Author’s response to reviews

Title: Do we know enough about the effect of low dose computed tomography screening for lung cancer on survival to act? A systematic review, meta-analysis and network meta-analysis of randomised controlled trials.

Authors:

Huiqin Yang (H.Yang@exeter.ac.uk)
Jo Varley-Campbell (jo.varley-campbell@ucl.ac.uk)
Helen Coelho (H.Coelho@exeter.ac.uk)
Linda Long (L.Long@exeter.ac.uk)
Sophie Robinson (S.R.Robinson@exeter.ac.uk)
Tristan Snowsill (T.M.Snowsill@exeter.ac.uk)
Ed Griffin (E.A.Griffin@exeter.ac.uk)
Jaime Peters (J.Peters@exeter.ac.uk)
Chris Hyde (C.J.Hyde@exeter.ac.uk)

Version: 2 Date: 03 Oct 2019

Author’s response to reviews:

Reviewer reports: Response

Reviewer #3: General comments

1 Overall, I think the author's conclusion that currently there is still uncertainty regarding the effectiveness of these screening strategies is an important one, that needs to be heard. Though, the review still lacks elaboration in some places; see specifics below. But first two general issues: No response required

2 In the abstract, the authors state "There was a non-statistically significant increase in all-cause mortality." This is a confusing statement: it suggests that the increase was significant, but not statistical. If the notion of significance is to be used, it is to be stated as "There was a statistically non-significant increase …". The same goes for similar statements throughout the text. We have implemented the suggested solutions throughout the paper
3 The abstract is missing more information regarding the interpretation of the results. For instance, the conclusion of the abstract does not mention all-cause mortality, which is mentioned as a primary outcome in the text. See more below. I'll start with some specific issues that have remained after the first round of reviewal, and continue with more issues. Added after first sentence of conclusions: “The uncertainty about the effect on all-cause mortality is even greater.”

Comment 5.

4 I agree with reviewer 1 that issues such as harms and overdiagnosis need to be considered when making statements such as "CXR screening had a 99·7% probability of being the worst intervention with usual care intermediate." Though, I also agree with the authors that it is still important to look at mortality outcomes in isolation. No response required

5 Yet, the conclusions, including those in the abstract, need to make clear that these issues have not been considered, i.e. that any statements are made only with respect to lung cancer or all-cause mortality. For instance, I suggest rephrasing this example sentence to: "The results showed that in terms of lung cancer mortality reduction LDCT was ranked as the best screening strategy, CXR screening as the worst strategy and usual care intermediate." and preferably include statements regarding the uncertainty of these ranks. We have made change suggested

Comment 6.

6 Considering that the research question is "Do we know enough about..." solely estimating a RR is not enough. Instead, I believe it certainly is necessary to grade the level of evidence, as this is precisely what the research question entails. I believe GRADE is the preferred method here. We have formally GRADED the evidence as suggested. According to this the RCT evidence would need to be downgraded on all three of the main features suggested: imprecision; inconsistency and indirectness. With the exception of one included RCT there is minimal risk of bias according to our quality assessment of the included studies. Publication bias is difficult to assess. It is unlikely at the trial level, but there is the possibility that there may be publication bias at the outcome level. On this basis we believe our main conclusion that there is uncertainty about the effect of LDCT screening for lung cancer is justified. We have added the following sentence at the end of the second paragraph of the conclusions to reflect this: “The uncertainty is confirmed if the evidence is GRADED, with downgrading for imprecision, inconsistency and indirectness.”

Comment 9.

7 Several elements have remained unclear regarding the NMA.

1. The authors mention it was a multivariate NMA, but does this mean they took the multivariate NMA approach (in contrast with the hierarchical approach or the meta-regression approach, see (2) ), or does this mean they analyzed the two outcomes simultaneously? Or the
multivariate NMA approach for both outcomes simultaneously? On page 5 we state that a "A multivariate random-effects meta-analysis using restricted maximum likelihood approach was performed". We did not perform a multivariate NMA.

8 2. Did the NMA method take within-study correlations into account? It’s not entirely clear to us what within-study correlations the reviewer is referring to. If this comment is in response to the reviewer assuming we did we a multivariate NMA then it is no longer relevant. It could relate to within-study correlations if studies with >2 arms are included (but this was not the case).

9 3. On what level of measurement has the meta-analysis been performed? No link function is mentioned. The NMA uses the point estimates (lnRR) and variance.

10 Line 48. Is it "CT has developed" or "CT has been developed"? On review of the sentence in question the original “CT has developed” makes best sense. An additional comma after developed may help however.

11 Page 4. Lines 8 -11 mention a study on LDCT and CXR. Then lines 12-14 state: "Investigators concluded that screening with LDCT reduces mortality from lung cancer." What happened to CXR? Does that mean LDCT reduces mortality from lung cancer compared to CXR? And then lines 14-15 mention only screening. Does that include CXR? Please clarify the text. We agree that there is a flaw in the logic applied to the interpretation of NLST as the reviewer points out. The problems associated with this are addressed later in the paper. However, in the introduction we merely wished to report what the original investigators had reported. To emphasise this we have made clear that “screening with the use of LDCT reduces mortality from lung cancer” is a direct quote from the paper cited.

12 Page 5. Lines 25 - 27. Please note that the Dersimonian and Laird method is not recommended. It is recommended to estimate the variance by REML, and apply the HKSJ modification to the confidence interval of the effect estimate instead.(1) Though, as REML has been used in the NMA I don't consider this a major issue. Thank you for pointing this out. We did consider reanalysing using REML, but were persuaded that this was unlikely to make a major difference to the result as indicated by the reviewer themselves.

13 Page 8, line 8. "a highly borderline non-statistically significant decrease". The p-value happens to be equal to the decision boundary. It cannot be "highly borderline" to the boundary. I suggest to remove "highly". "highly" removed as suggested

14 Page 8 "The main findings of this systematic review are a non-statistically significant decrease in lung cancer mortality (...) and a non-statistically significant increase in all-cause mortality outcome (...)" for what intervention and comparators? Also which method is used to estimate this RR, is it the MA or the NMA? Additional information added as suggested
Page 9 "The network meta-analysis confirms the likelihood that". The NMA was performed on largely the same data as the direct MA, making this is a misleading statement. I would rather say the results from the NMA were in agreement with those from the direct MA. Changed as suggested.

Page 9. In "This translates to a number needed to screen to avoid one lung cancer death of 357 (95% CI 82 to -113[...])" the point estimate falls outside the confidence interval, which makes no sense. The 95%CI includes infinity, and is a challenging feature of NNT arising from it being a reciprocal. As we had already added a note concerning interpretation of negative values of NNT, have added a further note, “..; the confidence interval includes infinity, equivalent to a risk reduction of 0, so the point estimate is encompassed by the interval although apparently lying outside it”

Page 10. The authors note in the manuscript that due to the possible ineffectiveness of CXR, the evidence from the NLST study that found a higher efficacy for LDCT (vs CXR) may not be considered evidence in the comparison of LDCT vs no screening. This has serious implications for the validity of the comparisons in the direct MA, and thereby the conclusions made in the manuscript and especially the abstract. We already recognise the serious implications by noting in the conclusion of the abstract that the first source of uncertainty is “the largest of the RCTs compared LDCT with CXR screening rather than no screening”. We are using uncertainty to mean “challenge to validity”. We would be happy to add this for emphasis, but are inclined against this as we believe most readers will understand that uncertainty implies challenge to validity.

Page 10. The authors mention statistical heterogeneity. This is at least partially a result of the mixing of the comparator: If the effect for LDCT vs CXR is truly different from the one for LDCT vs no screening, then the effect for LDCT vs (CXR or no screening) is guaranteed to be heterogeneous for a large enough number of studies. As indicated in the conclusion we believe the most likely source of the heterogeneity we identified is study quality. However, the reviewer is correct that in future meta-analyses, exploration of heterogeneity should include nature of the comparator, particularly whether the comparator was no screening (the most policy relevant comparator) or another active intervention like CXR screening. We have not made any changes responding to this comment.

Page 10. The 394 full text articles are accounted for by 18 systematic reviews + 128 citations in systematic review + 248 excluded articles = 394

References