Author's response to reviews

Title: Exploring the Uncertainties of Early Detection Results: Model-Based Interpretation of Mayo Lung Project

Authors:

Lu Shi (lushi.pku@gmail.com)
Haijun Tian (tianhajjun@gmail.com)
Barbara Berman (bberman@ucla.edu)
William J. McCarthy (wmccarth@ucla.edu)
Shinyi Wu (shinyiwu@usc.edu)
Robert Boer (fietsrob@gmail.com)

Version: 2 Date: 10 September 2010

Author's response to reviews: see over
Dear Editor,

In this document, I wrote item-by-item description of the revisions we made in responding to the reviewer comments. Please let me know if anything is still missing.

Regards,

Lu
Major Criticisms: The Marcus paper infers that the majority of excess screen-detected cancers represent over-diagnosis on the basis of long-term follow-up data results. This paper incorporates only the initial six years of data as a contribution to the deviance measure. It is unclear whether the initial trial data is alone sufficient to distinguish between the hypotheses under question.

It will be a worthwhile effort to model the long-term follow-up results as well. However, it is not very plausible that the intervention group would have experienced significantly more over-diagnoses than the control group after the screening intervention. 1. Not looking at mortality 2. Incidence data (lung cancer mortality remains higher, a random noise is thus unlikely)

The authors fail to mention the possibility that a risk difference between the two trial arms was accrued during the trial itself as a result of screening. Therefore, the list of alternative hypotheses to over-diagnosis given is not an exhaustive list of possibilities. A recent paper by Goldwasser et al. (2010) in the journal Cancer proposes that a risk difference may be attributable to the excess radiation risk received during the trial. This hypothesis was consistent with individual screen histories and the long-term mortality follow-up data. This paper addresses many of the same issues but is not in the list of citations.

We have added the Goldwasser article to the references as per the reviewer’s comment.

Incorporating over-diagnosis by the addition of a parameter for the detection of indolent cancers should improve the fit to the overall data-set when mortality is considered in addition to incidence. The fact that it did not do so in the model suggests a failure in the methodology and in particular, the choice of the deviance measure given. The baseline/simple model assumes perfect sensitivity. Given that the sensitivity of chest x-ray is far less than 100%, in order to achieve the goodness of fit demonstrated in table 2, the underlying natural history model must be unrealistic. Therefore, it may not be reasonable to expect that the simple model serves as a valid reference point against which to examine a distinct natural history model that incorporates over-diagnosis.

The reviewer here took it for granted that a significant number of indolent cancers did exist, while the very existence of indolent cancers under chest X-rays has not been verified.

Our model fitting here does not calibrate to the mortality outcome. Thus the reviewer’s comment would be more relevant for a calibration to mortality outcomes.

In the methods section, it would be useful to have a mathematical definition for
the deviance in the text as well as a measure of statistical significance for the deviance reduction, in particular for the risk difference model compared to the simple model.

*We add the mathematical definition for the deviance measure and describe the significance test in our method section. In the Results section we describe how we test the difference between the risk difference and the simple model.*

The authors conclude that the trial results are best explained by an initial risk difference between the study arms, in other words, a failure in randomization at the onset of the trial. The probability of randomization failure resulting in a mortality difference as large as that observed can itself be quantified in the context of simulation.

*In the first paragraph of the Discussion section, we have now included an estimate of the probability that the observed number of lung cancers in the screening intervention arm is the same as the model-expected number of lung cancers in the screening intervention arm.*

The authors report that individuals in the screening arm smoked on average half a cigarette more per day than individuals in the control arm. This finding is supportive of the authors’ suggestion that a risk difference at the onset of the trial may explain the trial results. It would be useful to look at the distribution of cigarettes per day, in particular as a cross-section with age in order to examine whether this difference was sufficient to explain a risk difference as large as the one observed in the MLP.

*The difference in smoking intensity is not necessarily a risk difference at the onset of the trial. In another paper currently under review, we track the smoking intensity in the two trial arms over the six years and find different temporal patterns. However, as we double-checked the distribution of cigarettes per day in that study, the magnitude of the difference (1/2 cigarette per day) is not big enough to explain the observed difference in lung cancer diagnosis in MLP.*

Minor Criticisms: There are several typos in the tables and some grammatical errors. In the introduction, it is implied that there is three years of data, when the initial trial lasted for six years with one additional year of follow-up.

*Fixed.*
Reviewer: Antonio Mutti

Reviewer's report:
The manuscript "Exploring the Uncertainties of Early Detection Results: Model-Based Interpretation of Mayo Lung Project" by Lu Shi et al. is a re-assessment of the Mayo Lung Project (MLP), a randomized controlled clinical trial of lung cancer screening among male smokers, exploring additional statistical models. The manuscript is interesting and acceptable for publication. However, the manuscript is very difficult to read and the message difficult to grasp by the average reader. Moreover, the performed simulation did not take into account other possible models, Sensitivity & Risk difference, Systematic error & Risk difference and finally Sensitivity & Systematic error & Risk difference. Authors should give an explanation about it. Even if the conclusions of the study are convincing, the parameters indolent cancer and risk difference were always considered as single factors, even if indolent cancer parameter is probably less interesting than risk difference.

We agree and we have incorporated this reviewer's comment in our discussion.
Reviewer's report
Title: Exploring the Uncertainties of Early Detection Results: Model-Based Interpretation of Mayo Lung Project
Version: 1 Date: 12 July 2010
Reviewer: Oscar M Rueda

Reviewer's report:
Shi et al. present a simulation model to explore the results of the Mayo Lung Project for lung cancer screening. The manuscript tries to explain why the results of that famous trial did not show a reduction in lung cancer mortality when a screening strategy was applied to a randomized group.

There are several theories about this that the authors try to address with the aid of MISCAN, a microsimulation model of early detection of chronic diseases adapted to lung cancer.

Major Compulsory revisions:

1. I would like the authors to summarize a little more MISCAN's assumptions. Figure I is not very clear and has no explanation. What is being represented in vertical and horizontal axes? Time? Evolution of the disease?

   We have added an explanation to Figure 1’s axes and have elaborated the title to make it more self-explanatory.

There is some information scattered through additional file number 4 and a reader guide referenced, but the manuscript would be much improved if a good summary of the model would be provided, because the assumptions (how do they model the times and probabilities of going through the different stages, etc.) can be relevant to the results.

   We have added more details about assumptions, e.g. the rationale for distribution assumptions and parameter calibration.

2. Related to that, I would like the authors to comment on the limitations of their model to detect pitfalls in the trial.

   At the end of the Discussion section we discussed the limits of the model in relation to the over-diagnosis/indolent cancer issue

Minor essential revisions:

3. I would include some of the results of the table into the manuscript, especially the information relevant to the fit of the different models.

   We have added the relevant statistics to the text in the method section.
My overall impression is that manuscript is interesting, because shows how simulation can help in the interpretation of complex biological processes and in this case, the results of clinical trials. But it the paper must state very clearly how good the simulated model reflects the biological process and if it takes any assumption that might benefit any of the explanations that are being tested.

As mentioned above, at the end of the Discussion section we discussed the limits of the model in relation to the over-diagnosis/indolent cancer issue.
Reviewer: Ramon Diaz-Uriarte

Reviewer's report:
This paper might have provided an interesting study, complementing previous ones that exist along similar lines. But it suffers from a lack of clarity in explaining the context and a lack of details in the methods.

Major compulsory revisions

1. First of all, the authors say, without much explanation, that "(...) there was no reduction in lung cancer mortality among subjects in the intervention arms as compared to subjects in the control arm, but there was an increase in lung cancer survival" (p. 3). And in the abstract we are told "(...) demonstrated an increase in lung cancer survival, but no reduction in lung cancer mortality". Anyone not familiar with the Mayo Lung Project (MLP) will be left wondering what gives here. These statements seem to say "A and not-A".

Since understanding the key findings of the MLP study is crucial to understanding this paper, this all needs to be explained more clearly. Summarizing quickly from the literature, we can say (for the sake of brevity I do not include details about the follow up periods to which these date refer, etc):

- There is no difference in lung cancer mortality between the screened and the control groups (in fact, there is a slight, but not statistically significant, larger lung cancer mortality in the screened group). The definition of what "lung cancer mortality" is can be found very clearly in, for instance, Marcus et al., 2000, "Lung Cancer Mortality in the Mayo Lung Project ...

- There is a survival advantage for those diagnosed with cancer in the screened group compared to those diagnosed with cancer in the control group. This is explained in detail, for example, in Marcus et al., 2000, "Lung Cancer Mortality in the Mayo Lung Project ...". JNCI and Flehinger et al., 1993, "Screening for Lung Cancer", Cancer.

- There is a larger incidence of cancer (i.e., number of cancer cases detected) in the screened group (all references say this explicitly).

- And, finally, if I remember correctly, there was a survival advantage, just plain, actuarial 5-year survival, for those in the screened group (Strauss et al., 1997, "Screening for Lung Cancer : Another Look", Chest).

Those papers, and others, show that we are not saying "A and not-A", since we are measuring different things. But this is not at all obvious from the current ms. Those papers, as well, give possible explanations of the results.

In fact, when the authors discuss the "key arguments from a modeling perspective", I'd say that they need to do a much more thorough job not just in terms of modeling perspective, but just in terms of the plain explanations of what might be happening. This
reviewer finds some of the explanations of some of the above papers to be much clearer than the ones in the present ms. In addition, when exploring the key arguments, more detailed review of previously published papers should be provided, emphasizing the pros- and cons- of the different arguments and what the actual consensus (if any) is regarding each argument.

We have reworded the sentences and have now included discussion of more previous papers as per the reviewer’s suggestion.

2. I would have liked to see more details about the MLP. For instance, I think that the control group participants received an annual set of screening tests, because that was the standard practice of the Mayo clinic in the 70s, but this is not said clearly in the current ms.

No, only half the unscreened group received an annual chest x-ray examination during the study and that was considered as contamination already.

3. Page 4, second line, says "(...) morbid lesions or so indolent that patients". There is some type there but, more importantly, the idea of the indolent problem is not explained in this paper until a few pages later (p. 6, third paragraph). And this is a key idea.

We have moved discussion of the idea of indolent cancers to Page 4 as per the reviewer’s suggestion.

4. Methods: we are given no references nor reasons for many of the modeling choices. This criticism applies to every single paragraph of pages 5 and 6. I am not questioning the choices. But I would like to know why.

We added the explanation of the statistical tests and some additional references about the distribution assumption of the cancer stages. Other than that, a full explanation of the MISCAN-lung is available through the referenced reader guide.

5. The first paragraph of the discussion is confusing. It is unclear whether you trying to explain the results of the MLP, or exploring the consequences of more sensitive screening. I understand that, ultimately, the objective is to explore the consequences of more sensitive screening, and you are using your model of the MLP to approach that question. But then, please say so. Start by explaining what your results say about the MLP, and then go into the bigger question.

We reworded the opening sentence of the paragraph to make the flow follow the reviewer’s suggestion.

6. Discussion: and how do your results compare to those from other similar exercises, including modeling ones? (See some of the references above).
We have added some discussion under the context of previous exercises.

7. Tables are currently part of the supplementary material, but they should be in the text, given their importance for the reported results. It is there where the parameters are shown, and the results reported. And, please, do not place suppl. mat. as word files when you can as easily provide a pdf.

We will discuss this with the editor about where to place these tables, as this is more of an editorial policy rather than a writing style. Our personal preference would be to include the tables in the text, as the reviewer has suggested, but we defer to the editor's decision on this point.