Peer Review File

Article information: https://dx.doi.org/10.21037/tlcr-23-548

Reviewer A

Comment 1. The LUSI investigators and current authors are to be commended on a well-conducted study and well-written manuscript. I have a few points as detailed below.

Reply 1: Thank you for these encouraging words, and for your valuable time taken to evaluate our work.

Comment 2. The introduction does not clearly outline the research questions that the present study was designed to address.

Reply 2: We have restructured the second half of our Introduction section to make it clearer. We have now reversed the order of our research questions addressed, as this seems to be more intuitive to several of the reviewers:

1. using data of the German Lung Screening Intervention ("LUSI") trial, assess whether a selected set of candidate markers for lung cancer risk (CRP, IL-6, GDF-15), identified by independent recent studies, can be used to stratify lung cancer (LC) screening participants by their short- to medium-term risk of having or developing lung cancer;
   and
2. in parallel, assess whether the same markers, plus additionally NT-proBNP, can also be used to predict all-cause mortality and mortality by causes other than lung cancer.

The overarching objective is to examine whether (or which of) these markers could be used as meaningful predictors for risk of LC and/or competing-cause mortality, in view of improved, risk-stratified LC screening approaches.

Changes made in text: We modified the Introduction, to articulate these objectives more clearly, as explained in our reply above. You can see the highlighted text on the revised version of our manuscript. (page 3 lines 89-105)

Comment 3. The methods section does not include any information about the sample size and/or power of the study for the primary outcomes.

Reply 3: The LUSI trial is randomized lung cancer screening trial, which aimed to examine whether LDCT screening can reduce lung cancer-related mortality. The findings of this randomized comparison have been published earlier (Becker et al., Int J Cancer 2020).

Our present results originate from secondary analyses that were not defined by any original study protocol. Thus, there were no prior power calculations for the present work. However, for all of our present analyses we present hazard ratio and odds ratio estimates with confidence intervals and p-values. Adding post-hoc calculations for power and/or sample size would not add any further information, beyond the already indicated confidence intervals and p-values, and is actually considered inappropriate.

Indeed, while the utility of prospective power analysis in experimental design is universally accepted, statisticians recommend strongly against any such post-hoc calculations, as these can be fundamentally flawed. Some articles in which this is extensively discussed, explaining why post-hoc power calculations are inappropriate, include those by Dziak et al. (Curr Psychol. 2020 June ; 39(3): 870–877, Fraser et al. (Journal of Rheumatology 2022), or Griffith et al. (J Surg Res. 2021; 259: A9–A11).

Changes in text: For the reasons explained above, we did not perform any post hoc power calculations. However, we did add a short paragraph to the Methods section about numbers of outcomes (lung cancer; cases of death by lung cancer and other causes), relevant to the study questions that are now outlined more systematically in the Introduction (page 5 lines 167-179). This information was previously in our Results section, from where it has now been deleted.

Comment 4. The primary outcomes and confounding factors and any efforts to overcome bias are also not explained.

Reply 4: The central aim for our analysis was the confirmation of utility of selected biomarkers as risk predictors for lung cancer, as well as for death by competing causes, but not whether these markers are themselves causally involved in lung cancer etiology or death, independently of other causes. In a prediction context, contrary to studies aiming to establish etiologic relevance for a given risk factor, bias by confounding is not really an issue. However, from a prediction utility perspective, it is important to note that our analyses were, actually, adjusted for age, sex and detailed smoking history, as we wanted to show that the markers do indeed have major prediction potential beyond these classical risk factors. The outcomes of interest are the short-to-medium term risks of lung cancer, and the risk of all-cause mortality and mortality by causes other than lung cancer. We have restructured the second part of our Introduction section to make this point (even) clearer (please see our response to comment #2) and also made this point clearer in the Methods section, paragraph on statistical analyses).

Changes in text: We added some further sentences to the Methods section, to mention the use of age, sex, BMI and detailed smoking history (lifetime duration, cigarettes/day, time since quitting for ex-smokers) as model covariates (page 5 lines 194-196).

Comment 5. Given the number randomised, the number subsequently excluded and the final numbers in the present analysis a consort or flow diagram would be a useful addition to the paper.

Reply 5: Again, it should be made clear that our current findings did not primarily result from a randomized, controlled comparison between the two LUSI study arms.
Nonetheless, we have now added a flow diagram to the Supplemental files, to describe the exact origin of the data used for our present analyses.

Changes in text: new flow diagram in Supplemental figure 1

Comment 6. The results are interesting, though I am wondering how these biomarkers may be used particularly given the current important issues in lung cancer screening as acknowledged by the authors in the present paper. Namely the need to identify the group of individuals most likely to BENEFIT from screening (i.e. high risk of lung cancer, low risk of competing death).

Given the present study suggests using the same markers for predicting lung cancer risk/ death vs all cause presumably identifies more or less the same participants at high risk for both. How can these models therefore be applied in order to select the desired population that is enriched for lung cancer though at low risk of competing mortality?

Reply 6: To have a likely benefit from lung cancer screening, essentially two risk-related criteria should be met:

a. The likelihood of having a curable lung cancer tumor should be sufficiently high;
   AND
b. The absolute risk of death by competing causes (i.e., other than lung cancer) should be sufficiently low, such that, in case of an early tumor detection the screening participant may expect a meaningful number of life years. The latter criterion also helps keeping the risk of over-diagnosis low.

In combination, the two complementary criteria – the absolute probability of having a detectable lung tumor (lung cancer risk) and risk of short- to medium-term mortality – are the key determinants for expected benefit of LC screening (expected gain in life years, in case of early detection, and cure, of lung cancer). For individuals at a comparatively younger age (e.g. below age 65-70), having a sufficiently high LC risk will be the key criterion for screening eligibility, as well as screening frequency, at it determines the average number of individuals that to be screened (NNS), and thereby exposed to the potential harms of screening, for a case of lung cancer to be detected. For individuals at older age (mostly, those 70 years and older), the estimated risk of short-to-medium term (5-10 year) mortality becomes the key criterion to decide, on a personalized basis, whether or not it should be recommended to discontinue screening.

Please note also that not only the serum biomarkers evaluated in our present study, but also the "classical" risk factor variables used in established models to predict lung cancer risk and for determining the eligibility for lung cancer screening (i.e., age, lifetime duration of smoking, average number of cigarettes per day, time since quitting for ex-smokers) are each also major predictors for all-cause and mortality by causes other than lung cancer. The main objective of our analyses was to show these parallel risk relationships for the selected serum markers, which have been proposed for improved identification of individuals at risk for lung cancer.

Changes in text: We restructured and partially rewrote our Discussion text, with paragraphs more clearly articulated by the "themes" LC risk vs mortality risk, with clearer introductory lines of explanation of the reasons why both risk axes are relevant. We now also indicate more explicitly that the risk stratification is to be based on absolute (and not relative) risks, and that the use of elevated mortality risk as reason for screening dis-continuation would be relevant mostly for more elderly individuals (e.g. above the age of 70). The development of quantitative, absolute risk models, however, is beyond the scope of our present analysis (and of the LUSI study as study resource).

Comment 7. The limitations are barely mentioned in the discussion- this section needs to be considered and expanded upon.

Reply 7: We thank the reviewer, and also reviewers B and C ((comment #11 And #18), for raising this point.

Changes in text: We have now added a section on paragraph about strengths, limitations and needs for further work to our Discussion (page 9 -- lines 366-376).

Comment 8. In table 4- what is the explanation for why the AUCs are higher for lung cancer when the follow up is limited to <5years? Also the caption describes the outcomes in reverse order to how they appear in the table - consider revising.

Reply 8: We think that the decrease in AUC over follow-up times longer than 5 years may have resulted from variability over time in serum biomarker concentrations.

changes in text: We added a short paragraph to the Discussion to interpret the higher AUCs for lung cancer when the follow-up is limited to 5-9 years, and what practical implications this might have for the use of risk models and markers for risk-stratified screening: page 8-9 -- lines 335-342

Reviewer B

Comment 9. The authors present the results with the future perspective to personalize lung cancer screening by selecting patients at high risk of lung cancer while excluding those at a high risk of mortality by competing causes. This reviewer has struggled with this dual objective since the elevated biomarkers presented indicating a high risk of lung cancer are also those associated with high risk of mortality by lung cancer or other causes and therefore do not allow the intended discrimination.

Reply 9: For an individual to have a likely benefit from LC screening, two basic criteria need to be met: a. He/she should have a sufficiently high risk of having, or soon developing, a detectable lung tumor; and b. He/she should have a sufficiently long remaining life expectancy (i.e., a sufficiently low absolute risk of mortality by any cause other than lung cancer). With our analyses, we aimed to show that selected biomarkers (especially GDF15 and IL-6) can be used for the improved estimation of both risks, when added to known risk factors such as age, sex and detailed smoking history.
Please see our reply to comment #6, of reviewer A.

Changes in text: We have restructured the Introduction section to outline general background and study purposes more clearly (page 3 – lines 79-88; 89-105). Furthermore, we added explanatory sentences to the Discussion section to indicate why, and how the two criteria, LC risk and risk of short- to medium-term mortality by competing causes, would be used in practice for improving risk-stratified screening scenarios (page 9 – lines 354-363).

Comment 10. To increase the clinical relevance, the authors might sharpen the scope of the study by removing focus on mortality, COPD, and aging to strengthen the focus on only detecting patients at high risk of lung cancer. This might in turn increase the specificity of enrolment for lung cancer screening.

Reply 10: In view of optimizing the benefit of LC, through improved risk stratification, the risk of having lung cancer, as well as the short- to medium-term risk of death, are both relevant. Both of these risk axes should be considered for improving the specificity of enrolment for lung cancer screening, or also to decide the best moment for a screening participant to stop screening (e.g. when the absolute risk of death by competing causes has risen beyond a recommended threshold level).
It is our explicit objective to show how candidate markers may used for the estimation of, both, LC risk as well as the risk of death by competing causes (other than lung cancer). Please see our response to comment #6, of reviewer A, and to comment #9, above.
The associations of the serum markers with COPD / PRISm, however, are indeed more accessory findings; this is why these results are reported in the Supplemental materials only, and why we do not discuss these findings at any greater length. However, we found it important to also show that serum markers were associated with the risks of LC and competing mortality independently of COPD or PRISm as potential risk predictors.

Comment 11. For such a purpose, one could use time-to-event analyses, excluding prevalent lung cancer cases, and develop an algorithm trained on the data on GDF-15, IL-6 and other relevant biomarkers.

Reply 11: Our immediate purpose is more modest, namely to document the potential utility of pre-selected biomarkers for the estimation of LC risk, as well as the risk of death by any cause other than lung cancer, in a typical population of long-term smokers meeting existing criteria for LC screening. Our focus is on the estimation of relative risks associated with markers individually and independently of age, sex and smoking history (Tables 2 and 3), and on the improvements in risk discrimination (i.e., AUC) when marker measurements are combined with age, sex and detailed smoking history (our Table 4).

The further development of risk algorithms would require the development of absolute risk models, with a definition of possible risk thresholds for screening eligibility or its discontinuation. Our current findings serve to indicate that such algorithm might usefully include serum measurements for GDF-15 and IL-6, e.g. as part of accompanying general health checks for LC screening participants. The latter objective, however, is beyond scope for our current manuscript, as it would require quantitative (micro-)simulation modeling. In addition, LUSI study is too small, and as a stand-alone study not sufficient, for the developments of absolute risk models that can be generalized to diverse populations.

Changes in text: In the newly added section on study limitations (see also our responses to reviewers' comments #7 and #19), we now comment on the need for further development of algorithms for the estimation of absolute risks of LC and short/medium term mortality, for improved identification of individuals eligibility for LC screening, or for assignment of individually more optimized screening intervals. We here also mention the limitations of LUSI, as a standalone study, for such further development (e.g., too limited study size; lack of representativeness, as a stand-alone study, for diverse populations).

12. The algorithm should then be validated on a separate dataset.

Reply 12: Our study is itself actually an external evaluation, of the utility of blood-based biomarkers that have been identified by other recent studies (cited in the Introduction) as potential candidates for improved lung cancer risk prediction. We have now stated this more clearly in our Introduction section, as well as in the Discussion.

Changes in text: In the Introduction and the Discussion section we now state more clearly that our study is itself actually an external evaluation, of the utility of blood-based biomarkers that were previously identified in independent studies. With regard to algorithm development (for estimation of absolute risks of LC and competing mortality), we now clear state in our Discussion (new section on strengths and limitations (page 9 -- lines 366-376) that this is beyond scope of our present manuscript, and of the LUSI data as basic study: see our reply to comment #11, above.

Reviewer C

Comment 13. The markers included in this study are all well-established inflammation markers but are referred to as "aging-related" biomarkers throughout the manuscript. It might be more accurate to call them inflammation-related biomarkers.

Reply 13: We accept this comment, also because "aging" itself is not meant to be a central topic of our current manuscript.

Changes in text: We have now removed reference to "aging-related" markers from the Title, running title, Introduction, and other parts of the manuscript.

Comment 14. While the authors refer to prevalent and incident lung cancer cases, my understanding is that individuals with a diagnosis of lung cancer were not eligible for the CT screening trial and thus prevalent refers to cases identified in the first round of screening – could the authors confirm this is correct?

Reply 14: Thank you for pointing to a potential confusion. We made some changes to the text.

Changes in text: In the Statistical Analyses part of the Methods section we have rephrased "*For the present analyses, we excluded participants who had reported a past diagnosis of cancer (n=105), ...*" (page 5 -- lines 167-170). Furthermore, we have changed wording in the Results section, where we now state more clearly that "prevalent" LC cases were those identified in the first round of screening (which was indeed the correct interpretation of this reviewer).

Comment 15. Please explain this in the statistical analysis section where it is stated that prevalent cases were included in the analyses (as most studies do not include prevalent cases for risk analyses).

Reply 15: please see our response to comment #14, above

Comment 16. Given the screening trial, lung cancer cases are likely early cancers; since stage is available in this study, could you restrict prevalent cases to those that were early-stage cancers?

Reply 16: We performed additional analyses restricting to early stage lung cancers (stages I and II) only, and report findings in an additional supplemental table (Supplemental Table 4). Of note, IL-6 remains strongly associated to early stage lung cancers, whereas GDF-15 may be associated more with risk of late stage diagnosis. However, due to the highly reduced numbers of early stage cases, especially in the control arm, these findings by tumor stage may need require confirmation by other studies.

Changes in text: We added a Supplemental Table 4.

Comment 17. Statistical methods section: The details on description in the methods are lacking and do not match up with the results. For example, results mention pairwise Spearman correlations and findings from logistic regression models, neither of those methods are described in the methods.

Reply 17: Thank you again for careful reading. We have now rewritten our sub-section on statistical analyses (Methods section), mentioning the Spearman correlations and logistic regression modelling.

Changes in text: Rewritten sub-section on statistical analyses (Study Population and Methods) (page 5 -- lines 184-199).

Comment 18. Person-time estimations are not defined for prevalent/incident lung cancer (used in Table 4) – the information should provide start date, and censoring dates. Also, the results describe AUC curves and forward selection modeling for joint effects, which are not mentioned in the methods.

Reply 18: Thank you for calling our attention to this aspect, too.

Changes in text: We have now added all of this information, too, to the sub-section statistical methods (page 5 -- lines 191-194).

Comment 19. Discussion should include section with study strengths and limitations. There should also be some discussion of findings by follow-up and implications for the different findings.

Reply 19: Thank you for this raising this point. We added a separate section about strengths and limitations in the discussion.

Changes in text: see our response to comment #7 from reviewer A (page 9 -- lines 366-376).

Reviewer D

Comment 20. Please discuss more about the methodology for sample size estimation in the manuscript.

Reply 20: Please see our response to comment #3, from reviewer A

Comment 21. In the discussion section, please provide more information about how the serum GDF-15 and IL-6 could be useful for lung cancer screening and the relation with COPD and PRISm. It would be helpful if the authors give example or scenario to support its description.

Reply 21: Thank you; we added more information on these points to the discussion section.

Changes in the text: Changes to the Discussion section, lines 317-334, and 354-359.

Reviewer E

Comment 22. The authors assess the power of cancer prediction in an all-participant group ("entire follow-up") and in the subgroup "≤5 years follow-up". The presented results suggest (indirectly) significantly lower prediction power for cases diagnosed more than 5 years from baseline. Hence, additional analysis concerning the ability to predict cancer depending on the period from the time of the blood collection to the time cancer manifestation would be of great importance (e.g., separately in subgroups 0-2, 2-5, 5-10, >10 years).

Reply 22: We considered additional analysis for different time frames but our main restriction was sample size in each scenario. For instance, we only have 19 cases of lung cancer mortality in the first 2 years and 13 for mortality by causes other than lung cancer.
However, we did add results to Table 4, for the first 2 and first 5 years of follow up for lung cancer, and for the first 5 years for the mortality risk estimates. Is worth to mention that prediction did not differ substantially across the various time-restricted scenarios.

Changes in text: To Table 4, we added estimates for the prediction of lung cancer risk for the first 2 years of follow up, for all endpoints (i.e. lung cancer risk, mortality) over the first 5 years of follow-up.

Comment 23. A similar analysis of death prediction's time-dependent power would also be important. These time-refined analyses could be important for the decision-making process regarding the practical effectiveness of the potential biomarker.

Reply 23: Please see our response to comment #22, above.