Comments on the referee's report on our paper entitled

"Can heterogeneity in reporting behavior explain the gender gap in self-assessed health status?"

We are very grateful to the referee for carefully reading our paper and we appreciate the insights and feedback received.

Comments:

1. *Connection of this model with reporting behaviour:*
   Possible gender differences in reporting behaviour, and their relation to the gender health paradox are used to motivate this paper and the proposed methodology. However, the connection between the two is not made sufficiently clear, which weakens the motivation and clarity of the contribution. I started by finding the title (about reporting behaviour) and the abstract (about heterogeneity in discount factors) contradictory. And this remained unclear throughout the paper. Besides improving clarity in this respect, it would also be important to explain to what extent this methodology has better explanatory power than others.

   Women generally report poorer health than men on self-reported measures of health although they live longer than men. Therefore, it can be argued that there is a gender health paradox in health economics literature. We argue that the gender gap in SAH may result from the differences in discount rates of women and men. We will emphasize the connection between the two in the revised version of the study. Further, the most important feature of our study is the methodology that we used. First, we proposed a theoretical identification mechanism through a dynamic structural model which allows for heterogeneity in discount factors of individuals in order to explain the gender differences in SAH. Second, we used smoking as a proxy for individual discount rate. Although this method has some advantages and disadvantages as compared to other methods in the existing literature, this study is the first attempt to investigate the gender gap in SAH by providing a theoretical model and by testing empirically the theoretical predictions in the context of adjusting for the heterogeneity in reporting behavior.

2. *Descriptives:*
   The paper should contain descriptive statistics of all the data used, not just of SAH. A related comment, in the introduction the authors say that the gender differences are more pronounced in Turkey compared to other countries. It would be good to see some data about this too.

   The point is well taken. We will provide descriptive statistics of all the data used in the revised version of the paper. Furthermore, we will give some statistics relating to gender differences in health for Turkey.

3. *Discounting:*
   The authors hypothesize that heterogeneity in discount factors might drive gender differences and partly explain the gender health paradox. However, it remains unclear to me what the "correction" proposed is actually doing and implies. Do the authors mean that one should compare current health rather than current combined with discounted expected future health (ie, "uncorrected" SAH)? Is this what the model is doing? All this should become much clearer. Given the role that the discount factors play in this methodology, the respective estimation results should also be presented. There should also be an explanation and discussion of how those results relate to the differences obtained with the base model and the extended model. I

also believe that identification of the individual specific discount factors from smoking behaviour should at least be better justified and discussed. Finally, I wonder if there is/should be some rescaling of the predictions of equation (14), with explanatory variable smoking behaviour, into the discounting factor.

The paper proposes the theoretical framework in order to account for the possibility that current health valuation (SAH later used as a measure for this) can be a combination of current and future expected health outcomes as the referee expressed. Viewing current health valuation as this has the potential to explain the differences in valuation of health that purely caused by economic behaviors of agents rather than the differences in their fundamental objective health.

The reporting bias literature has already realized this possibility and accounts for this through allowing the responses to be heterogeneous (i.e. benchmarking different thresholds for reporting). In this respect, the current paper brings a theoretical justification for the existence of the reporting bias. In the empirical application male and female SAH outcomes when estimated via a traditional model (Base Model in Table 1: estimated via ordered probit), versus the other two (Model 1 and 2, there are two models since the interactions implied by the theoretical model are fully accounted in Model 2, whereas Model 1 is a full interaction model produced from the variables in the base model; there is a type in footnote 7 which describes the models' specifications) models reveal that the Female variable's coefficient which is significant and negative in Base model becomes insignificant in Model 1 and 2.

However, we agree with the referee that this finding should be made clearer n the paper. "Correction" in the empirical application is firstly revealed through the insignificant Female coefficient. Therefore, SAH when taken as the proxy for health valuation demonstrates no significant health differences between male and female once the model implied variables are included in the estimation. In terms of the referee's clarifying questions; one implication is that one should compare the "implicit current health" (which is not what SAH measures, but can be obtained through model's period utility coefficients) if the aim is to obtain a measure for current health. Otherwise if SAH is used, then one should be aware that SAH measures the expected discounted value of health valuation, and any comparison with the measures of current health can be misleading.

We agree with the comment that "***Given the role that the discount factors play in this methodology, the respective estimation results should also be presented***.", and in the revised version those estimations are will be included and discussed.

Regarding the comment "***There should also be an explanation and discussion of how those results relate to the differences obtained with the base model and the extended model. I also believe that identification of the individual specific discount factors from smoking behaviour should at least be better justified and discussed***.", there will be a clear extended discussion of why the specifications in the Base Model and the extended ones (Model 1 and 2) are proposed (Also see the responses to comments 5). These are basically obtained by including the interaction variables (theoretical model implied derivations leads to those interactions once the discounts obtained by smoking equation is replaced for the discount factors in the utility valuation in equation (4). Equation 12 describes the procedure and the derived variables from this replacement. We agree on the comment that these derivations and comparison to the base model should be discussed better. In the revised version there will be extended sections to explain the details of this replacement and how this procedure produces the interactions in the extended models (also see the comments to 5) which are absent in the base model. The section on identification of the individual specific discount factors from smoking behavior will be better discussed in the revised version.

4. *Other comments on (presentation of) methodology:*
   I would have liked to see a better explanation of what the assumptions described under equation (4) mean, as well a discussion of their plausibility and possible implications. Is it plausible to assume that the income growth will be the same for both gender and education groups? Again, what are the possible implications of such assumption? The authors repeat a couple of times that SAH is categorical and that an ordered model is needed because of that (somehow presented as a limitation). I think it would work better to establish this at the beginning and avoid repeating. Why are there no interactions in equations (12) and (13)?

   The assumptions basically state that shocks to health valuation are independent over time and individuals with mean zero conditional on observed state at time t (There is a typo in the last condition such that it is not conditioned on x, it will be fixed in the revised version). Of course the implications (or limitations) of those assumptions might have consequences for the identification and relatedly estimation of the model. For instance, with an autocorrelation structure of the shocks, the last term in equation (4) will not be wiped out as zero. In the revised version, those assumptions with their limitations will be discussed further as a separate paragraph. Income growth is assumed to be constant in the theoretical model. This form helps us to derive closed form solutions and consequently better represent the role of discount factors in our discussion. However, it is in general not necessary for the arguments of the paper. With a stochastic growth rate, the derivations are needed to account for the distribution of the growth rate and therefore the estimation should rely on some numerical integration procedure to calculate the expectations. This will be discussed in the revised version of the paper. The simple form assumed in the theoretical model for the income growth is to present the health valuation model's contribution to the existing literature as simple as possible as an alternative way to account for life-cycle effects. As can be seen in Table 1, in the estimation of the lambda parameter, it is differentiated across the different discrete income levels to account for some of the heterogeneity in the growth rate. However, a fully stochastic growth can also be considered as an extension or robustness of the findings, though a standard ordered probit will no longer will be an adequate estimation routine for such an extension.
   Regarding the referee's comment "***The authors repeat a couple of times that SAH is categorical and that an ordered model is needed because of that (somehow presented as a limitation). I think it would work better to establish this at the beginning and avoid repeating.***", the revised version will be written accordingly.
   There are interactions in those equations but might be needed to better explained in the text. This we will do in the revised version. In equation (12) and (13), the constructed variables z are in fact empirical counterparts of the theoretical model implied variables. As can be seen in the previous paragraph before equation (12), the zs are constructed as combinations of rho and x variables. X includes age, education and income and pho is predicted from equation (8) so in the final estimation in Table 1, pho itself is a constructed variable. Therefore, in equation (12) we present as zs and those zs include interactions and squared terms as presented in the previous paragraph before equation (12) and also in equation (15). In the estimation, the variables that constructs zs are used, so the interactions show up there. This can be better stated and explained, and that is what we will do in the revised version.

5. *Other comments on (presentation of) results:*
   What are the four lambdas in Table 1? In the equations, there is only one lambda but I do not see the relation with the four estimated ones. Why is there a constant term in Model 2 but not in Model 1? (on the other hand, it is clear that it is not identified in the base model, the standard ordered logit) I am puzzled by the fact that the base model, supposedly more restrictive, has a slightly higher likelihood than Model 2. I also do not understand why that of Model 1 is better than Model 2. As a more general comment, it should become much clearer what Model 1 and Model 2 actually are. Even more generally, it should become much clearer what parameters are presented in Table 1, in relation to the equations.

   We agree with the referee that those parameters and equations should be discussed clearly. The revised version will address all of the questions raised by the referee in this comment. Let us briefly describe the answers to those questions here. The theoretical derivation equation (4) treats income as a continuous variable. However, in the data, income is available only as discrete outcomes (very poor, poor, medium, rich and very rich as can be seen in Table 1). Therefore, in the estimation, very poor is taken as the base group (therefore omitted) and the remaining income variables are constructed as dummy variables which therefore produced the four lambdas in Table 1. This definitely should be explained in the main text and will be explained accordingly in the revised version. In fact, those reported coefficients are not directly lambdas, but the corresponding coefficients (lambda*delta3) as described in equation (15). The delta3 (the income coefficients) can be used along with those coefficients to identify lambdas. The below description summarizes what those constructed variables are based on equation (15):
   gen zet0 = 1 + discount_rho
   gen zetF = female + female*discount_rho
   gen zet1_1 = incomplete + incomplete*discount_rho
   gen zet1_2 = primary + primary*discount_rho
   gen zet1_3 = secondary + secondary*discount_rho
   gen zet1_4 = highschool + highschool*discount_rho
   gen zet1_5 = tertiary + tertiary*discount_rho
   gen zet2_1 = discount_rho + discount_rho^2 + ageG2 + ageG2*discount_rho
   gen zet2_2 = discount_rho + discount_rho^2 + ageG3 + ageG3*discount_rho
   gen zet2_3 = discount_rho + discount_rho^2 + ageG4 + ageG4*discount_rho
   gen zet3_1 = poor + poor*discount_rho
   gen zet3_2 = medium + medium*discount_rho
   gen zet3_3 = rich + rich*discount_rho
   gen zet3_4 = veryrich + veryrich*discount_rho
   gen zet4_1 = poor*discount_rho + poor*(discount_rho^2)
   gen zet4_2 = medium*discount_rho + medium*(discount_rho^2)
   gen zet4_3 = rich*discount_rho + rich*(discount_rho^2)
   gen zet4_4 = veryrich*discount_rho + veryrich*(discount_rho^2)
   These variables are used to estimate Model 2. Therefore, in Model 2, the constant actually corresponds to the coefficient of the variable zet0 (see above definition).
   There is a typo in the footnote 7 which states that "The difference between Model 1 and 2 is that Model 1 includes less interaction terms whereas Model 2 includes all interaction terms". Model 2 is actually composed of what is described above; theoretical model implied variables in the ordered probit model, which is run by bootstrap algorithm since there is an estimated

proxy for the discount rate in the estimation (therefore standard errors should account for this fact).

However, Model 1 is a model with full interactions. It takes the base model and produces all possible interactions (some of them drops out due to collinearity obviously) from the variables used in the base model. As expected it produces a higher likelihood and Pseudo R2. The number of variables used in Model 1 is 72 compared to 13 in the base model and 18 in Model 2. This is why model 1 is better than Model 2 and base model in terms of likelihood basically. Coming to the question why base model is slightly better in terms of likelihood, since two models do not use the same variables for the estimation and in fact they are producing very close fit. However, the main message of the Table 1 is about the gender paradox which can be seen from the coefficient of Female in the estimations. The full interactions Model 1 produces an insignificant female coefficient (without any rationale for including the interactions), compared to the base model which produces a significant negative coefficient (gender paradox).

Model 2 on the other hand produces an insignificant female coefficient with far less variables than Model 1 (with a rationale for including the constructed less variables), compared to the base model which produces a significant negative coefficient (gender paradox). Therefore, this result we interpret as the validity of the proposed theoretical model in the paper.

The referee suggests "***As a more general comment, it should become much clearer what Model 1 and Model 2 actually are. Even more generally, it should become much clearer what parameters are presented in Table 1, in relation to the equations.***". We agree on this comment, and the above explanations will be in the revised version with better supplementary Tables.

6. *Writing:*
   Several sentences and paragraphs should be better written. Besides some small issues with punctuation and with articles, the lack of clarity in parts of the text impairs important messages related to the general motivation and contribution of the paper, as well as the proposed methodology. These are some examples but advise the authors to revise the whole text carefully and possibly have it proofread:

   - The first sentence of the second paragraph of the introduction is too long. - I found the third paragraph quite confusing in some respects. - In the second full paragraph of page 3, there seems to be some confusing between risk and time preferences. - In the first paragraph of Section 3, it is unclear whether or not the authors find SAH a problematic measure. - I also found the first paragraph of sub-section 3.2 quite confusing. It is perhaps best to start with saying what is now in footnote 7, to avoid giving the initial impression that the model estimated will be much simpler than what was presented previously. - At the beginning of sub-section 3.3, it seems better to present clearly what is done in this paper, rather than starting with referring to the most general case in the paper of Kose and Soytas (this should perhaps be in a footnote instead).

   The point is well taken and the paper will be thoroughly proof read. Further, the paper will be revised and organized carefully.

7. *Additional references:*
   The paper would benefit from some additional references on gender health differences and gender specific health reporting such as: Case, A.and C. Paxson. 2005. "Sex differences in morbidity and mortality." Demography 42(2):189-214. Peracchi, F.and C. Rossetti. 2012.

"Heterogeneity in health responses and anchoring vignettes." Empirical Economics 42(2):513-538.

Thank you for your suggestions. The references that you mentioned will be considered in the revised version.