PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	The effect of information about overdetection of breast cancer on women's decision making about mammography screening: study protocol for a randomised controlled trial
AUTHORS	Hersch, Jolyn; Barratt, Alexandra; Jansen, Jesse; Houssami, Nehmat; Irwig, Les; Jacklyn, Gemma; Dhillon, Haryana; Thornton, Hazel; McGeechan, Kevin; Howard, Kirsten; McCaffery, Kirsten

VERSION 1 - REVIEW

REVIEWER	Jo Waller University College London UK
REVIEW RETURNED	20-Mar-2014

GENERAL COMMENTS	This paper describes the protocol for a much-needed trial investigating the impact of information on overdiagnosis on women's decision-making about breast screening. The primary outcome is informed choice, but information on other psychological variables and screening uptake will also be collected. The paper is clearly written and describes the trial design and procedures very comprehensively. Ethical approval has been granted and the trial has been registered. A few thoughts came to mind as I read the paper and I will note these in case they are useful. 1) The issue of defining adequate knowledge to facilitate informed choice is, as the authors are well-aware, a difficult one. It wasn't completely clear from reading the paper whether a cut-off score was going to be used to define 'good' knowledge, or whether certain elements of knowledge would be considered essential for informed choice (so that if these were incorrect, knowledge would be inadequate, even if the overall score was high). I suppose my key question is whether all women deemed to be making an informed choice will have to demonstrate a good understanding of over-detection? 2) How will the study be described to participants? Will it be clear that the main aim it to evaluate the impact of information on over-detection, and if so, might this in itself have an impact on outcomes? Or will the aims be described in a less specific way? 3) I was interested that the authors have decided to exclude women who don't read the information materials. I can see the logic of this, as the aim is to evaluate the impact of the information, but in a national screening programme, it is likely that many women do not read all the materials they are sent, so the study might over-estimate the true impact of changing communication materials.

	4) I may have missed this, but are the intervention and control leaflets identical in their format? I think it would be useful to make this clear. 5) I'm sure the authors have considered this, but is there no possibility of assessing screening uptake objectively using medical records rather than (or as well as) self-report? 6) An interesting question that isn't currently addressed, is whether over-detection information could have a differential impact on psychological outcomes in women who attend for screening depending on their screening result, e.g. if women understand about overdiagnosis, they may find an abnormal result more distressing because they realise their treatment could be unnecessary. I realise this isn't the core aim of the trial, but it might be something that could be explored in the qualitative stream.

REVIEWER	Karsten Juhl Jørgensen The Nordic Cochrane Centre, Copenhagen
REVIEW RETURNED	26-Mar-2014

GENERAL COMMENTS	This study aims to assess information at the time of invitation to the first screening round about the major harm of breast screening influence the decision to participate or not, as well as measures of quality of life and decision regret etc. Really, it is unethical to even ask the study question and do this trial: women should be informed about the major harms of all medical intervention, including screening, period. This is our obligation as Doctors, required by law. Sadly, it is probably necessary to do the trial regardless, as (to my knowledge) the UK is currently the exception in informing women about overdiagnosis, the remainder of screening programmes opting to value high participation over basic ethical considerations. The ethical merit of this study should therefore be seen in the light of past and current violations of basic ethical standards. I don't think I have much to criticize or contribute, as this protocol seems to be very thoroughly thought out. I do have a few comments, of course. 1: Regarding the evaluation of screening knowledge (pp 9, 20-29), I am not quite sure that I understand (or agree with) the rationale behind excluding those women who do not read the information material from the analysis. This would lead to selection bias, and likely overestimation of the effect of the material. Rather, I would note whether the woman had or had not read the material, analyse the data intention to treat style, and then use the information about how many read the material to say something about how likely it is to obtain informed choice in a real life setting. Later on (pp 15, it is stated that all analyses will be ITT. I find this a bit confusing and rather irreconsilable, but perhaps that's just me. 2: The authors should note that one of their previously studies found that the influence of overdiagnosis on the perceptions about breast screening is dependent on previous screening participation. It iseems a very human (and important) characteristic to aim to justify one's previous decisions. It is a bias that explains a lot in the screening controversy, also among experts, and is an important justification for this study to focus on women who receive their first invitation, rather than include women who have already made up

	their mind. 3: I can't see from the protocol how the information on benefits and harms will be presented. I assume absolute numbers will be a part of it, as this was the case in a previous study by this group, but is wuld be nice to make sure that this is the case. There is good evidence showing that absolute numbers are necessary for understanding the belance between benefits and harms. 4: Overdiagnosis is important and has been left out by all invitations in the past. It is therefore important to explore its role, as in this study. However, other harms are also important and have likewise been downplayed, e.g. the long term psycholsocial consequences of the many, many thousands of false positives. And what about all the "benign" surgical biopsies? According to the Marmot report, the absolute numbers break down as follows (all numbers per year in the UK): Number of lives extended: 1,400 Number overdiagnosed: 4,000 Number of false positives: 65,095 Benign core needle biopsies: 19, 467 Benign surgical biopsies: 1,539 (Note that the Marmot report provides these numbers in a "consequence-tree", per 100,000 screened. I multiplied these numbers by 27, as 2.7 million women were screened the relevant year in the UK. Note also that while the two first numbers are estimates with a huge amount of uncertainty attached, the latter numbers can be known precisely). My question is: would these additional numbers for other harmful consequences matter for decisions? Can we talk about informed choice at all without providing them? The authors have chosen to look at one single harm (arguably the most important one) in isolation and have left out other harms that are also important (but arguably less so). This choice should be discussed and motivated in the protocol, and ensuing papers.

VERSION 1 – AUTHOR RESPONSE

REVIEWER #1

1: The issue of defining adequate knowledge to facilitate informed choice is, as the authors are well-aware, a difficult one. It wasn't completely clear from reading the paper whether a cut-off score was going to be used to define 'good' knowledge, or whether certain elements of knowledge would be considered essential for informed choice (so that if these were incorrect, knowledge would be inadequate, even if the overall score was high). I suppose my key question is whether all women deemed to be making an informed choice will have to demonstrate a good understanding of over-detection?

RESPONSE: Overdetection is an important part of the knowledge scale. We will compare the groups on all subscales (general and overdetection-related) and expect them to differ on the overdetection subscales (conceptual and numerical). We have not designated specific items (e.g. on overdetection) as essential to achieving an adequate overall knowledge score. However, as overdetection is an important element of knowledge, women will have to demonstrate at least a basic conceptual understanding of overdetection in order to achieve an adequate overall knowledge score.
To clarify this point, we have added the following to the section on knowledge measurement (p.11):
"Women will have to demonstrate a basic conceptual understanding of overdetection, false positives, and the mortality benefit from screening to be considered as having adequate knowledge."

2: How will the study be described to participants? Will it be clear that the main aim is to evaluate the impact of information on over-detection, and if so, might this in itself have an impact on outcomes? Or will the aims be described in a less specific way?

RESPONSE: To clarify this, we have added the following to the 'Participant recruitment' section (p.7): "The trial's aims will be described in a general way, as 'a study to make sure that written information about breast cancer screening is clear and helpful to women', without specifically referring to overdetection."

3: I was interested that the authors have decided to exclude women who don't read the information materials. I can see the logic of this, as the aim is to evaluate the impact of the information, but in a national screening programme, it is likely that many women do not read all the materials they are sent, so the study might over-estimate the true impact of changing communication materials.

RESPONSE: This trial is designed to detect any intervention effect under ideal circumstances, which is appropriate for the first randomised trial in the world of consumer information about overdiagnosis. Pragmatic studies may follow depending on the results of this trial. We highlight that engagement will be equivalent between the two arms given the randomized design.
To clarify our reasoning, we have expanded our explanation in the 'Pre-intervention procedure and measures' section as follows (p.8):
"…maximise the likelihood that the individuals randomised will read their allocated intervention and complete the trial, which is designed to detect any intervention effect under ideal circumstances. This is appropriate for the first randomised trial of consumer information on overdetection in breast screening."
We are now in the field with the trial. Of the first 250 women recruited, all have read the initial materials in a timely manner. Therefore to date no one has been excluded on this basis. We will continue to monitor and report on these exclusions so that they can be taken into account during analysis and interpretation of the findings.

4: I may have missed this, but are the intervention and control leaflets identical in their format? I think it would be useful to make this clear.

RESPONSE: To clarify, we have added to the 'Intervention and control arms' section (pp.9-10) as follows:
"Both are identical in format; the control version was produced directly from the intervention booklet by simply deleting the two pages on overdetection and all other references to it (e.g., in Q & A and summary table). The sections on benefit and false positives are identical across versions in terms of content and format."
We have also slightly modified Table 1 such that the common content lines up across the columns.

5: I'm sure the authors have considered this, but is there no possibility of assessing screening uptake objectively using medical records rather than (or as well as) self-report?

RESPONSE: The reviewer raises an important question which we have considered at length. We would like to assess screening uptake objectively in addition to by self-report, but obtaining access to the necessary medical records would require participants' written consent. We are concerned that seeking this consent may influence screening behaviour and thereby bias the trial results, which we consider an unacceptable risk. Therefore we have decided to delay assessing screening from medical records (and requesting participants' permission to do so) until the end of the two-year follow-up period. We accept that there will be some attrition by then, but we consider that to be more acceptable than taking the risk of biasing the trial results by seeking this consent earlier.
We have added some more detail to the section about measurement of screening participation (p.13)

as follows:
"…this is a reliable indicator of actual breast screening behaviour in Australia (91% of women reported a mammogram accurately to within a year of the recorded date).[Barratt]" and
"At the end of the trial we intend to assess participants' screening attendance from screening records as well."

6: An interesting question that isn't currently addressed, is whether over-detection information could have a differential impact on psychological outcomes in women who attend for screening depending on their screening result, e.g. if women understand about overdiagnosis, they may find an abnormal result more distressing because they realise their treatment could be unnecessary. I realise this isn't the core aim of the trial, but it might be something that could be explored in the qualitative stream.

RESPONSE: We agree that this is an interesting question and it is our intention to explore this in the qualitative stream. We have slightly modified the following sentence in the 'Stream B: Qualitative study' section (p.14) to highlight this:
"…whether the information has any positive or negative impact on women's screening experience (e.g., the way in which women interpret, cope with, and act upon their screening results)."

REVIEWER #2

1: Regarding the evaluation of screening knowledge, I am not quite sure that I understand (or agree with) the rationale behind excluding women who do not read the information material from the analysis. This would lead to selection bias, and likely overestimation of the effect of the material. Rather, I would note whether the woman had or had not read the material, analyse the data intention to treat style, and then use the information about how many read the material to say something about how likely it is to obtain informed choice in a real life setting. Later on (p15), it is stated that all analyses will be ITT. I find this a bit confusing and rather irreconcilable, but perhaps that's just me.

RESPONSE: Please see our response to Reviewer 1's point #3 regarding the rationale behind excluding these women. We note that these exclusions occur prior to randomisation, therefore we can still include in the analysis all participants as randomised.
We have slightly modified the following sentence in the 'Pre-intervention procedure and measures' section (p.8) to add this clarification:
"If she has still not read the leaflet by the next contact, she will be excluded from the trial (prior to randomisation)."

2: The authors should note that one of their previous studies found that the influence of overdiagnosis on the perceptions about breast screening is dependent on previous screening participation. It seems a very human (and important) characteristic to aim to justify one's previous decisions. It is a bias that explains a lot in the screening controversy, also among experts, and is an important justification for this study to focus on women who receive their first invitation, rather than include women who have already made up their mind.

RESPONSE: We thank the reviewer for this important observation supporting our choice of target population. We have added the following to the 'Setting and participants' section (pp.6-7):
"This study focuses on women facing an initial decision about whether to screen, as our qualitative study found that women's perceptions of overdetection were influenced by their previous screening participation.[Hersch]"

3: I can't see from the protocol how the information on benefits and harms will be presented. I assume absolute numbers will be a part of it, as this was the case in a previous study by this group, but it

would be nice to make sure that this is the case. There is good evidence showing that absolute numbers are necessary for understanding the balance between benefits and harms.

RESPONSE: We agree with the reviewer and have slightly modified the following sentence in the 'Intervention and control arms' section (p.9) to clarify this:
"The expected frequencies of outcomes are illustrated and contextualised using icon arrays depicting the absolute numbers affected per 1000 women screened over 20 years.[Trevena]"

4: Overdiagnosis is important and has been left out by all invitations in the past. It is therefore important to explore its role, as in this study. However, other harms are also important and have likewise been downplayed, e.g. the long term psychosocial consequences of the many thousands of false positives. And what about all the "benign" surgical biopsies? According to the Marmot report, the absolute numbers break down as follows (numbers per yr in UK):
Lives extended: 1,400; Number overdiagnosed: 4,000; False positives: 65,095; Benign core needle biopsies: 19,467; Benign surgical biopsies: 1,539…
My question: would these additional numbers for other harmful consequences matter for decisions? Can we talk about informed choice at all without providing them? The authors have chosen to look at one single harm (arguably the most important one) in isolation and have left out other harms that are also important (but arguably less so). This choice should be discussed and motivated in the protocol, and ensuing papers.

RESPONSE: Both the intervention and control booklets present false positives information. This takes the form of a text description of what a false positive result is (including an acknowledgement that 'some women find that they keep worrying about breast cancer for a while afterwards') as well as an icon array illustrating the following cumulative absolute numbers among 1000 women screened in Australia over 20 years: (a) the total number of women who experience a false positive, and (b) the subset who undergo any biopsy as part of their follow-up diagnostic assessment.
We have edited our description of booklet contents (p.9) to clarify this via the text in parentheses:
"…estimates of breast cancer mortality benefit, false positives (including total number of women with a false positive, and number having a biopsy), and overdetection".
In addition, to expand on the brief summary of booklet contents given in Table 1 we have now included a supplementary file containing a table that provides a much more detailed description of booklet contents.
The reviewer also raises a point about the trial having been designed to focus on exploring the role of information about overdiagnosis rather than other important outcomes of screening. We feel this is justified given that previous breast screening decision aid trials (Mathieu et al, 2007 & 2010) have already examined the effects of numeric information about false positives and the mortality benefit from screening on women's decision making.

ADDITIONAL CHANGES

1: Following further discussion among the project team members and collaborators, we have changed our quality of life measure from the SF12 to the Consequences of Screening – Breast Cancer questionnaire (p.13).

2: We have slightly modified the text (p.14 under 'Methods of data collection') and Figure 1 to clarify that the qualitative interviews will take place either face to face or by telephone.