, I think a more thorough description of the randomization procedure is warranted (it is now one line in the last section on page 8). As the manuscript stands, this study would not rate as very high quality if included in a Cochrane-review and subjected to the Cochrane Risk of Bias tool – most items would be rated as “unclear”. This is likely entirely unjustified. I suggest you use the tool and other reporting guidelines as inspiration for items to mention about the randomisation process in your Methods section.

RESPONSE: This paper is focused on development of the decision aid, including a small pilot of RCT procedures using draft materials. We have already published the RCT protocol and referred readers of the current paper to that one for additional methodological detail. Nonetheless, to address this issue we have added a bit of extra information to the ‘Stage 2 interviews’ section (p.8) as follows:

"HVRF interviewers telephoned women, invited those eligible to participate, and obtained oral consent. The interviewers were not aware of the randomisation sequence. Exclusion criteria were: personal history of breast cancer; increased risk of breast cancer; any mammogram in the past two years; or insufficient fluency in English. Table 3 shows Stage 2 sample characteristics. Although 36 women were randomised, two (6%) were lost to follow-up – one in each arm.
Procedure
Using a computer random number generator, participants were randomised to be sent either the intervention or control decision aid by post. Participants had been told that they would receive one of two versions of the booklet, but they were not aware of how the versions differed or which was the intervention arm."

2.5: I support your use of absolute numbers rather than relative risks, but suggest you expand a bit on this choice and use the research by Gerd Gigerenzer to motivate it.

RESPONSE: To highlight this important aspect of the decision aid design, we have added a sentence to the 'Key design features' section (p.4) as follows:
"Quantitative screening outcome information is stated transparently using absolute frequencies with a clearly specified reference class.[Gigerenzer et al, 07] The expected frequency of each outcome is illustrated by an icon array..."

2.6: The Results section may become even more clear if more distinctively divided into results for Stage 1 and Stage 2, as these are two quite different studies.

RESPONSE: Although Stages 1 and 2 involved different study methods, the findings presented in this paper were drawn from examining the data collectively. We feel the findings are most effectively presented in the current structure – that is, divided into sections about knowledge/understanding and about acceptability, with integrated learnings from the combination of qualitative interviews and quantitative survey data. Therefore we would prefer not to divide the Results section into stages. To make this clearer, we have slightly modified the opening sentence under Results (p.9):
"The Stage 1 (qualitative) and 2 (quantitative) interviews together highlighted several important challenges in the communication of information about unfamiliar aspects of screening…"

2.7: Regarding the results for stage 2, the percentages are now a bit confusing. I know the underlying numbers are in Table 5, but suggest "26 of 34 (76%) found the booklet to be about the right length" etc. Also, some of the results could do with a simple p-value, e.g. to test if there was a significant difference between the number who found the leaflet slanted either for or against screening.

RESPONSE: Regarding the first point – Because the underlying numbers are presented in Table 5, and to aid readability, we would prefer not to also add these numbers to the text. On the second point – This pilot study was not designed or powered to be able to conduct any statistical analysis, but rather to test the trial procedures and provide additional data on the draft materials. However, the reviewer's suggestion has helpfully prompted us to revise our statements about the 'balance' item. To avoid making inappropriate inferences from these data in the absence of statistical testing, we now simply direct the reader to the table and comment briefly, as follows (p.11):
"A question about whether the booklet was balanced or slanted towards or away from screening elicited a range of responses in both groups (see Table 5)."

2.8: One information item I missed in the decision-aid was the issue of whether screening reduces the need for invasive surgery. I am sure the authors know all the evidence on this. Unfortunately and counter to the evidence, many women are told and believe that they reduce their risk of losing a breast by screening. This is an important topic and the authors need to discuss why this was not included in their material. One could also discuss why it is not mentioned that screening does not measurably lower the risk of dying from all causes, which would likely also come as a surprise to many women. The leaflet was likely long enough as it is, but please go over these important topics. Should future leaflets be expanded as understanding of overdiagnosis would increase? Should such information be explained in more detail in apps or on the internet and referred to in the leaflet?

RESPONSE: We agree that these are important topics that warrant exploration but, as the reviewer notes, it is very difficult to cover everything in one accessible resource, particularly at this stage while

public understanding is so minimal. To address these points, we have added the following to the final paragraph (p.14) before 'Implications and conclusions':
"Although our current focus is on introducing to women the novel concepts of overdetection and overtreatment, as public understanding increases over time future decision aid developers might consider also trying to address the difficult issues of how screening may affect the extent of treatment women receive and the risk of dying from all causes."
In response to this comment, we felt it would also be helpful to underline the trial's purpose more clearly, early on in the paper. We have therefore slightly expanded our description of the trial at the end of the Introduction (p.3), as follows:
"The goal was to produce materials that we could then use in a randomised trial to assess whether information on overdetection makes a difference to women's views and decisions about screening,[17] with the potential for future adaptation into a resource suitable for distribution within organised screening programs."

VERSION 2 – REVIEW

REVIEWER	Karsten Juhl Jørgensen The Nordic Cochrane Centre, Rigshospitalet, Department 7811, Copenhagen, Denmark
REVIEW RETURNED	05-Sep-2014

GENERAL COMMENTS	The authors have responded adequately to my previous concerns and redacted the manuscript as necessary. The authors have not accomodated all the concerns raised in my review but have argued well why they have chosen not to do so. While we disagree on some of the choices, they are minor issues and a good case can be made for either point of view, in my opinion. I therefore have no further queries.