PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Treatment delay affects clinical severity of tuberculosis; A longitudinal cohort study
AUTHORS	Virenfeldt, Jakob; Rudolf, Frauke; Camara, Cidia; FURTADO, Alcino; Gomes, Victor; Aaby, Peter; Petersen, Eskild; Wejse, Christian

VERSION 1 - REVIEW

REVIEWER	Chandrashekhar T Sreeramareddy Faculty of Medicine and Health Sciences University Tunku Abdul Rahman Bandar Sungai long Selangor, Malaysia
REVIEW RETURNED	27-Feb-2014

GENERAL COMMENTS	There are some more recent literature that could be of importance to cite within this article.

REVIEWER	David Dowdy Johns Hopkins Bloomberg School of Public Health, USA
REVIEW RETURNED	17-Mar-2014

GENERAL COMMENTS	This manuscript describes an analysis of 973 TB patients in Guinea-Bissau, comparing clinical severity at presentation to self-reported duration of symptoms. The primary conclusions are that treatment delay is decreasing over time in Bissau (a nice result for the TB program in Bissau, but not necessarily a piece of generalizable knowledge) and that long treatment delay is associated with more severe clinical presentation. While this latter finding is not particularly novel, the large sample size, duration of follow-up and population-based sample are favorable qualities. Major comments: 1. The primary conclusion seems to be based upon a test for trend (done with what test? an ordinal regression?), shown in Table 3. If this is to be the main conclusion of the manuscript, I would recommend that the authors explore this in a little more detail than simply proportion with TBScore >=8 in each quartile of the population. What is the actual difference in TBScore? What degree of symptomatology does this difference correspond to? Is there evidence of a threshold effect (in the simple analysis provided, it seems as if there might be, around the median value)? Why not perform a logistic regression with "severe TB" as the outcome and

	"treatment delay" as the relevant exposure, thus using data on the specific timing of treatment delay and adjusting for other potential confounders (e.g., calendar time)? The authors have a large dataset at their fingertips; collapsing that dataset into a simple analysis of "does the proportion of patients with TBScore >=8 change with quartile of treatment delay" seems to waste a lot of available data. 2. I would like to see a little more information on how patients were asked about their symptom duration. Is this a single question asked of all TB patients, and if so, is it asked at the time of TB diagnosis, initiation of therapy, etc.? I find it hard to believe, for example, that a majority of patients developed cough, fever, and chest pain simultaneously (as implied by Table 1). The fact that these are retrospectively collected data is a known and acceptable limitation, but it is important to know just how much of a limitation this is. I would also suggest to the authors that they not conflate "self-reported duration of symptoms" with "treatment delay," for example in the title and abstract of the manuscript. Patients with more severe symptoms are quite likely to recall a longer duration of symptoms (i.e., recall bias), and unless "treatment delay" means something other than "self-reported duration of symptoms," the authors should not imply that patient recall necessarily reflects reality. Minor comments, in order of appearance in the manuscript: 3. Statistical analysis: The authors state that, "to account for outliers in the reported treatment delay medians were chosen and logarithmic transformation was performed." What is meant by the term "medians were chosen" in this sentence? Is the regression not performed on individual values of treatment delay as the dependent variable - how is a median value used? 4. In the statistical analysis, please provide the method of testing for trend, as there are many such "trend tests." 5. 364 patients "did not show up for enrollment." Does this mean that these patients were not treated, or that they were treated but not enrolled in the research portion of the study? 6. In the results section under "risk factors for treatment delay," it seems unnecessary to recap all of the statistically significant univariate results in the table. I find the magnitude of the differences in the table to be much more informative, in any case. 7. The authors state that "a delay above 21 weeks had a 59% significantly higher mortality than quartile 1" - this misinterprets a hazard ratio as an incidence ratio, and furthermore, only the unadjusted results are presented. The adjusted HR should also be included, rather than simply stating that it was not significant. 8. The authors present the change in TBScore for smear-negative patients (with a quote that p<0.00, presumably a typo). Can they present the corresponding change for smear-positive patients? And was TBScore (as used for the primary analysis) universally measured at treatment initiation? 9. In the section on changes in treatment delay, a figure showing mean or median treatment delay over time would be extremely helpful in assessing whether delay was truly linear over time. It appears that most of this decline occurred over a two-year period

	during which a "TB incidence study" was undertaken. Further details of this "incidence study" would be helpful -- if the authors mean a TB prevalence survey, then I would argue that the finding of shorter time to treatment during a period of active case finding is not a scientifically relevant result and should be removed from the manuscript. 10. Given the important trend in treatment delay over time, why have the authors not adjusted for calendar time in most of their analyses? It seems likely that calendar time (during which TB services may have improved in Bissau) could be an important confounder. 11. The discussion should start with a description of the strength of association between treatment delay and severity - how much more severe were the cases who were delayed? What are the likely clinical implications of shortening treatment delay? 12. I find it challenging to appropriately interpret Figure 2, including the confidence bands - the problem being that the authors only included patients who ultimately presented for treatment. For all they know, a substantial number of TB patients remain untreated at 60 weeks from symptom duration. I would find these data more palatable in a table describing the median (and other percentiles of) symptom duration, rather than in a figure that implicitly suggests that, after 1 year, all TB patients with symptoms are under treatment. The authors did not perform a prospective study of people from the time they initiated treatment. 13. In Figure 3, the captions "short delay," "moderate delay," "long delay," and "very long delay" should be replaced with actual numbers of self-reported symptom duration. Furthermore, if anything, this analysis shows that there is no mortality difference during the first year - i.e., almost certainly no mortality difference due to TB. This seems much more likely to reflect that patients with underlying illnesses (e.g., HIV) are more likely to present with severe TB, and then more likely - as a result of their underlying illness - to die after completing TB therapy. But this increase in mortality risk should not be attributed to TB, especially if the proportions dying while on TB therapy are similar (as they are - no difference in the first 6 months!). 14. In Table 2, the use of the double asterisk, with the footnote to "95% confidence interval", seems non-standard to me. Seeing these asterisks makes me think of statistical significance. Better to simply title the column header "Adjusted RR (95% CI)".

REVIEWER	Ying Li Third Military Medical University, Chongqing, China
REVIEW RETURNED	24-Mar-2014

GENERAL COMMENTS	1. In the section of introduction, the author used TB epidemiology data of 2010, but WHO 2013 TB report had new data 2. Methods: why not ulitilize the same criteria (such WHO regarding inclusion criteria, clinic criteria) to diagnose PTB? 3. methods: Bandim TBscore was used to measure clinical severity at diagnosis, but in the section of Abstract, secondary outcomes included TB morbidity. What is the TB morbidity used to measure? 4. Page 6, line7, "all patients diagnosed with TB" should be all

	pulmonary TB patients? 5. As for the "non-residents" 6.Table II, head of the table, it is best to use "RR(95%CI) "and "Adjusted RR(95%CI)" in stead of the note"**95% Confidence Intervals" 7.Page10, line 37-38," ....than patients with a higher educational level (9.7 weeks; 10.0 weeks)...." is not clear. This study had a long perspective follow-up data, the data should be very interesting and important. however, the author need perferct the data presentation. Methods: 1. study palce selection, followup process, outcome and data collections methods should be described more clearly. Discussion, it should describe the assoication between treatment delay anad clinic severity and its possible explanation for this association.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
R1.1: There are some more recent literature that could be of importance to cite within this article.
C1.1: We appreciate the reminder of recent publications. We have now included newer literature in the manuscript (ref 19, 20, 25, 27 and 31)

Reviewer: 2
R2.1: The primary conclusion seems to be based upon a test for trend (done with what test? an ordinal regression?), shown in Table 3.
If this is to be the main conclusion of the manuscript, I would recommend that the authors explore this in a little more detail than simply proportion with TBScore >=8 in each quartile of the population. What is the actual difference in TBScore? What degree of symptomatology does this difference correspond to? Is there evidence of a threshold effect (in the simple analysis provided, it seems as if there might be, around the median value)? Why not perform a logistic regression with "severe TB" as the outcome and "treatment delay" as the relevant exposure, thus using data on the specific timing of treatment delay and adjusting for other potential confounders (e.g., calendar time)? The authors have a large dataset at their fingertips; collapsing that dataset into a simple analysis of "does the proportion of patients with TBScore >=8 change with quartile of treatment delay" seems to waste a lot of available data.
C2.1: We appreciate the opportunity to elaborate on the issue of clinical severity and delay. According to the validation studies of the severity index (ref 18), the TBscore=8 is the best cutoff for TB severity (the top quarter in the initial validation dataset) and has strong prognostic capacity for mortality. We have clarified this in the methods section on p 8, line 161-3. We have also added the numerical differences in TBscore in means in Table 3. We also appreciate the suggestion of using logistic regression as promising and a good supplementary way to describe our findings. This has now been included in our results section on p 12, line 231-236. The test for trend was an inappropriate label since the test performed is a Fishers exact test. This has now been revised and the implications of the association has been described in the results section p. 11, line: 227-228.

R2.2: I would like to see a little more information on how patients were asked about their symptom duration. Is this a single question asked of all TB patients, and if so, is it asked at the time of TB diagnosis, initiation of therapy, etc.? I find it hard to believe, for example, that a majority of patients developed cough, fever, and chest pain simultaneously (as implied by Table 1). The fact that these are retrospectively collected data is a known and acceptable limitation, but it is important to know just

how much of a limitation this is. I would also suggest to the authors that they not conflate "self-reported duration of symptoms" with "treatment delay," for example in the title and abstract of the manuscript. Patients with more severe symptoms are quite likely to recall a longer duration of symptoms (i.e., recall bias), and unless "treatment delay" means something other than "self-reported duration of symptoms," the authors should not imply that patient recall necessarily reflects reality.
C2.2: We thank the reviewer for highlighting this important caveat in our data collection. We have attempted to clarify how the information on delay was obtained on p.7, line 133-136. The symptoms listed in Table 1 were symptoms indicated by the patients as initial symptoms, and no attempt was made to differentiate which was first. We have now clarified that on p 10, line 203-6. We appreciate the reviewers point on self-reported duration of symptoms not being equal to delay, and we have clarified that as a limitation on p 14, line 309-311 under limitations. Yet, we have chosen not to substitute “delay” with “self-reported duration of symptoms” in title and abstract, because we hold that is a generally accepted norm that delay is based on the patient’s recall of duration of symptoms, and there is no other way of measuring this variable.

R2.3: Statistical analysis: The authors state that, "to account for outliers in the reported treatment delay medians were chosen and logarithmic transformation was performed." What is meant by the term "medians were chosen" in this sentence? Is the regression not performed on individual values of treatment delay as the dependent variable - how is a median value used?
C2.3: This sentence has now been revised in the manuscript p. 8, line 167-168. The statement concerned the choice of medians in the presentation of data (table 1 etc) instead of means because of the outliers in the dataset. The regression was performed on individual values of treatment delay as the dependent variable.

R2.4. In the statistical analysis, please provide the method of testing for trend, as there are many such "trend tests."
C2.4: The mentioned trend test is no longer present in the manuscript. See C2.1.

R2.5: 364 patients "did not show up for enrollment." Does this mean that these patients were not treated, or that they were treated but not enrolled in the research portion of the study?
C2.5. The patients not enrolled received treatment at the local health facility, but were not included in the epidemiological study, so they did not undergo the clinical evaluation and had the questionnaire completed, hence we do not have data on them. This is now clarified on p 6, line 120-123.

R2.6. In the results section under "risk factors for treatment delay," it seems unnecessary to recap all of the statistically significant univariate results in the table. I find the magnitude of the differences in the table to be much more informative, in any case.
C2.6: We thank the reviewer for noting redundancy in the manuscript; we have removed this part from the Results section, p. 11 line 211-13.

R2.7: The authors state that "a delay above 21 weeks had a 59% significantly higher mortality than quartile 1" - this misinterprets a hazard ratio as an incidence ratio, and furthermore, only the unadjusted results are presented. The adjusted HR should also be included, rather than simply stating that it was not significant.
C2.7: This misinterpretation has now been removed from the manuscript. The adjusted HR rate is now included p 12, line 238-241.

R2.8: The authors present the change in TBScore for smear-negative patients (with a quote that $p<0.00$, presumably a typo). Can they present the corresponding change for smear-positive patients? And was TBScore (as used for the primary analysis) universally measured at treatment initiation?
C2.8: The mentioned typo has been deleted from the manuscript. The TBscore was universally measured between subgroups at inclusion in the study. The corresponding change for smear-positive

patients in TBscore is now presented on p. 12, line 245-247.

R2.9: In the section on changes in treatment delay, a figure showing mean or median treatment delay over time would be extremely helpful in assessing whether delay was truly linear over time. It appears that most of this decline occurred over a two-year period during which a "TB incidence study" was undertaken. Further details of this "incidence study" would be helpful -- if the authors mean a TB prevalence survey, then I would argue that the finding of shorter time to treatment during a period of active case finding is not a scientifically relevant result and should be removed from the manuscript
C2.9: We have added a figure demonstrating log delay over time (Figure 4). We thank the reviewer for highlighting an inappropriate description, we have corrected an error regarding the prevalence survey which was labeled as an incidence study on p 15, line 297-299, we have removed this speculating part from the Result section and in the Discussion the survey is now described in more detail (ref 29) as well as the possible effects on p 15, line 297-303. This survey affected only a minor part of the population in the study area (aprox 5% of the adults), and was therefore not a period of active case finding in the entire study area.

R2.10: Given the important trend in treatment delay over time, why have the authors not adjusted for calendar time in most of their analyses? It seems likely that calendar time (during which TB services may have improved in Bissau) could be an important confounder.
C2.10: The point raised is very relevant and has now been taken into the risk factor analysis (Table 2). Correcting for calendar time has also been addressed in the severity analysis see C2.1.

R2.11: The discussion should start with a description of the strength of association between treatment delay and severity - how much more severe were the cases who were delayed? What are the likely clinical implications of shortening treatment delay?
C2.11: We have now added a section in the Discussion on p 14, line 263-267 on the association between delay and severity.

R2.12: I find it challenging to appropriately interpret Figure 2, including the confidence bands - the problem being that the authors only included patients who ultimately presented for treatment. For all they know, a substantial number of TB patients remain untreated at 60 weeks from symptom duration. I would find these data more palatable in a table describing the median (and other percentiles of) symptom duration, rather than in a figure that implicitly suggests that, after 1 year, all TB patients with symptoms are under treatment. The authors did not perform a prospective study of people from the time they initiated treatment.
C2.12: We thank the reviewer for raising this point and see that it is a possible misunderstanding of Fig. 2. Yet, although we appreciate the point raised, we hold that possible misinterpretations of the figure can be avoided by a more detailed description of the analysis in the text and a clearer figure legend. We have therefore chosen not to omit the figure, we hold that it should be quite clear from the study description, that this is a study on notified TB patients on treatment, not a prospective follow-up of TB suspects, and to the best of our knowledge we do not imply anywhere that we have information on all the undiagnosed and diagnosed untreated cases. Our group has recently published a prospective study on TB suspects which we now refer to in the clarification inserted on p 10, line 198-200 and in the Discussion p 16, line 320-323.
Similar figures of treatment delay have been used in delay studies earlier (e.g. ref 11).

R2.13: In Figure 3, the captions "short delay," "moderate delay," "long delay," and "very long delay" should be replaced with actual numbers of self-reported symptom duration. Furthermore, if anything, this analysis shows that there is no mortality difference during the first year - i.e., almost certainly no mortality difference due to TB. This seems much more likely to reflect that patients with underlying illnesses (e.g., HIV) are more likely to present with severe TB, and then more likely - as a result of their underlying illness - to die after completing TB therapy. But this increase in mortality risk should

not be attributed to TB, especially if the proportions dying while on TB therapy are similar (as they are - no difference in the first 6 months!).
C2.13: We recognize the value of further discussing the results and implications of Figure 3. The analysis does not show any significant differences between the different groups of delay, which we agree can indicate that other factors (HIV, TB severity, age etc.) are more important in terms of mortality. It is of notice that there is no comparison to the background population and that the figure simply describes the differences in mortality of the strata's of delay p. 14, line 271-73.
The actual numbers of the treatment delay in each group is now included in the legend of Figure 3.

R2.14: In Table 2, the use of the double asterisk, with the footnote to "95% confidence interval", seems non-standard to me. Seeing these asterisks makes me think of statistical significance. Better to simply title the column header "Adjusted RR (95% CI)".
C2.14: This suggestion has now been implemented in Table 2.

Reviewer: 3
R3.1: In the section of introduction, the author used TB epidemiology data of 2010, but WHO 2013 TB report had new data.
C3.1: We thank the reviewer for mentioning such data. It has now been implemented in the manuscript, p. 5, line 81-83.

R3.2: Methods: why not ulitilize the same criteria (such WHO regarding inclusion criteria, clinic criteria) to diagnose PTB?
C3.2: We have used exactly the WHO criteria as referred to in Ref 13 for diagnosing TB.

R3.3: methods: Bandim TBscore was used to measure clinical severity at diagnosis, but in the section of Abstract, secondary outcomes included TB morbidity. What is the TB morbidity used to measure?
C3.3: Bandim TBscore is used as "pseudo"measure for TB-morbidity. This is now clarified in the abstract.

R3.4: Page 6, line7, "all patients diagnosed with TB" should be all pulmonary TB patients?
C3.4: We included all types of TB in the cohort but excluded extra-pulmonary TB in this study due to a low number of patients in this group and presumably considerable differences between the two groups.

R3.5: As for the "non-residents"
C3.5: Typo?

R3.6: Table II, head of the table, it is best to use "RR(95%CI) "and "Adjusted RR(95%CI)" in stead of the note"**95% Confidence Intervals"
C3.6: This has been revised in the manuscript (see Table 2).

R3.7: Page10, line 37-38," ....than patients with a higher educational level (9.7 weeks; 10.0 weeks)...." is not clear.
C3.7: This has been changed p. 11, line 215.

R3.8: This study had a long perspective follow-up data, the data should be very interesting and important. however, the author need perferct the data presentation.
C3.8: We have now included one more figure (4) to explain the decrease in treatment delay during the study period. The results section has been revised in an attempt to make the understanding of the main outcomes of this study more presentable.

R3.9: Methods: 1. study palce selection, followup process, outcome and data collections

methods should be described more clearly.
C3.9: We thank the reviewer for these general comments on the methods section. The data collection methods and follow up process has now been described in further detail p. 7, line 124-139. Concerning the study place we find that the setting is well described on p. 6, line 108-117. The main outcome is TB-morbidity assessed by Bandim TBscore, which has been extensively described in former studies. We find that the description on p.8, line 155-163 is sufficient for the understanding of this study.

R3.10: Discussion, it should describe the assoication between treatment delay anad clinic severity and its possible explanation for this association.
C3.10: We have now added a description of clinical implications of treatment delay in the Discussion, p 14, line 261-70.

VERSION 2 – REVIEW

REVIEWER	David Dowdy Johns Hopkins Bloomberg School of Public Health, USA
REVIEW RETURNED	27-Apr-2014

GENERAL COMMENTS	I would like to thank the authors for their careful revisions on this resubmission, which I do believe have improved the manuscript substantially. I still have a few concerns, however, numbered according to their number in my original review. Those that are not listed have been appropriately resolved by the authors. 1a. Fisher's exact test only tests for whether there is any difference, comparing any of the four strata of delay. (Thus, for example, if the percentage of severe TB cases was 20% in the lowest quartile, 20% in the 3rd quartile, and 20% in the 4th quartile, but 80% in the 2nd quartile, the p-value for Fisher's exact test over the entire dataset would be significant.) Can the authors consider statistical tests that might more accurately support their primary conclusions? Fisher's exact tests comparing each stratum to the baseline, or to the stratum immediately underneath, might be one way of doing this (though there are many others). 1b. I appreciate the use of TBscore >8 as a cutoff, but could the authors give some description - either in the methods or the discussion - of how severe this is? For example, describing a TB patient with a TBscore of 8 might be helpful to readers (like myself) who are not intimately familiar with this score. 1c. Thank you for adding the logistic regression to the Results. Could you also please adjust the adjusted results for age (at least >/<45 years old), as this seems to be an obvious and potentially strong confounder? 2. I also appreciate the authors inclusion of additional text to clarify their definition of "treatment delay," and I will not press this point if the authors disagree (i.e., OK for the authors not to reply to this comment on re-revision if they choose not to), but "treatment delay" means something very different to me than self-reported symptom duration. I think of "treatment delay" as having 2 components: patient delay and health system delay, and patient delay may not necessarily begin from the very onset of symptoms but perhaps from the time at which those symptoms became sufficiently bothersome to the patient to merit consideration of care-seeking. (For example, if

	a patient with asthma has been coughing for 20 years prior to TB diagnosis, this would not equate to a treatment delay of 20 years to me.) Again, I am willing to trust the authors as to what they believe to be the best way to resolve this issue -- but I would still contend that symptom duration and treatment delay are not necessarily the same thing. 9. Thank you for adding Figure 4. I might recommend simply showing a bar graph, with each bar corresponding to the mean or median delay over a 6- or 12-month period, rather than a series of dots that are very difficult for the average reader to interpret. I would also recommend using a non-log scale for this analysis (as log-base-e of weeks is not a value that most readers will be able to interpret). 11. I appreciate the authors' addition of the first four sentences of the Discussion in response to my prior comment. But these four sentences are now followed by discussion of the fact that TBscore "translates directly into a mortality risk", based on prior evaluations that TBscore is associated with increased mortality. However, the authors have direct mortality data on their population here, and after adjustment for other confounders, they find no change in mortality as a function of delay/symptom duration. I would suggest removing the 5th and 6th sentences in the Discussion from the manuscript, as I prefer the later discussion of the mortality results as actually observed. 12. I can appreciate the authors' point about Figure 2 and am willing to accept it, but would recommend that they include a legend for this figure (legends for the other figures could be useful as well!), that clearly describes what they are showing. For example, this legend could include a statement to the effect of, "All patients were in treatment by the end of the analysis period because only those patients who ultimately began treatment were included in this analysis." As a minor point, please correct minor typos in the title (to "proportion of TB patients in treatment, according to self-reported delay"). 13. As with comment 12 above, I would encourage the authors to add a legend (i.e., few sentences of written text) to Figure 3. I would also recommend that in the paragraph of the Results section starting "in figure 3…", the authors consider adding a statement that the majority of the excess mortality in people with a high TBscore occurred after the first 6 months following presentation. This might also be useful (though not mandatory) to add to the sentence in the Discussion regarding mortality differences ("there was a significant difference in mortality…")

VERSION 2 – AUTHOR RESPONSE

Reviewer: 2
R2.1a: Fisher's exact test only tests for whether there is any difference, comparing any of the four strata of delay. (Thus, for example, if the percentage of severe TB cases was 20% in the lowest quartile, 20% in the 3rd quartile, and 20% in the 4th quartile, but 80% in the 2nd quartile, the p-value for Fisher's exact test over the entire dataset would be significant.) Can the authors consider statistical tests that might more accurately support their primary conclusions? Fisher's exact tests comparing each stratum to the baseline, or to the stratum immediately underneath might be one way of doing this (though there are many others).
C2.1a: We thank the reviewer for pointing out this important statistical aspect. We have now changed the Fishers exact test to compare each stratum with the baseline (table 3).

R2.1b: I appreciate the use of TBscore >8 as a cutoff, but could the authors give some description - either in the methods or the discussion - of how severe this is? For example, describing a TB patient with a TBscore of 8 might be helpful to readers (like myself) who are not intimately familiar with this score.
C2.1b: We acknowledge the value of such an example of the typical TB patient with a specific TBscore, though we find it difficult to describe in detail because the score combines clinical findings and signs of TB. The score includes the following signs: Cough, haemoptysis, dyspnoea, chest pain and night sweating; and the following symptoms: Anaemia, tachycardia, lung-auscultation finding, fever, low body-mass index, low mid-upper arm circumference (MUAC) giving patients a TBscore from 0 to 13 (up to two points for BMI and MUAC). As a consequence the “typical” patient with a TBscore=8 can have any number of different combinations of these signs and symptoms.

R2.1c: Thank you for adding the logistic regression to the Results. Could you also please adjust the adjusted results for age (at least >/<45 years old), as this seems to be an obvious and potentially strong confounder?
C2.1c: We thank the reviewer for highlighting this typo in the newly revised manuscript. The results included were already corrected for age (e.g. strata’s from table 2), though this was not mentioned before. (p. 12, line 232).

R2.2: I also appreciate the authors inclusion of additional text to clarify their definition of "treatment delay," and I will not press this point if the authors disagree (i.e., OK for the authors not to reply to this comment on re-revision if they choose not to), but "treatment delay" means something very different to me than self-reported symptom duration. I think of "treatment delay" as having 2 components: patient delay and health system delay, and patient delay may not necessarily begin from the very onset of symptoms but perhaps from the time at which those symptoms became sufficiently bothersome to the patient to merit consideration of care-seeking. (For example, if a patient with asthma has been coughing for 20 years prior to TB diagnosis, this would not equate to a treatment delay of 20 years to me.) Again, I am willing to trust the authors as to what they believe to be the best way to resolve this issue -- but I would still contend that symptom duration and treatment delay are not necessarily the same thing.
C2.2: We thank the reviewer for highlighting this important issue and though we follow the arguments given in the review we hold that it is a generally accepted norm that delay is based on the patient’s recall of duration of symptoms, and there is no other way of measuring this variable.

R2.9: Thank you for adding Figure 4. I might recommend simply showing a bar graph, with each bar corresponding to the mean or median delay over a 6- or 12-month period, rather than a series of dots that are very difficult for the average reader to interpret. I would also recommend using a non-log scale for this analysis (as log-base-e of weeks is not a value that most readers will be able to

interpret).
C2.9: We appreciate the comments on the layout and analysis behind Figure 4. This has now been implemented and Figure 4 should be less complicated to comprehend.

R2.11. I appreciate the authors' addition of the first four sentences of the Discussion in response to my prior comment. But these four sentences are now followed by discussion of the fact that TBscore "translates directly into a mortality risk", based on prior evaluations that TBscore is associated with increased mortality. However, the authors have direct mortality data on their population here, and after adjustment for other confounders, they find no change in mortality as a function of delay/symptom duration. I would suggest removing the 5th and 6th sentences in the Discussion from the manuscript, as I prefer the later discussion of the mortality results as actually observed.
C2.11: We acknowledge the wording on mortality risk was too strong, and we have now rephrased on p.14, line 266-268

R2.12. I can appreciate the authors' point about Figure 2 and am willing to accept it, but would recommend that they include a legend for this figure (legends for the other figures could be useful as well!), that clearly describes what they are showing. For example, this legend could include a statement to the effect of, "All patients were in treatment by the end of the analysis period because only those patients who ultimately began treatment were included in this analysis." As a minor point, please correct minor typos in the title (to "proportion of TB patients in treatment, according to self-reported delay").
C2.12 We thank the reviewer for commenting on this important aspect. Legends for the figures have now been included in the manuscript accordingly.

R2.13. As with comment 12 above, I would encourage the authors to add a legend (i.e., few sentences of written text) to Figure 3. I would also recommend that in the paragraph of the Results section starting "in figure 3…", the authors consider adding a statement that the majority of the excess mortality in people with a high TBscore occurred after the first 6 months following presentation. This might also be useful (though not mandatory) to add to the sentence in the Discussion regarding mortality differences ("there was a significant difference in mortality…")
C2.13 We believe that this comment partly relies on a misunderstanding. The mortality related to TBscore is not directly analyzed in this manuscript. The figure only shows the relationship between the delay strata's and mortality. We have revised the section mentioned and included a statement on treatment delay, which we hope is what the reviewer was referring to (p. 12, line 238-240). As mentioned in 2.11 this part of the discussion section and the related conclusions has been revised.