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)	Do Psychological Harms Result from Being Labeled with an Unexpected Diagnosis of Abdominal Aortic Aneurysm or Prostate Cancer Through Screening? A Systematic Review
AUTHORS	Cotter, Anne; Vuong, Kim; Mustelin, Linda; Yang, Yi; Rakhmankulova, Malika; Barclay, Colleen; Harris, Russell

VERSION 1 – REVIEW

REVIEWER	Philippe Autier International Prevention Research Institute, Lyon, France
REVIEW RETURNED	09-Jun-2017

GENERAL COMMENTS	The paper of AR Cotter et al examines in detail the literature on "labelling" of subjects being told they have an asymptomatic condition detected by a test. A "test" includes the screening tests, but also any test susceptible to inform on the presence of an asymptomatic condition. I agree with authors that labelling is a largely underappreciated harm, and that guidelines panels should incorporate this issue in their algorithms for recommendations. The two examples chosen by authors are relevant and the topic should certainly attract more attention from the medical community. The methodology is outlined with sufficient detail and authors must have identified most (if not all) eligible studies. The analytic method is also well described. Authors have adopted a rational methodology for analysing the more qualitative literature. I have few comments: 1. Introduction: "Diagnosis given to a patient not destined to benefit from earlier detection (i.e., due to unavailability of effective treatment)". This aspect deserves some more attention because other issues than the availability of effective treatment threaten the benefit: PSA screening in men older than 75 years is questionable because they are unlikely to "benefit" from early detection. Subjects with many co-morbidities and thus short life-expectancy may not benefit from early detection. However labelling may be harmful in all subjects unlikely to "benefit". 2.The literature search was on English language articles only. The mother tongue of this reviewer is not English and the reviewer is aware of quite a number of publications in non-English languages on the labelling topic, mainly in the psychological literature. I fully understand the focus on the literature in English. I would just encourage authors to point at the need to examine the works done on labelling in countries where English is not the national language, as reviews based on the methodology adopted in this paper may further substantiate the knowledge base on labelling.

	3. Suicide rates are known to be higher in subjects diagnosed with a cancer. A large literature exists on this association. Hence, cancer labelling due to cancer screening will contribute to increase suicide rates associated with the discovery of a cancer. This indirect way to point at the harmful effect of labelling should probably be stressed in the Discussion.

REVIEWER	Ingrid Flight, PhD, Research Fellow Flinders Centre for Innovation in Cancer, Flinders University of South Australia, Australia
REVIEW RETURNED	12-Jun-2017

GENERAL COMMENTS	Overall, this is a paper that addresses a very pertinent question, ie what if any psychological harm is caused by being told that, although there are no symptoms and you were completely unaware, you have a potentially life-threatening condition? How does one cope with being told that yesterday you were healthy, today you are [labeled as] not healthy? The authors provide a thoughtful introduction and discussion; my comments relate more to the conduct and presentation of the review. My major reservation is whether the authors have adhered to their research question and examined only the literature that examined psychological outcomes relating to diagnosis through screening (the patient was asymptomatic prior to screening) and excluded diagnosis through follow-up of symptoms. I acknowledge that this distinction may not always have been made clear and a quick dip in to a few papers addressing prostate cancer included in the review show that there is not always clarity about how patients in the study population were diagnosed. For example, Oba et al (2014) included patients who were scheduled for prostate biopsy to rule out cancer. But there is no description that I could find of why in the first place they were scheduled for biopsy. Similarly for Selli et al 2014. As a third example, it isn't clear when looking at Appendix E, population based studies, that only data for those diagnosed through screening are included; Carlsson et al (2013) outcomes appear to include other results in addition to suicide "within 6 months of diagnosis among those detected by health control" (which I understand means through screening). If the object of the review is to specifically address diagnosis through screening, the authors should include information about how this criterion was ascertained, including whether any attempt was made to retrieve data from authors. They would then have to decide whether they would exclude certain papers, include results from sub-groups if available or provide caveats to their results. Also, as a general comment, results would be easier to follow if papers were referenced in the text more than they currently are. For example, in their discussion of the results of quantitative studies (p.10), the authors do not differentiate between AAA or prostate cancer literature but provide aggregated, unreferenced data; they mention that only one of 6 studies presented results in terms of frequencies and severity but don't reference that paper to make it easier to locate, nor do they reference papers in the relevant Table 3

	(and Table 3 results differ from this statement in that it appears that n=6 prostate cancer papers provide frequency and severity of psychological state; this seeming contradiction may be a result of not referencing). Some specific comments follow. 1. Please provide a reference for the fact that only one of 19 screening recommendations has addressed labeling (p.6 line 53). 2. P.7 line 6, 'widespread screening programs'. It would be helpful if the authors expanded on where the screening programs are in place, what they entail and what the threshold is for a diagnosis of cancer. For example, I am not familiar with screening for AAA and a screening result may be a dichotomous yes/no for evidence of aneurysm, but for prostate cancer, is a screening result of a certain PSA level necessary for further investigation and diagnosis? 3. P.7 line 33. Why select only papers published since 1 Jan 2002? 4. Tables – the authors should define the various instruments that have been used to measure psychological states so that readers can understand the acronyms, and ideally with references. 5. Tables – extraction of data seems inconsistent – sometimes mean score and other quantitative data is included, sometimes not. Where there are quantitative results, include p-value and CIs where provided. Indicate if these data are not reported (where they should be). For example, the comment against Korfage et al (2006) in Appendix C states "… vitality scores significantly decreased at 2 weeks…" should include a p-value. 6. Tables – Appendix C Oba et al (2014). The table records n=184 participants, however Table 2 of this paper shows that n=99 prostate cancer patients participated in the study (the higher number indicates the number of couples participating). 7. Tables – population based studies. Relating to my earlier comment about focusing on only the asymptomatically screened population, do these data reflect results for only that population?

VERSION 1 – AUTHOR RESPONSE

Reviewer 1

Comment: Issues other than availability of effective treatment threaten the benefit from screening.

Response: We agree with the reviewer, and had thought that our reference to "overdiagnosis" would include the cases noted. To make sure, however, we added text (page 5) to include patients with limited life expectancy.

Comment: English language search.

Response: We agree with the reviewer and have tried to make this clear in the text Page 6). We also added text in the Discussion (page 13) to recommend a non-English language review.

Comment: Suicide rates increase with cancer.

Response: We agree and have added text in Discussion (page 12) to emphasize the issue of suicide.

Reviewer 2

Comment: Have the authors included only the literature related to the psychological outcomes related to diagnosis through screening rather than literature involving patients diagnosed by symptoms?

Response: We appreciate the reviewer's comment and have added text to the paper in both the Methods (page 6) and Discussion (page 14) sections to address this issue. As our review only examined results in studies of early-stage PCa (i.e., localized) and AAA (i.e., less than 5.0cm, when surgery is beginning to be considered), conditions that are almost always asymptomatic in the early stage, we have ensured that our results refer primarily if not entirely to patients diagnosed asymptomatically – i.e., by screening. As we note in the Discussion, however, even if a few of the participants in these studies were diagnosed by symptoms, this would likely bias the results in the direction of less psychological harm from screening. Thus, if anything, our review may under-estimate the psychological harms of labeling due to diagnosis by screening. To be as clear as possible, we have added references to "unexpected diagnosis" rather than only screening in the title and Introduction section (page 5).

Comment: Increase use of reference numbers in Results section.

Response: We have done this, see especially page 10 on results from quantitative studies and Table 3. There is no discrepancy in the references and the numbers, as is shown by the references. No previously included papers need to be excluded. We did not contact authors for further data, but have explained in the text (pages 6 and 14) and above how we included data only for early stage AAA and PCa (i.e., did not require immediate treatment), almost all of whom for these conditions were likely detected by screening (or when they were asymptomatic, thus giving an unexpected diagnosis). Table 3 does not conflict with the text. In doing this, we found a small error in Table 3 (over-counting of studies finding no evidence of harm) and have corrected this throughout the manuscript.

Comment: Provide reference for only one of 19 screening recommendations by the US Preventive Services Task Force has addressed labeling.

Response: This reference (now given on page 5 in the text) is from a careful review of USPSTF recommendation statements done by the authors at the start of this review, and updated throughout the review period. There is simply no referral to the issue of labeling in almost all of these documents. The senior author is a former member of the USPSTF (and a co-author of its Procedure Manual) and, in a consulting role, has attended USPSTF meetings for some years.

Comment: Identify where widespread screening programs are in place.

Response: It is common knowledge that screening for AAA and PCa has been recommended by such guideline groups as the USPSTF and specialty societies within the USA, and that screening for these conditions has also been discussed and sometimes implemented (at least on an opportunistic basis) in Europe as well. The UK National Screening Program, for example, endorses screening for AAA and shared decision-making for PCa screening. Some of the trials of screening for AAA have been done in the UK and in Denmark. The largest trial of screening for PCa was conducted in Europe. All of these trials were considered by many to be positive. Because we are uncertain which countries have formally adopted national screening programs for these conditions, we have changed the wording to refer to widespread recommendations for screening (page 6) rather than widespread screening programs. Note that AAA screening does identify “aneurysms” when the aortic diameter is 3.0 cm or greater, although surgical treatment is not recommended until (and if) the AAA diameter is 5.5 cm. Thus many people are identified as having AAA although they do not receive immediate surgical treatment, and many never require treatment at all.

Comment: Why select papers published only since 2002?

Response: See our answer to this query above.

Comment: Tables – define the instruments that have been used.

Response: All instruments are identified in the footnotes to the Appendix Tables. It is beyond the scope of this review to itemize the reliability and validity of each one. Many of these instruments are well known.

Comment: Tables - make extraction data consistent.

Response: Extraction data in the Appendices are now consistent, with p values where available. The papers rarely used confidence intervals. Note that there can be no confidence intervals or significance tests when no comparison is being made.

Comment: Appendix C, Oba paper – make sure correct number of participants is recorded.

Response: The correct number of participants is recorded in the table – adding men with and without prostate cancer to sum to the total number of participants.

Comment: Tables – population-based studies; are these studies referring only to men with PCa detected by screening?

Response: We agree that it is not clear that all of these cases of suicide and other serious outcomes are of men detected by screening, although several of the studies do provide data on men with localized prostate cancer – likely diagnosed unexpectedly by screening. With the large increase in PCa incidence since screening became common, it is clear that many (probably the great majority) of these cases were detected by screening.

But even if some were detected by work-up following symptoms, the bias would likely be toward minimizing psychological harm from labeling (as many would have been expecting this diagnosis). Thus, if anything, these studies likely under-estimate the effects of labeling by screening on negative psychological state (see added text, page 14).
New comments from BMJ Open

Comment: Provide a clean copy

Response: We have done so, although it is not clear whether you want this in Word or pdf format.

Comment: Contributorship statement

Response: We have made sure that the statements in the paper and in ScholarOne are the same.

Comment: References

Response: We have reviewed the references and corrected the ordering. All references are cited and coordinate with the text.

Comment: PDF format

Response: We have uploaded the Appendices in pdf format. We assumed you did not want the other documents in pdf, although we will provide them if you wish.

VERSION 2 – REVIEW

REVIEWER	Dr Ingrid Flight, Research Fellow Flinders University of South Australia, Flinders Centre for Innovation in Cancer, Bedford Park, SA 5042, Australia
REVIEW RETURNED	16-Aug-2017

GENERAL COMMENTS	I am sorry to seem very 'picky'; my aim is to not be critical but provide what I hope are constructive comments. In the end, it is up to the journal editor's discretion as to whether they are passed on. Revised manuscript title: I find it hard to ascertain, from the revised title, what the article is about. The comma is an unexpected place, which results in a not very clear sentence; I assume the words 'unexpected diagnoses by screening' are intended to define 'labelling'? If so, would be better to put the former words in parentheses. I have taken the liberty of suggesting some alternatives for the authors' consideration; they are only suggestions and can be ignored! They include the conditions to make it easier for database (eg PubMed) indexers to allocate terms and subsequently easier for users to locate relevant literature. • Do psychological harms result from being labelled with an unexpected diagnosis of abdominal aortic aneurysm or prostate cancer through screening? A systematic review. • A systematic review of psychological sequelae from unexpected diagnosis via screening of abdominal aortic aneurysm or prostate cancer.

	• Are there psychological harms as a result of labelling asymptomatic individuals as having abdominal aortic aneurysm or prostate cancer following an unexpected diagnosis through screening? A systematic review. Authors' response that "It is common knowledge that screening for AAA and PCa has been recommended by such guideline groups as the USPSTF …" and revised submission Page 5, "widely recommended screening programs" I'm not sure that this terminology is quite correct given that the USPSTF in 2012 recommended against routine PSA screening and I understand that the most recent draft guidelines recommend that the decision of whether or not to be tested must be individualised for men 55-69 years. I don't know what the situation is with AAA. I think it is the word 'programs' that implies population-based screening, so this concept could be re-worded. Appendix C, Oba paper It appears that the authors are reporting results for Time 1 in this study, which I understand from the paper is prior to biopsy, with Time 2 representing 1 month following diagnosis (methods section of the Oba paper, p.464), the outcome of interest (and the results show there is a statistically significant difference at one month between those with PCa and those without). Appendix E, population based studies I apologise for not originally commenting on the data entered in to these tables but felt I should at this stage. Although the authors state that the population studies cover a longer period, I felt the authors should be consistent with their stated objective of reporting data for psychological harms and newly diagnosed, early stage cancer or otherwise expand and justify the criterion for population studies. • The authors state on page 11 of the revised submission that cardiovascular outcomes were a reported outcome. While of itself potentially interesting information, this outcome is not pertinent to the stated objectives of the review, ie psychological harms. If the authors believe this is an important outcome within the bounds of this review, this outcome should be mentioned and justified earlier. • Carlsson et al, Bill-Axelson et al: Results seem to be included in the incorrect column, ie under 'Recruitment, comparison group, time point' rather than 'outcomes and reported results' columns. • Bill-Axelson et al: Separate to the above comment – reporting the method of suicide does not seem to be pertinent to the objectives of the review. If this information is deemed pertinent, efforts should be made to find out suicide methods for the other population studies. • Bill-Axelson et al: Some of the reported results (ie, including data for men with locally advanced or metastatic disease) are at odds with the fact that the authors state that they intended to focus only on studies with newly diagnosed, early stage PCa and AAA (eg, revised submission pp. 6, 12, 14); also the abstract states suicide results from population-based studies for "patients recently diagnosed with PCa". • Klaassen et al: Similar comment to that for Bill-Axelson above; the data presented indicates higher rates of suicide, "especially □15 years since diagnosis". This information isn't pertinent to the "newly diagnosed, early stage" study population criterion. I would have thought only the 0-5 years data would be considered, particularly as the authors conclude that population studies show increased suicide rates in men 'recently' diagnosed with PCa (p.12).

VERSION 2 – AUTHOR RESPONSE

Responses for Reviewer 2

- Manuscript title: We have changed the title as you suggest (see above)

- Re prostate cancer screening recommendations: In the US, specialty organizations have recommended screening. You are correct, however, that the USPSTF new recommendation is for shared decision making. This is also the case in several other developed countries. Nevertheless, there is widespread screening in the US and, perhaps to a different degree, in other developed countries, especially encouraged by the results of the ERSPC trial. We do agree, however, that the word “program” is not correct, and thus we have changed the wording to indicate there is widespread screening, deleting the word “program”. (page 2 – abstract – and page 6 in text)

- Re: AAA screening recommendations: Both the USPSTF and the UK National Screening Program do recommend screening for AAA in men ages 65-74. As above, however, we have eliminated the word “program”.

- Appendix C, Oba paper: Thank you for pointing out this transcribing error in summarizing the results of the Oba study. The correct comparison for our outcome of interest is the change in the distress score for the PCa group vs the no-PCa group. This statistical test is not given in the paper. However, the PCa patients scored higher than the no-PCa patients both before and after biopsy. There was a slight increase in distress in the PCa group and a slight decrease in the no-PCa group after biopsy compared with before biopsy. (Note that this distress scale is a general measure that is not sensitive to the types of changes in distress expected with labeling, as we point out in the paper.) Thus we have conservatively concluded that psychological harm is uncertain in this paper. We think this is the best classification in this situation, given our explicit classification scheme. We have revised the results section in Appendix C to reflect this analysis. (see highlighted section of Appendix C)

- Appendix E, population-based studies have different criteria for stage of prostate cancer:

As some advanced and metastatic prostate cancers at diagnosis are still detected by screening, and as some of the population-based studies did not stratify their results by stage at diagnosis, we have explained in our methods section (highlighted section, page 7) that we still accepted results from longer-term population-based studies, as long as they also provided results from short-term follow-up. We think this is consistent across population-based studies and does not change any conclusions. We have also been careful to state the results and conclusions correctly. (highlight section, p 12 and 14)

- Appendix E, use of cardiovascular outcomes:

We have added in our Methods section (highlighted p 7) an explanation that we reported conditions such as cardiovascular disease that could be a result of psychologic distress, although our focus was on more direct psychologic outcomes. We have removed a mention of cardiovascular outcomes in our Results and Discussion sections. (P 12 and 14)

- Appendix E, Carlsson et al and Bill-Axelson et al, results in wrong column

We appreciate this comment and have changed the columns appropriately.

- Appendix E, method of suicide

We agree and have deleted this report.

- Appendix E, Bill-Axelson et al, locally advanced and metastatic PCa

As above, we have reported results from more advanced PCa in population-based studies as at least some of these cancers were likely detected by screening. We only report these results, however, when early stage results are also presented in the same paper. The most important finding in Bill-Axelson is that the RR for suicide was increased especially shortly after diagnosis. We have amended our Methods section to explain this inclusion. (p7) These inclusions do not add more studies to our review, but only corroborate the primary conclusions. We have been careful to state this correctly in our Results and Discussion sections. (p 12 and 14)

- Appendix E, Klassen et al, results from longer time period after diagnosis

As above, we have amended our Methods section (p 7) to explain our reporting of these longer-term outcomes, noting also results from the shorter term outcomes. In Klassen et al, the suicide risk was elevated in all follow-up periods.

VERSION 3 – REVIEW

REVIEWER	Ingrid Flight, PhD Flinders University of South Australia, Flinders Centre for Innovation in Cancer, Australia
REVIEW RETURNED	20-Sep-2017

GENERAL COMMENTS	Thank you for your response to my earlier comments.

Correction: *Do psychological harms result from being labelled with an unexpected diagnosis of abdominal aortic aneurysm or prostate cancer through screening? A systematic review*

Cotter AR, Vuong K, Mustelin LL, *et al*. Do psychological harms result from being labelled with an unexpected diagnosis of abdominal aortic aneurysm or prostate cancer through screening? A systematic review. *BMJ Open* 2017;7:e017565. doi:10.1136/bmjopen-2017-017565

The first five authors share first authorship: Anne R Cotter, Kim Vuong, Linda L Mustelin, Yi Yang and Malika Rakhmankulova.

Author 'Linda L Mustelin' should be written without a middle initial as 'Linda Mustelin'. She is also has the following additional affiliation:

University of North Carolina, School of Public Health, Chapel Hill, North Carolina, USA

BMJ Open 2018;**8**:e017565corr1. doi:10.1136/bmjopen-2017-017565corr1