Peer Review File

Article Information: Available at http://dx.doi.org/10.21037/tlcr-20-727

Review Comments

Reviewer A

Comment 1: LDCT is an effective tool to screen for lung cancer; however, it comes with known risks including a high rate of false-positive results, false-negative results, potential for unnecessary follow-up testing, radiation exposure, overdiagnosis, changes in anxiety level and quality of life, and substantial financial costs. There is a lack of these information in the introduction section, in the context of the urgent need of looking for a specific and sensitive tool in lung cancer diagnosis.

Reply 1: We added a sentence to the Introduction to mention not only false-positive tests, but also radiation-induced cancer, unnecessary follow-up testing and financial costs, and over-diagnosis as possible risks of LDCT screening.

Changes in the text: Sentence added: "it comes with risks of radiation-induced cancer, false-positive test results, unnecessary follow-up testing and increased financial costs, as well as overdiagnosis." (**Page 5 Lines 60 - 63**).

Comment 2: Moreover, autoantibodies are not described sufficiently in the overall context of lung cancer diagnosis, and commercial panel used in the study.

Reply 2: We added a more comprehensive description of tumor-specific autoantibodies in the context of lung cancer diagnosis and the commercial panel used in this study to the Introduction.

Changes in the text: *"… increased titers in cancer patients with various forms of solid tumors, including lung cancer. In fact, a number of individual TAAbs and multi-TAAb panels have been evaluated for its ability to discriminate lung cancer patients from cancer-free individuals {**Chapman 2008, PMID: 17932110; Broodman 2016, PMID: 27769114; Du 2017 PMID: 29356386; Qin 2018 PMID: 30562746; Yang 2019 PMID: 31396403**}. Panels have outperformed individual TAAbs in all studies, which has been attributed to the heterogeneity of lung cancer tumors {**Broodman 2016, PMID: 27769114**}. However, while most of the panels show good specificity, their sensitivity was generally only modest {**refs: Broodman 2016, PMID: 27769114; Qin 2018 PMID: 30562746; Yang 2019 PMID: 31396403**}. One well-established panel is EarlyCDT®-Lung (Oncimmune Ltd, Nottingham, United Kingdom) which, in its most recent version, iscomposed of 7 different antibody assays(CAGE, GBU4-5, HuD, MAGEA4, NY-ESO-1, p53 and SOX2), most of them not specific for lung cancer but arising also in other cancers such as breast, colorectal,*

gastric, prostate, liver and testis, as well as in autoimmune diseases ***{refs: Chapman 2008, PMID: 17932110; Broodman 2016, PMID: 27769114; Du 2017 PMID: 29356386****}.".* **(Pages 5-6 Lines 76-91)**

Comment 3: I would consider relocation of the lines 78-83 (about Scottish trial) into the discussion section.

Reply 3: We have followed this suggestion, and moved these lines to the Discussion.

Changes in the text: the paragraph: **"***EarlyCDT®-Lung is currently being tested as a first-line population screening tool in a Scottish trial (N=12,209) for the identification of subjects likely to harbor a lung tumor, who are then further examined by LDCT* ***[refs]****. In addition, EarlyCDT®-Lung has been evaluated as a confirmatory test in clinical settings, to decide whether a biopsy or surgical intervention for definitive diagnosis is indicated for subjects presenting with incidentally observed pulmonary nodules* ***[refs]***" was moved from **Page 6 Lines 96-101, to Page 15 Lines 297-302**.

Comment 4: At the end of the introduction section, there is an information about accordance of the article with the STARD reporting checklist. Authors should briefly describe what STARD is, present its statement or at least explain the abbreviation.

Reply 4: We took the original sentence as stated in the Guidelines for Authors from the TLCR journal website (http://tlcr.amegroups.com/pages/view/guidelines-for-authors#content-3-12), section 2.1.1. For clarity, we have now expanded the abbreviation and added a reference to the reporting checklist

Changes in the text: We expanded the STARD abbreviation "Standards for Reporting of Diagnostic Accuracy Studies" and added a citation to the corresponding reference {**PMID: 26511519}.** (**Page 7, lines 109-110**).

Comment 5: Authors should consider to present "nodule management protocol" in the table or in other schematic way.

Reply 5: We replaced/added detailed information on the nodule management protocol to the supplementary information in a tabular form.

Changes in the text: "*A schematic presentation of the nodule management protocol used in LUSI is given in Supplemental Table 1*" (**Page 8 Lines 142-143**)

Changes in the supplementary material: Added Supplemental Table 1. Supplemental Table 1 in the previous version became Supplemental Table 2.

Comment 6: Links to the websites should be placed in the references section, not in the text. Moreover, websites may change over time or disappear, so authors should create an archive of the cited websites.

Reply 6: The URLs of websites were replaced by links to the references section.

Changes in the text: The URLs *http://www.isrctn.com/ISRCTN30604390* (**Page 7 Line 126**); *https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Lung-Rads* (**Page 8, line 132**); and http://www.isrctn.com/ISRCTN30604390 (**Page 20 line 391**) were moved to the references.

Comment 7: More editing is needed in the text because some sentences read awkward and some of them are incomprehensible e.g. "At time of blood donation, lung cancer cases (n= 46, 32 of them males) were significantly older (median: 63.0 years, range: [51.9, 74.5]) than the control subjects (BC: median: 56.8, range: [50.9, 69.7], p<0.001); SNC: median: 55.8, range: [50.6, 70.0], p<0.001)."

Reply 7: Thank you for bringing this to our attention. We have edited our text to improve the reading experience. The sentence in the reviewer's comment was modified as described below. Additionally, a native English speaker (and Statistician) had a further look at the manuscript and suggested some additional linguistic edits, which are traced in the text.

Changes in the text: The sentence in the reviewer's comment was modified as follows: "*At the time of blood collection, the 46 participants eventually diagnosed with lung cancer (32 of them males) were significantly older (median: 63.0 years, range: [51.9, 74.5]) than those in both the BC (56.8 [50.9, 69.7], p<0.001) and the SNC groups (55.8 [50.6, 70.0], p<0.001)*"**. (Pages 12-13 Lines 243 - 246).**

All additional small linguistic edits are traced in the text.

Comment 8: Abbreviations should be checked because not all of them are explained (for example STARD, ISRCTN, ICD) and used consistently thereafter (VATS).

Reply 8: We checked and made sure that all abbreviations are explained / added list of references to manuscript.

Changes in the text: "*Standards for Reporting of Diagnostic Accuracy Studies"* (**Page 7 lines 109-110**), "*International Standard Randomized Controlled Trial Number*" (**Page 7 Line 125**), "*Lung Imaging Reporting and Data System*" (**Page 8 Line 131**), "*computed tomography*" (**Page 8 Line 145**), "*positron emission tomography*" (**Page 8 Line 145**), "*International Classification of Diseases*

*for Oncology, version 3" (**Page 9 Lines 166 -167**), "video assisted thoracoscopic surgery" deleted (**Page 9 Line 172**), "enzyme-linked immunosorbent assay (ELISA)"* (**Page 10 Line 183**) *"Odds Ratio" (**Page 14 Line 269**).*

Reviewer B

Comment 9: This is a timely and well-conceived study intended to determine the potential contribution of an autoantibody panel to lung cancer early detection. The study relied on an established German cohort and the analysis showed modest contribution of the marker panel. Given the interest in autoantibodies and the prior studies in this regard, this paper provide novel insights regarding the limitations of the panel.

Reply 9: We thank the reviewer for her / his appreciation.

Comment 10. Of course validation studies differ with respect to the subject population, the platform utilized to test biomarkers and the status of samples analyzed from collection to storage etc nevertheless the study is sufficiently informative to merit publication

Reply 10: We agree. Nonetheless, we would like to note that our study provides a representative case of performance of the Early®CDT test in a setting of population screening. The LUSI trial was conducted as a typical population-screening trial, amongst men and women at increased risk of having lung cancer in view of their age and smoking history, invited from the general population around Heidelberg, Germany. The Early®CDT test was performed by an experienced immunoassay lab, according to the manufacturer's instructions with strict adherence to the timing, development and measurement of results.

Changes in the text: none

Reviewer C

Comment 11: The application of biomarkers to lung cancer screening is currently a subject of considerable interest, so the subject matter is topical and medically relevant. The manuscript is well-written and the material is clearly laid out. The samples were selected from the German Lung Cancer Screening Intervention (LUSI) trial. The design and conduct of the LUSI trial are not covered in this review.

Reply 11. As mentioned in the text, the design and conduct of LUSI have been described in great detail elsewhere ([**Becker et al 2012 and Becker et al 2019**]).

Changes in the text: none

The data seem in general to be valid, although there are a few anomalies in the presentation that need correcting:

Comment 12: Figure 1: A few of the arrows and numbers in the top three or four lines of the Consort diagram do not seem to be correct.

Reply 12: We thank the reviewer for pointing to this. We corrected the arrows and numbers in the Consort diagram.

Changes in the table: Please see previous and revisited versions of the Consort diagram.

Comment 13: Supplemental Table 1: Available and Unsuspicious samples at Baseline Round 1 has n=1362. Are these the same as the n=1362 samples "provided independently of LDCT" stated in the "Blood sample collection protocol section"? If so, then the description is confusing.

Reply 13: Thank you also for this observation. We have corrected the description accordingly.

Changes in the text: the phrases "independently of findings on their LDCT examination" (**Page 8 Line 150**) and "irrespective of their LDCT scan results" (**Page 10 Line 176**) have been deleted.

Comment 14: Supplemental Table 1: The numbers in lines 4 and 5 of the table do not add up.

Reply 14: We have corrected the numbers in the "Total" column entries of rows 4 and 5.

Changes in the Table: The counts N=236 was replaced by 235, and 1791 replaced by 1675.

Comment 15: Supplemental Table 1: The section label "Taken at the time of suspicious CT scan findings" is a bit confusing as the table also includes non-suspicious findings.

Reply 14: We corrected our wording in the table. Please also notice the footnote: "*** Red cells indicate samples taken as replacement for unsuccessful baseline blood draws. These were taken even if the CT-Scan results were non-suspicious."

Changes in Supplemental Table 1: "*Taken at the time of suspicious CT scan findings*" replaced by "*Rounds 2 to 5*". The footnote was edited by adding: "*These were taken even if the CT-Scan results were non-suspicious."*

Comment 16: To facilitate interpretation of the results the manuscript needs to make it clear in which screening round the lung cancers were diagnosed, and thus possibly by different criteria? If they were diagnosed in rounds 2 to 5 then they may have been based on new or existing nodules. It is quite unusual to use different diagnostic methods in successive rounds. Also, the average time interval between the suspicious scan and confirmed diagnosis should be stated.

Reply 16: In the Results section we added further details about nodules detected by CT either during the first ("prevalence") screening, or during any of the four subsequent ("incidence") screenings; as well as the median and the range of the time between the suspicious scan and confirmed diagnosis. These statistics are also shown in the new version of Table 2.

Regarding the round of detection: During all screening rounds, the same criteria for lung cancer diagnosis and detection were used, with the difference that newly appearing nodules not observed in earlier annual screenings received specific attention (the schematic overview of the nodule management protocol used in **LUSI Supplemental Table 1** in the revisited version of the manuscript; as requested by reviewer A, further clarifies this). The latter is common procedure in all more recent screening trials, e.g. NELSON, DLCST, DANTE, ITALUNG, MILD. Thus our study provides a representative example of whether the Early®CDT-Lung test has sufficient sensitivity for identifying malignant pulmonary nodules in an equally stage as in CT screening with more recent nodule management protocols.

Please note that, the time between suspicious scan/blood draw and confirmed diagnosis is not very informative, since detection and diagnosis dates differ mostly due to logistics (participants agreeing on undergoing further diagnostics, availability of the facilities, etc), and not due to biological reasons. It is also worth clarifying that the blood sample was always taken at the time of the suspicious CT findings that triggered further diagnostic work-up.

Changes in the text: The following paragraph was added to the results section / supplement (**Page 13, Lines 247-255**): *"Lung cancer detection occurred on the first ("prevalence") screening round for 19 (41.3%) of cases and on subsequent second to fifth ("incidence") rounds for 27 (58.7%) of cases. As described in the nodule management protocol, all tumors detected in the first screening round were deemed suspicious based exclusively on their size. For one of the participants, lung cancer was detected on the second screening round in the absence of pulmonary nodules, due to the identification of atelectasis in the scan images. 21 (80.8%) of the remaining 26 detections in the incidence rounds were done in known nodules already observed in previous screening rounds; with 7 of these immediate recall decisions based solely on nodule volume doubling time (VDT)."*

Table 2 now includes the rows "Round of detection" and "Time between detection and diagnosis". Additionally, the case for which detection was done in the absence of nodules is now shown in the description of "largest diameter".

Comment 17: No sample size or statistical power calculations are presented presumably due to the limited number of cases available. This situation is quite normal, but it does mean that statistical power was not very high for some parts of the analysis Therefore the study can only really be considered preliminary and a larger cohort would be required to make an accurate assessment of the diagnostic performance of the test. For example, the Introduction suggests that the study is focusing on nodules <10mm in size, but only 11 out of the 46 eligible cancers were in this category (Table 2) which is not enough to allow firm conclusions to be made. It needs to be clarified how the study design addresses this particular issue.

Reply 17: Our first focus is on the estimation of lung cancer detection sensitivity, which depends exclusively on the number of cases. While the number of cases in our study may seem low (N = 46), please note that the 95% confidence intervals for our overall sensitivity estimate (13.0% [95% CI: 4.9% - 26.3%]) are reasonably narrow and do not include levels of detection sensitivity (upper limit of only 26.3%) that could be considered useful. This also shows that the sample size of our study (number of CT-detected lung cancer cases) was not so much a limiting factor.

Regarding specificity, the estimates and confidence intervals we obtained are in line with those found in other studies using EarlyCDT-Lung, showing a good performance of the test.

Post-hoc power calculations do not add much further information, over and above confidence limits for our actual study estimates.

Changes in the text: none

Comment 18: Also, it just needs to be clarified why no matching was performed (Table 1). With so many controls to choose from one would normally use a degree of matching, say by age, gender and nodule size.

Reply 18: Matching usually is performed to account for potential confounding factors. In studies for evaluation of diagnostic testing performance, having a random sample of all screening participants (as in our "baseline control group" or alternatively, of all disease-free participants showing indeterminate nodules, as in our "suspicious nodules control group") allows a more straightforward estimation of test specificity. As a matter of fact, methodologic work has shown that using a matched case-control design for the evaluation of a screening test may actually lead to biased estimates of specificity {Janes and Pepe 2008, **PMID:17501939;** Brenner et al. **PMID: 23257151**}. Brenner et al (2013, **PMID: 23257151**) concluded that "*for valid judgment of the specificity of cancer early detection markers, the controls should be representative of cancer-free people from the screening population, who might differ from the cases with respect to sex and age*" and "*Ideally, controls should be a random sample of the cancer-free screening population".*

Changes in the text: none

Comment 19: A previous study to which the manuscript refers (Massion et al) showed in a cohort of 269 nodules that the addition of a positive test result to nodule size significantly increased the PPV for malignancy prediction with a relative risk of 2.7 fold for nodules 4mm to 20mm in the largest diameter. Interestingly, the manuscript summarizes analysis of only High positive results (Table 1), and for these data the relative risk is 2.5 for all nodules, and 2.9 for nodules >=10mm, so well in line with the aforementioned publication. So it should be stated that there is some evidence that this study performed in line with the literature.

Reply 19: In our manuscript, (Table 1) we only summarize "High" level results because, as stated in the Results section, we did not obtain any "Moderate" level results from EarlyCDT-Lung when applied on samples from lung cancer cases.

Regarding the association between test results and malignancy: please note that the study from Massion et al uses data from a cohort design, whereas ours is a nested case-control design. While relative risks can be estimated from cohort designs, the same cannot be done using case-control data since the incidence cannot be estimated. We therefore report odds ratios.

While trying to replicate the RR values reported by the reviewer, we found out that these were calculated by combining the two control groups BC and SNC. If this is the case, these RR cannot be directly compared to those reported by Massion. Massion et al included only subjects who showed at least one non-calcified nodule in the CT scans; whereas our BC group includes subjects with "non-suspicious" findings in their CT scans (78 out of 90), of which 52 showed no nodules at all (this information has been added to Table 1). A more appropriate comparison with the results from Massion et al would be to measure the association between a positive test result and lung cancer when looking at cases and controls with nodules observed on their CT scans. To make such comparison possible, we now included Supplemental Table 3, which shows ORs and positive likelihood ratios calculated amongst cases and controls with suspicious nodules, by nodule size (< 10mm, >= 10mm). There is indeed some level of association; however, it is not statistically significant (all 95% CI include 1), possibly because of the limited sample size of our study.

About the added value of the test on top of nodule size, a way of comparing our results to those of Massion would be to contrast the positive diagnostic likelihood ratio (DLRp, LR+ in our manuscript) reported by Massion et al (DLRp = 2.3 (1.3–4.1) overall, 2.5 (1.1 – 5.4) in nodules 4 to 20 mm) with our estimates shown in Supplemental Table 3. Our point estimates of LR+ (1.92 overall, 1.93 for nodules < 10 mm), are somewhat in line with those from Massion et al, and show some degree of association between a positive test result and presence of lung cancer, though not always statistically significant. Again, the lack of statistical significance might be attributed to small sample size.

Changes in the text: Table 1 was modified and Supplemental Table 3 was added. The following sentences were added to the Discussion **(Page 16 Lines 336 – 340):** *"Regarding the association between a positive test result and malignancy among subjects with suspicious nodules on their CT-scan images, our positive likelihood ratio estimates (1.5 to 1.9 depending on positive test*

definition) are in line with those from Massion et al (LR+: 2.3 (1.3, 4.1), all nodules combined) ***{PMID: 27615397}.*** *Our results, however, were not statistically significant which might be due to the small sample size."*

Comment 20: In the Methods section it is stated that a negative EarlyCDT-Lung test should not affect the clinical management plan. This means that EarlyCDT-Lung is not a "rule-out" test. However, the Abstract conclusion and the Discussion state that the test may be unsuitable as a rule-out test. So the Authors need to explain why they are assessing EarlyCDT as a "rule-out" test when it is in fact a "rule-in" test.

Reply 20: We agree with the reviewer, in that the sentence in the conclusion section of the abstract was misleading. The sentence has been deleted.

The statement that negative EarlyCDT-Lung should not affect the clinical management plan is from the manufacturer's side, and is also stated in Healey et al 2017 [https://doi.org/10.4236/jct.2017.85043. Our findings essentially confirm this.

Changes in the text:
We deleted the following sentence: "*For individuals with small pulmonary nodules, the test sensitivity appears too low to help rule out malignancy, in view of avoiding invasive diagnostic work-up*" in the Conclusions section in the Abstract (**Page 3 Lines 54-56**).

We added the sentence: "*As mentioned in the Methods section, the Early®CDT test is being proposed as a "rule-in" test to identify patients with increased risk of having a pulmonary malignancy, whereas a negative Early®CDT test should not affect the clinical management plan; that is, Early®CDT is not meant to be used as a "rule-out" test. The low sensitivity of the test as estimated from our data confirms the recommendations from the providers*." (**Page 16 Lines 325-330**)

Comment 21: Additionally, to assess a rule-in test the positive predictive value (PPV) normally needs to be considered but there is no mention of PPV in the manuscript.

Reply 21: We agree with the reviewer in that additional statistics are helpful when evaluating a test (in particular a "rule-in" test). However, we consider reporting the positive likelihood ratio to be more appropriate and more informative.

Although the positive predictive value is widely used in clinical practice, it has some disadvantages. The PPV is in itself not an invariant property of the test, since it depends on the prevalence of the disease in the population tested **(Sedighi 2013; PMID: 24910762).** This in turn means that the PPV cannot be transferred to or compared with other populations or settings. Furthermore, its dependency on the prevalence makes it impossible to estimate based on a case-control study like ours, since the observed prevalence only reflects our sampling and it is not

representative of the population **(Kohn 2013; PMID: 24238322**). The alternative of using estimates of lung cancer prevalence in the population eligible for screening from either external sources or from the LUSI trial would make the estimated PPV strongly dependent of the selected data (i.e. prevalence in the first screening round vs follow-up rounds).

Based on the arguments explained above, we decided to report estimates of the positive diagnostic likelihood ratio (LR+) for each of the two control sets, and additionally contrasting lung cancer patients vs controls with suspicious nodules, as shown in Supplemental Table 3. The LR+ shows how much more likely someone is to get a positive test in the presence of the disease, compared to a person without disease. The LR+ has the advantage that it can be directly derived from the estimated sensitivity and specificity without the need of estimates of prevalence and can therefore be compared in various populations and settings.

Changes in the text: *"Based on these estimates, the LR+ was then calculated at 1.17 [0.46, 3.03] in the BC group and at 1.47 [0.54, 3.98] in the SNC group.*" added to the Results section, (**Page 13 Lines 265-267**); "*specificity of 96.7% [90.6%, 99.3%]) in the BC group and 4.4% (specificity of 95.6% [89.0%, 98.8%]) in the SNC group, yielding estimates of LR+ of 3.91 [1.03, 14.94] and 2.93 [0.87, 9.88] respectively*" (**Results section Page 14 Lines 275-277**).

"*… in comparison to the BC group (OR: 4.35 [1.04, 18.28], p=0.04) and not enough evidence of association in the SNC group (OR: 3.22 [0.86, 12.07], p=0.08).*

*Among lung cancer patients showing nodules on their CT-Scans (N=45) (Table 2) and the SNC group, we couldn't find enough evidence between test results and malignancy, with OR of 1.58 [0.51, 4.86] and 3.31 [0.88, 12.39] depending on the definition of a positive test (Supplemental Table 3). Similar results were obtained when stratifying by nodule size (<10 mm, >= 10 mm). Regarding positive test results and malignancy, there was only weak evidence of association when considering only "High Level" results as significant with LR+: 1.92 [1.09, 3.40] overall, and 1.17 [1.02, 1.35] among participants with nodules >= 10 mm, but not among those with nodules < 10 mm (OR: 1.93 [0.24, 15.77]." (***Page 14 Lines 278-285***)*

Additionally the sentence: "*Regarding the association between a positive test result and malignancy among subjects with suspicious nodules on their CT-scan images, our positive likelihood ratio estimates (1.5 to 1.9 depending on positive test definition) are in line with those from Massion et al (LR+: 2.3 (1.3, 4.1), all nodules combined)* ***[12]****. Our results, however, were not statistically significant which might be attributed to the small sample size."* was added to the Discussion (**Page 16, Lines 336-340**)

Comment 22: Finally, it may well be that the marker test has reduced sensitivity for smaller nodules, but this study is insufficiently powered to confirm that. This should be stated in the Discussion as a limitation of the study.

Reply 22: Considering "smaller nodules" as in the introduction section of our paper ("*small malignant nodules (< 10 mm in diameter*)", the results are: 1 high level test result vs 10 NS test results, for an estimated sensitivity of 9.1% [0.23%, 41.3%]. In nodules *>= 10 mm in diameter*, the estimated sensitivity was 14.7% [4.9%, 31.1%]. While we agree with the reviewer that analyses and estimates for nodule of different sizes (subcategories) are somewhat hampered by more limited CT-detected case numbers, it should be noted that, still, confidence intervals for sensitivity of the Early®CDT test do not cover a level of sensitivity that would be considered sufficient even when we focus more specifically on nodules >=10mm.
The misleading sentence in the abstract referring to the low sensitivity of the test in small nodules was removed.

Changes in the text: The sentences "*Within the subset of participants with nodules <10 mm in diameter, the test produced "High Level" results for 1 out of 11 CT-detected lung cancer patients, yielding a sensitivity of 9.1% [0.23%, 41.3%]. For participants with nodules >= 10mm, the estimated sensitivity was 14.7% [4.9%, 31.1%]*." were added to the Results section (**Page 13, Lines 259-262**)

We deleted the following lines "*For individuals with small pulmonary nodules, the test sensitivity appears too low to help rule out malignancy, in view of avoiding invasive diagnostic work-up*" in the Conclusions section in the Abstract (**Page 3 Lines 54-56).**

Comment 23: The conclusion in the manuscript seems to show a commercial product in a poor light. Therefore the authors should contact the Manufacturer of the product to ensure all procedures were carried out correctly.

Reply 23: We were hoping to have more positive results for the Early®CDT test, as this might have opened up avenues for further screening trials also in Germany. We have actually had several contacts by e-mail as well as a telephone conference with four employees of Oncimmune to seek scientific cooperation. Amongst others, we requested help in understanding better the exact individual TAAb marker thresholds used and the weighting of TAAb-positive scores in the test, but Oncimmune did not help us further.

Changes in the text: none.