PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form ([http://bmjopen.bmj.com/site/about/resources/checklist.pdf](http://bmjopen.bmj.com/site/about/resources/checklist.pdf)) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

<table>
<thead>
<tr>
<th>TITLE (PROVISIONAL)</th>
<th>Valuing Health Related Quality of Life among the Indian Population: A Protocol for the Development of an EQ-5D Value-set for India using an Extended design (DEVINE) Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUTHORS</td>
<td>Jyani, Gaurav; Prinja, Shankar; Kar, S; Trivedi, Mayur; Patro, Binod; Purba, Fredrick; Pala, Star; Raman, Swati; Sharma, Atul; Jain, Shalu; Kaur, Manmeet</td>
</tr>
</tbody>
</table>

VERSION 1 – REVIEW

<table>
<thead>
<tr>
<th>REVIEWER</th>
<th>Bram Roudijk</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EuroQol Research Foundation, the Netherlands</td>
</tr>
<tr>
<td>REVIEW RETURNED</td>
<td>03-Jun-2020</td>
</tr>
</tbody>
</table>

GENERAL COMMENTS

General comments:
I think this is a very important study, which should be done meticulously. Furthermore, this study represents an enormous logistical feat involving sampling in multiple regions with different cultures and languages. Because of this, I think it is a great idea to publish the study protocol first as it could serve as an example protocol for valuation studies in large and heterogeneous countries. I think the authors have done a great job in writing this protocol paper. The sampling strategy has been described into the smallest detail, which is a great asset. However, some points will need to be addressed before it can be accepted for publication.

In the subsecond section of the section “valuation methods”, the cTTO is explained. I think it may be better to rewrite this section, as it is currently not very clear to the reader how this method works. I would suggest using examples such as “Respondents are asked whether they prefer to live for 10 years in good health or 10 years in some inferior state B.” Then explain the subsequent steps as the time in good health varies. Once you have explained these steps in detail, you can start explaining the lead time TTO part of the cTTO.

I think it would be a good idea to mention that respondents complete warmup tasks before starting the “real” valuation tasks. (e.g. the two wheelchair examples and three practice states (mild, severe, difficult to imagine)).

I think it would make sense to move the “quality control” section before the “testing the extended EQVT design” section. The quality control is an element of the EQ-VT protocol, while the extended design is an extension. Therefore it would make sense to write about the quality control first.

In the quality control section, interviewer effects are not explained in detail. I would add a couple of sentences to describe how these will be addressed.

The manuscript may benefit from some English language edits. Although the manuscript generally reads well, some
improvements in the grammar and interpunction could improve the overall quality of the manuscript.

Specific comments:

Article summary, strengths and limitations.

First bullet: actually, there is a value set for the EQ-5D-3L for Sri Lanka: Kularatna, S., Whitty, J. A., Johnson, N. W., Jayasinghe, R., & Scaffham, P. A. (2015). Valuing EQ-5D health states for Sri Lanka. Quality of Life Research, 24(7), 1785-1793. However, this is a VAS-based value set, which is not a preference based assessment. Maybe you could state instead that this is the first preference based valuation study/ or first valuation study using cTTO and DCE in South Asia?

Fourth bullet: This is a problem among all EQ-5D valuation studies, which use the same protocol, and possibly for the valuation of other instruments as well.

Introduction

First section, beforelast sentence. It is mentioned that utilities are likely to be affected by social and cultural factors. There is a broad scala of evidence that social factors affect utilities, and it may be good to provide literature references for that.


Furthermore, there is actually evidence that suggests that even though there are large differences in utilities between countries, this is unrelated to cultural values. (this does not change your conclusion though) This can be read in:


Valuation methods

First section, second sentence. I think it would suffice to say that the EQ-VT sofware, developed by the EuroQol Group is used.

The Valuation Working Group can be removed from this sentence. The Modelling

Equation 1 represents an OLS regression, as there is no random intercept specified. If you want to use a random intercept, this needs to be specified in equation 1 as an extra error term (you will need for example an \( \varepsilon_{i,t} \) and a \( \mu_{0i} \), where \( t \) is the response and \( i \) the respondent.) It will then be a Generalised Least Squares regression instead of an OLS regression though.

I think it would be a good idea to clarify that equation 2 uses the same structure as equation 2, regarding the parameters for the level-attribute combinations, so that it is in fact a 20 parameter model as well.

The description of the hybrid regression is overly complicated in my opinion. I would keep the first sentence of this section, and then replace the second sentence with something like: The hybrid model will combine the likelihood functions of a linear model for the cTTO data and a logit model for the DCE data. Then continue with the third sentence as currently written. I would then remove the following sentence: “The dependent variable in the c-TTO part of the model will be defined as 1 minus the c-TTO observed values for a given health state to indicate disutility, and therefore
coefficients will express utility decrements. For DCE, the dependent variable will have a binary outcome 0/1 indicating the respondent’s choice for each pair of EQ-5D-5L states."

EQ-5D reference values
I would move the last sentence, “Modelling will be.... package.”, and make it the first sentence of the Modelling section.

Testing the extended EQ-VT design
Maybe it would make sense to discuss this in relation to the following paper:
This paper actually looks at whether we can use even smaller designs to generate tariffs for the Eq-5D-5L.

REVIEWER
Juan M. Ramos-Goñi
Axentiva Solutions S.L. (Spain)
I was freelancer for the EuroQol Research Foundation and I’m member of the EuroQol Group

REVIEW RETURNED
18-Jun-2020

GENERAL COMMENTS
Dear authors,
Thank you for giving me the opportunity of reviewing your manuscript. The manuscript is well written and straightforward to follow. However, I regret to say that I do not fell that the manuscript adds enough value to the current literature as it mainly focused in describing an existing and well documented EQ-5D-5L valuation study protocol. There 2 main points making your manuscript to differ from the existing literature, namely the specific sample strategy for the study in India and the 8 cTTO blocks added to the standard protocol. I do not consider that being enough value. Therefore, I recommend to reject this manuscript and wait until having the results of the study and integrate the information provided in this manuscript with the value set results.

REVIEWER
Koonal Shah
PHMR, UK

REVIEW RETURNED
03-Aug-2020

GENERAL COMMENTS
General comments:
1. More information about the selection of geographic states would be helpful, particularly for readers who are not very familiar with the Indian setting. What is the average income and health status in these states, and how does that compare with the country as a whole?
2. Different languages are spoken in the various states, but the manuscript is missing a description of how this affects the study design. Are different versions of the EQ-5D-5L and EQ-VT being used? Will each state be assigned its own set of interviewers who possess the necessary language skills?
3. The authors should consider citing other EQ-5D-5L valuation studies that employed similar data analysis and modelling purposes.
4. The manuscript contains a number of typographical and grammatical errors. For example, many phrases and sentences are missing an article such as ‘the’ and ‘a’. This includes the title, which would read better as: Valuing health-related quality of life among the Indian population: a protocol for the development of an
Specific comments:
1. Page 3 (article summary), line 9. The authors claim that this is the largest health state valuation of the world, but I believe the UK valuation study of the EQ-5D-3L (Dolan, 1997) used a larger sample. Also, in Brazil a saturation study was undertaken in which EQ-5D-3L health states were valued by over 9,000 respondents (Santos et al., 2016).
3. Page 5, lines 48-51. I don’t think it is accurate to describe the 86 health states as ‘representative’ of the 3,125. My understanding is that they were selected for their statistical properties rather than to be representative of the full set of health states. I also understand that while there are 86 health states that are valued directly, the final values for those health states are those predicted via the model rather than the directly observed values.
4. Page 7, line 54. Not all readers will be familiar with these acronyms.
5. Page 9, lines 49-57 (and elsewhere). ‘better/worse than death’ should be ‘better/worse than dead’.
6. Page 9, lines 49-52. Lead time is added to both Life A (which involves time in full health) and Life B (which involves time in the health state under evaluation). This is not the same as what is described by the authors here.
7. Pages 9-10. Will be ordering of TTO and DCE tasks be randomised?
8. Page 10, lines 3-6. The health state value formulae do not make sense without defining x and t. The TTO protocol paper in Pharmacoeconomics by Oppe et al. (2016) may be worth citing to support the description of the TTO task.
10. Page 13, lines 19-22. The explanation of how self-reported EQ-5D profiles are converted into values could be improved.
11. Page 13, line 26. The statement about undertaking modelling in STATA should appear elsewhere (perhaps in the modelling section). This particular sub-section does not refer to any modelling work.

VERSION 1 – AUTHOR RESPONSE

Comments by Reviewer- 1:
Comments to the Author: I think this was a well written and useful manuscript. Some improvements will need to be made before it is acceptable for publication though. Please see attached file for detailed comments.

I think this is a very important study, which should be done meticulously. Furthermore, this study represents an enormous logistical feat involving sampling in multiple regions with different cultures and languages. Because of this, I think it is a great idea to publish the study protocol first as it could serve as an example protocol for valuation studies in large and heterogeneous countries. I think the authors have done a great job in writing this protocol paper. The sampling strategy has been described into the smallest detail, which is a great asset. However, some points will need to be addressed before it can be accepted for publication.

Reviewer 1- General comment 1: In the sub-second section of the section “valuation methods”, the cTTO is explained. I think it may be better to rewrite this section, as it is currently not very clear to the reader how this method works. I would suggest using examples such as “Respondents are asked whether they prefer to live for 10 years in good health or 10 years in some inferior state B.” Then explain the subsequent steps as the time in good health varies. Once you have explained these steps in detail, you can start explaining the lead time TTO part of the cTTO.

Authors’ response: We thank the reviewer for highlighting this and suggesting the approach to make the explanation comprehensive. Complying with the suggestion of the reviewer, we have rewritten this part of the manuscript. The revised text is read as follows:

Lines 180-219 on pages 11-12: “In the standard design of c-TTO, there will be 10 blocks of health states. Each block will contain 10 health states which includes one anchor state (55555). The blocks used for interview will be randomly selected by the EQ-VT software. In TTO valuation, the respondent is asked to indicate the amount of time he/she is willing to give up to attain perfect health. The respondent will be asked to imagine two alternative health states (life A and life B) described on screen and express the preference using TTO. The respondents will be asked whether they prefer to live for 10 years in perfect health (life A) or 10 years in some inferior health state (life B). It will also be explained to the respondents that at the end of the stated time, there will be an immediate painless death in both the lives. As a rational choice, the respondents would prefer good health (life A) over the state of inferior health (life B) when the time available in both the alternatives is equal (10 years). Thereafter, the time available in life B will be kept constant at 10 years, while the time available in life A will be decreased sequentially, and the respondent will be asked to select the better alternative between life A and life B. Thereby, the respondent will be asked to state its preference between ‘living for 10 years in an inferior health state’, or ‘living for less than 10 years in perfect health’. This exercise will be done till the point of indifference is achieved (when the respondent feels that both life A and life B are of equal value). At this point of indifference, the traded-off time in life A will be recorded, which reflects the time in perfect health the respondent is willing to give up in order to avoid living in the inferior health state (life B). The severe the health state, the more is the time the respondent wants to give up to avoid it. This exercise is known as conventional TTO.
Nevertheless, there are certain health states, for which the respondent prefers to die immediately rather than living in that health state. These health states are known as worse than dead (WTD), and their valuation will be done with the help of c-TTO. The c-TTO approach will begin with the conventional TTO for all health states, followed by a lead time TTO (LT-TTO) in the scenario where the participants’ response will indicate the health state to be worse than dead (WTD). The c-TTO involves adding healthy life years (‘lead time’) before both the alternatives (life A and life B) being compared. This will allow the respondent to trade-off these additional years when he or she considers the health state in life B to be worse than dead. As per the EuroQol group’s recommendations, a lead time of 10 years will be used. The value of health will be calculated as $x/t$ for better than dead health states and $(x-10)/t$ for worse than dead health states, where ‘$x$’ is the time remaining in life A at the point of indifference, and ‘$t$’ is the time offered in life B, i.e., 10 years. This being a cognitively demanding exercise, first a small training exercise using an example of ‘being in a wheelchair’ as life B will be performed with the respondent to make sure the respondent understands the concept of TTO. The concept trading-off the time in both ‘better than dead’ and ‘worse than dead’ health states will be explained in this exercise. This will be followed by three practice tasks in which the respondent will be asked to value three health states of varying severity (mild, severe and difficult to imagine). Once the wheelchair example and practice exercises get over, the respondents will be assigned a block of 10 health states, on which the valuation will be done.

**Reviewer 1 - General comment 2:** I think it would be a good idea to mention that respondents complete warmup tasks before starting the “real” valuation tasks. (e.g. the two wheelchair examples and three practice states (mild, severe, difficult to imagine))

**Authors’ response:** We thank the reviewer for suggesting this change. As per the suggestion, we have added the relevant lines as follows:

Lines 211-219 on page 12: “This being a cognitively demanding exercise, first a small training exercise using an example of ‘being in a wheelchair’ as life B will be performed with the respondent to make sure the respondent understands the concept of TTO. The concept trading-off the time in both ‘better than dead’ and ‘worse than dead’ health states will be explained in this exercise. This will be followed by three practice tasks in which the respondent will be asked to value three health states of varying severity (mild, severe and difficult to imagine). Once the wheelchair example and practice exercises get over, the respondents will be assigned a block of 10 health states, on which the valuation will be done.”

**Reviewer 1 - General comment 3:** I think it would make sense to move the “quality control” section before the “testing the extended EQVT design” section. The quality control is an element of the EQ-VT protocol, while the extended design is an extension. Therefore, it would make sense to write about the quality control first.

**Authors’ response:** We completely agree with the suggestion of the reviewer regarding reordering the abovementioned sections. Complying with the suggestion, we have moved the section of ‘Quality Control’ before the section of ‘Testing the extended EQVT design’. In the revised manuscript, the
section of ‘Quality control’ is at lines 299-339 on pages 17-18, whereas the section of ‘Testing the extended EQVT design’ is at lines 340-399 on pages 18-21.

**Reviewer 1- General comment 4:** In the quality control section, interviewer effects are not explained in detail. I would add a couple of sentences to describe how these will be addressed.

**Authors’ response:** We thank the reviewer for seeking more clarification on this aspect. We have added following lines in the quality control section to describe interviewer effects, and their potential resolution. The added text reads as follows:

Lines 323-339 on page 17-18: “As a part of quality control, interviewers’ effect will also be assessed in addition to the protocol compliance. The presence of interviewers’ effect in the data will be assessed by indicators like distribution of TTO responses with respect to different health states for each interviewer, presence of clustering the TTO responses, health states given a value of ‘zero’ in the TTO tasks, health states given value of ‘less than zero’ in the TTO tasks, and proportion of non-traders (individual who refuse to give up any amount of time in the TTO, thus giving all health states the value of 1) in the respondents. The distribution of TTO responses will be interpreted by comparing the data of a specific interviewer with the pooled data from all interviewers. Any interviewer reflecting interviewers’ effect will be assisted by local team of investigators via phone and video calls during the conduct of next round of pilot interviews. During this process, it will be observed if the interviewer is influencing respondent by its interviewing style, shortening the task due to laziness, or facing difficulty in explaining WTD element of the c-TTO task. Personalized feedback will be provided to interviewers to overcome any such difficulty. Poor performing interviewers will be retrained, and removed from the team if no improvement is seen after retraining. The interviewers will be allowed to start the real data collection once they will have achieved a stable performance on the QC protocol. This QC check and personalized feedback process will constantly be followed throughout the process of real data collection.”

**Reviewer 1- General comment 5:** The manuscript may benefit from some English language edits. Although the manuscript generally reads well, some improvements in the grammar and punctuation could improve the overall quality of the manuscript.

**Authors’ response:** We thank the reviewer for this comment. As per the suggestion, we have corrected the minor grammatical errors found in the manuscript.

**Reviewer 1- Specific comment 1 (Article summary, strengths and limitations):** First bullet: actually, there is a value set for the EQ-5D-3L for Sri Lanka: Kularatna, S., Whitty, J. A., Johnson, N. W., Jayasinghe, R., & Scuffham, P. A. (2015). Valuing EQ-5D health states for Sri Lanka. Quality of Life Research, 24(7), 1785-1793. However, this is a VAS-based value set, which is not a preference-based assessment. Maybe you could state instead that this is the first preference-based valuation study/ or first valuation study using cTTO and DCE in South Asia?
Authors’ response: We are thankful to the reviewer for bringing forward this fact. Complying with the edit suggested by the reviewer, we have revised this point in the manuscript. The revised text reads as follows:

Lines 51-53 on page 3: “This is the largest EQ-5D-5L valuation study of the world, and the first preference based valuation study in the South-Asia.”

Reviewer 1- Specific comment 2 (Article summary, strengths and limitations): Fourth bullet: This is a problem among all EQ-5D valuation studies, which use the same protocol, and possibly for the valuation of other instruments as well.

Authors’ response: We completely agree with the reviewer on this point that is a common problem among all valuation studies. Nevertheless, we wanted to specifically highlight this point in the manuscript due to the socio-demographic context in which this study will take place. The literacy level in India is not as good as it is in the developed countries, and almost two-third of the population of the country is living in the rural areas. The tasks of time trade off and discrete choice experiments proposed as a part of this study are cognitively demanding. It would definitely be a challenge for the study team to successfully execute these complex exercises in a sample of around 3000 such individuals. These factors make Indian valuation study much more challenging and difficult as compared to its Western counterparts. Therefore, authors feel that it is worthwhile to mention this logistic challenge beforehand. Hence, this point is highlighted specifically in the India context that ‘Due to the exhaustive and lengthy process of interviewing, the respondent fatigue may set in, which may adversely impact the valuation of health states during the latter part of the interview.’

Reviewer 1- Specific comment 3 (Introduction): First section, before last sentence: It is mentioned that utilities are likely to be affected by social and cultural factors. There is a broad scale of evidence that social factors affect utilities, and it may be good to provide literature references for that.

(e.g. Jakubczyk, M., Golicki, D., & Niewada, M. (2016). The impact of a belief in life after death on health-state preferences: True difference or artifact?. Quality of Life Research, 25(12), 2997-3008.


Furthermore, there is actually evidence that suggests that even though there are large differences in utilities between countries, this is unrelated to cultural values. (this does not change your conclusion though) This can be read in:

Authors’ response: We are grateful to the reviewer for suggesting the relevant studies, and we appreciate the efforts of the reviewer for providing us the detailed references. In accordance with the suggestion of the reviewer, we have added the literature references for these studies in the manuscript. In the bibliography, these studies are mentioned as reference numbers 8, 9, 10 and 11 in the revised version of the manuscript.

Reviewer 1- Specific comment 4 (Valuation methods): First section, second sentence: I think it would suffice to say that the EQ-VT software, developed by the EuroQol Group is used. The Valuation Working Group can be removed from this sentence.

Authors’ response: We agree with the point mentioned by the reviewer. Complying with the suggestion, we have removed the mention of Valuation Working Group. The revised text read as follows:

Lines 174-175 on page 10-11: “We will be using EuroQol Group’s Valuation Technology (EQ-VT) software generated by the EuroQol Group.”

Reviewer 1- Specific comment 5 (Valuation methods): Equation 1 represents an OLS regression, as there is no random intercept specified. If you want to use a random intercept, this needs to be specified in equation 1 as an extra error term (you will need for example an $\epsilon_t$ and a $\mu_0i$, where $t$ is the response and $i$ the respondent.) It will then be a Generalized Least Squares regression instead of an OLS regression though.

I think it would be a good idea to clarify that equation 2 uses the same structure as equation 2, regarding the parameters for the level-attribute combinations, so that it is in fact a 20 parameter model as well.

The description of the hybrid regression is overly complicated in my opinion. I would keep the first sentence of this section, and then replace the second sentence with something like: The hybrid model will combine the likelihood functions of a linear model for the cTTO data and a logit model for the DCE data. Then continue with the third sentence as currently written. I would then remove the following sentence: “The dependent variable in the c-TTO part of the model will be defined as 1 minus the c-TTO observed values for a given health state to indicate disutility, and therefore coefficients will express utility decrements. For DCE, the dependent variable will have a binary outcome 0/1 indicating the respondent’s choice for each pair of EQ-5D-5L states.”

Authors’ response: We thank the reviewer for highlighting this point. As per the suggestion of the reviewer, we have edited this section of the manuscript. The revised section reads as follows:

Lines 227-277 on page 13-15: “Modelling will be undertaken using the STATA statistical package. TTO data will be modelled using the response values as dependent and the health states as explanatory variables. A main effects model will be employed that will include a constant and 5 main effects derived from the EQ-5D-5L descriptive system, using generalized least squares (GLS) and tobit models. The constant will reflect the utility decrement associated with any deviation from full health. Random effects
will be included to account for the panel structure in the data. The basic equation for the random-effects GLS regression with random intercept will be as follows:

\[ Y_{it} = \beta_0 + \beta_{MO} MO_{it} + \beta_{SC} SC_{it} + \beta_{UA} UA_{it} + \beta_{PD} PD_{it} + \beta_{AD} AD_{it} + \epsilon_{it} + \mu_{0i}, \]  

(1)

where \( Y_{it} \) refers to the TTO values dependent variable, \( \mu_{0i} \) will be the individual specific error component and \( \epsilon_{it} \) refers to the combined time series and cross-section error component, \( i \) indicating the respondent, and \( t \) accounting for the panel structure of the dataset (because there are 10 cTTO questions per respondent). The terms MO, SC, UA, PD and AD refer to five dummy-coded regressors for mobility, self-care, usual activities, pain/discomfort and anxiety/depression, each representing the five levels of the EQ-5D-5L. So in the equation 1, each dimension has four coefficient with first level as baseline is

\[ \beta_{MO} MO_{it} = \beta_{M1} MO_{it} + \beta_{M2} MO_{it} + \beta_{M3} MO_{it} + \beta_{M4} MO_{it}, \]

which is similar for SC, UA, PD and AD, leading to a total of 20 regressors plus the constant. The tobit model will assume a latent variable \( Y_{it}^* \) underlying the observed \( Y_{it} \) cTTO values. This will match with the censored cTTO data, which by nature of the applied cTTO task will be left-censored at -1. The tobit model will account for this censored nature of the data by estimating the latent variable \( Y_{it}^* \), which can take on predicted preference values extrapolated beyond the range of the observed values. A likelihood function will be used to adjust the parameter estimates for the probability of \( Y_{it} \) being above the censoring value. Hence, in the tobit model, the observed value \( Y_{it} \) will have the following properties when the censoring value is -1:

\[ Y_{it} = \begin{cases} Y_{it}^* & \text{if } Y_{it}^* > -1 \\ -1 & \text{if } Y_{it}^* \leq -1 \end{cases} \]

The equation for \( Y_{it}^* \) will be linear. The DCE data will be modeled under random utility using the conditional logit model. The model will include the same 5 parameters as the cTTO model, reflecting utility decrements associated with levels 1, 2, 3, 4 and 5 for each of the five domains: MO, SC, UA, PD and AD. This model will have same structure as equation 1 regarding the parameters for the level-attribute combinations, so it will be a 20 parameter model as well. The regression equation is given below.

\[ U_{js} = \beta_1 MO_{js} + \beta_2 SC_{js} + \beta_3 UA_{js} + \beta_4 PD_{js} + \beta_5 AD_{js} + \epsilon_{js}, \]  

(2)

where \( js \) will be the choice alternative in the choice sets.

As both TTO and DCE data provide information about the values of health states, we will also implement a hybrid modelling approach that will make use of both c-TTO and DCE datasets to estimate the potential value sets. This approach has been used in several national EQ-5D-5L valuation studies. The hybrid model will combine the likelihood functions of a linear model for the c-TTO data and a logit
model for the DCE data. As the coefficients will be estimated from a conditional logit and expressed on a latent arbitrary utility scale, we will use a rescaled parameter \( \theta \), which will assume that the c-TTO model coefficients are proportional to DCE model coefficients. This method will combine the utility values elicited in the c-TTO for the 150 health states with utility values elicited in the DCE experiment for 196 pairs of states. We will use cluster estimation to acknowledge that for each participant included in the models, 10 c-TTO and 7 DCE responses are available. We will also estimate adjusted hybrid model which adjust the social demographical variables like age, sex etc."

**Reviewer 1 - Specific comment 6 (EQ-5D reference values):** I would move the last sentence, “Modelling will be... package.”, and make it the first sentence of the Modelling section.

Complying with the suggestion of the reviewer, we have moved this sentence to the starting of the Modelling section. This reads as follows in the manuscript:

Line 227 on page 13: “Modelling will be undertaken using the STATA statistical package.”

**Reviewer 1 - Specific comment 7 (Testing the extended EQ-VT design):** Maybe it would make sense to discuss this in relation to the following paper: Yang, Z., Luo, N., Bonsel, G., Busschbach, J., & Stolk, E. (2019). Effect of health state sampling methods on model predictions of EQ-5D-5L values: small designs can suffice. Value in Health, 22(1), 38-44.

This paper actually looks at whether we can use even smaller designs to generate tariffs for the EQ-5D-5L.

**Authors’ response:** We thank the reviewer for suggesting the relevant resource to discuss this point. Accordingly, we have also discussed the extended design in relation to the research paper which the reviewer has suggested. The revised section in the manuscript reads as follows:

Lines 350-372 on page 19-20: “In contrast to the conventional EQ-VT protocol, which is optimized for a sample size of around 1000, the current study aims to collect data from 2700 respondents. This offers an opportunity to add more health states and assess the additional value of using a richer number of health states in predicting the utility value of all 3125 health states. In the conventional EQ-VT design, for the method TTO, 10 blocks of health states are used, which account for 86 different health states. These health states are selected using DCE technique, combining orthogonality with priors. Each block includes one most severe health state (55555) as anchor state, and one of the five very mild health states (which demonstrates slight problem in any one of the five dimensions, i.e, 11112, 11121, 11211, 12111 and 21111). The remaining eight unique health states in each block (in total 80 health states in 10 blocks) are selected using Monte Carlo simulations to predict the prior values obtained from the multinational pilot study.19 This set of 80 states is selected on the mean squared error (MSE) between the prior parameters and estimated parameters from a main effects model, and level balance, but without making orthogonality an explicit criterion.29, 43 A dedicated direct EQ-VAS valuation study employing saturated VAS dataset compared the prediction performance of the 86 health states subset with alternative smaller subset of health states.43 The study found that the orthogonal design with 25
states performed closely to the standard EQ-VT with 86 states. However, a caveat to the use of the small orthogonal design lies in the large mispredictions in case of mild health states. Therefore, the current study aims to assess the added value of increased number of health states and increased number of observations per health state using extended design. In the extended design for the current study, eight additional blocks have been added, consisting of 64 new health states. This selection was guided by added-value considerations, taking the initial ten blocks as point of departure. Hence, we have a conventional 10 blocks design, and an extended 10+8 blocks design."

Comments by Reviewer- 2:

Comments to the Author: Thank you for giving me the opportunity of reviewing your manuscript. The manuscript is well written and straightforward to follow. However, I regret to say that I do not feel that the manuscript adds enough value to the current literature as it mainly focused in describing an existing and well documented EQ-5D-5L valuation study protocol. There 2 main points making your manuscript to differ from the existing literature, namely the specific sample strategy for the study in India and the 8 cTTO blocks added to the standard protocol. I do not consider that being enough value. Therefore, I recommend to reject this manuscript and wait until having the results of the study and integrate the information provided in this manuscript with the value set results.

Authors’ response: We thank the reviewer for taking time to read our manuscript and we are pleased to know that you found it well-written and straightforward to follow. Further, regarding the point that some of the parts of the manuscript may appear describing the existing protocol, we would like to mention that as we will strictly be following it throughout the study, so we thought it would be good to explain all the planned steps in a context specific format, rather than just explaining the additional aspects of the DEVINE study. This was done to make the research protocol comprehensive, without leaving behind any of the methodological aspect, even if it has been previously reported, so that it becomes complete and easy to follow for the readers. Due care has been followed while writing the manuscript to correctly cite the available protocol. Moreover, even after the availability of the EQ-5D-5L valuation study protocol, it is a challenge to implement it in country of 1.4 billion people with inimitable linguistic, cultural and geographic diversity. Therefore, as its execution will involve an enormous logistical feat, we aspired to explain all the steps to be followed using the context of a heterogeneous country. Further, as the study will have a bearing on the financing decisions for such a large population (the study has been funded by the central HTA agency of the Government of India, which will be using the evidence generated as a part of this study), it would be a good idea to document each of the step transparently. Lastly, as you have mentioned, there are two unique areas of research which we have reported in detail in this protocol paper.

Comments by Reviewer- 3:
Reviewer 3- General comment 1: More information about the selection of geographic states would be helpful, particularly for readers who are not very familiar with the Indian setting. What is the average income and health status in these states, and how does that compare with the country as a whole?

Authors’ response: We thank the reviewer for seeking more clarification on this aspect. As per the suggestion, we have provided the information of these states on the indicators of income, health and education. We have also compared these indicators of different states with the country as a whole. This information has been mentioned as Table-1 in the manuscript, which reads as follows:

<table>
<thead>
<tr>
<th></th>
<th>Per Capita State Domestic Product (in INR)$^{23}$</th>
<th>Infant Mortality Rate$^{24}$</th>
<th>Literacy Rate$^{25}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uttar Pradesh</td>
<td>66512</td>
<td>41</td>
<td>73</td>
</tr>
<tr>
<td>Meghalaya</td>
<td>89024</td>
<td>39</td>
<td>75.5</td>
</tr>
<tr>
<td>Odisha</td>
<td>95164</td>
<td>41</td>
<td>77.3</td>
</tr>
<tr>
<td>India</td>
<td>126406</td>
<td>33</td>
<td>77.7</td>
</tr>
<tr>
<td>Tamil Nadu</td>
<td>193750</td>
<td>16</td>
<td>82.9</td>
</tr>
<tr>
<td>Gujarat</td>
<td>197447</td>
<td>30</td>
<td>82.4</td>
</tr>
<tr>
<td>Haryana</td>
<td>236147</td>
<td>30</td>
<td>80.4</td>
</tr>
</tbody>
</table>

Reviewer 3- General comment 2: Different languages are spoken in the various states, but the manuscript is missing a description of how this affects the study design. Are different versions of the EQ-5D-5L and EQ-VT being used? Will each state be assigned its own set of interviewers who possess the necessary language skills?

Authors’ response: We thank the reviewer for pointing-out this missing piece of information from the manuscript. We would like to mention that given the linguistic diversity among the states, every state will be assigned its own set of interviewers and separate training sessions will be organized for all the states. Proficiency in English as well as local language will be one of the essential criteria to be qualified for the post of interviewer. These training sessions will be bilingual in nature, so that the interviewers...
get to know the nuances of interview process in both English as well as local language. The duration of this training would be five days per site. Investigators from all the sites (proficient in their respective local language) had been trained at the EuroQol head-office. The training to local interviewers will jointly be conducted by two members from PGIMER Chandigarh (coordinating site), and one local investigator (proficient in local language). To ensure the uniformity of the training process across the states, all the training sessions will be using the same training agenda and will be coordinated by same trainers from PGIMER Chandigarh. For this purpose, EQ-VT and EQ-5D-5L have been translated into five different Indian languages (Hindi, Gujarati, Tamil, Odia and Assamese).

Keeping in mind the restrictions of word-count, this description was omitted from the earlier version of the manuscript. However, complying with the suggestion, we have made a brief mention of this process in the revised manuscript. The revised text reads as follows:

Lines 308-311 on page 17: “Given the linguistic diversity among the states, every state will be assigned its own set of interviewers and separate training sessions will be organized for all the states. For the purpose of data collection, EQ-VT and EQ-5D-5L have been translated into five different Indian languages (Hindi, Gujarati, Tamil, Odia and Assamese).”

Reviewer 3- General comment 3: The authors should consider citing other EQ-5D-5L valuation studies that employed similar data analysis and modelling purposes.

Authors’ response: We thank the reviewer for this comment. Complying with the suggestion of the reviewer, we have cited the relevant studies in the manuscript, which have used the hybrid model to develop the value-set. In the section of modeling in the manuscript, we have added the following line:

Lines 262-263 on page 15: “This approach has been used in several national EQ-5D-5L valuation studies. 30-38

The following studies have been cited for this purpose:


Reviewer 3- General comment 4: The manuscript contains a number of typographical and grammatical errors. For example, many phrases and sentences are missing an article such as ‘the’ and ‘a’. This includes the title, which would read better as: Valuing health-related quality of life among the Indian population: a protocol for the development of an EQ-5D value set for India using an extended design (DEVINE) study. The use of hyphens is also inconsistent, e.g. EQ-5D sometimes contains a hyphen and sometimes does not (EQ5D), and in some cases hyphens are followed by an unnecessary space.

Authors’ response: We are thankful to the reviewer for highlighting this aspect. As per the suggestion of the reviewer, we have done the language edits and corrected the grammatical errors found in the manuscript. The mention of ‘EQ-5D’ has been corrected using a uniform style throughout the manuscript. We have also edited the title of the manuscript, which now reads as:

“Valuing health-related quality of life among the Indian population: a protocol for the development of an EQ-5D value set for India using an extended design (DEVINE) study”

Reviewer 3- Specific comment 1: Page 3 (article summary), line 9. The authors claim that this is the largest health state valuation of the world, but I believe the UK valuation study of the EQ-5D-3L (Dolan, 1997) used a larger sample. Also, in Brazil a saturation study was undertaken in which EQ-5D-3L health states were valued by over 9,000 respondents (Santos et al., 2016).

Authors’ response: We thank the reviewer for this comment. Complying with the suggestion of the reviewer, we have rephrased the line as follows:

Lines 51-53 on page 3: “This is the largest EQ-5D-5L valuation study of the world, and the first preference based valuation study in the South-Asia.”

Reviewer 3- Specific comment 2: Page 4, lines 51-53. EQ-5D is the proper name for the instrument – it is not shorthand for ‘EuroQol 5 dimension’. See recent paper on nomenclature by Brooks et al. (2020) in the Journal of Patient-Reported Outcomes.
Authors’ response: We thank the reviewer for highlighting the error. Complying with the recommendations of Brooks et al, we have corrected all the pertinent nomenclature throughout the manuscript.

Reviewer 3- Specific comment 3: Page 5, lines 48-51. I don’t think it is accurate to describe the 86 health states as ‘representative’ of the 3,125. My understanding is that they were selected for their statistical properties rather than to be representative of the full set of health states. I also understand that while there are 86 health states that are valued directly, the final values for those health states are those predicted via the model rather than the directly observed values.

Authors’ response: We thank the reviewer for throwing more light on this aspect. In accordance with the suggestion of the reviewer, we have removed the phrase ‘representative’, and revised the line as follows:

Lines 110-112 on page 6: “Therefore, in the current design, only 86 health states are valued directly using a 10 blocks design, and the values of other 3039 health states are predicted using statistical modelling.”

In order to bring in more clarity on the aspect raised by the reviewer, we have added the following lines in the manuscript (in the section describing the extended EQ-VT design):

Lines 353-362 on pages 19: “In the conventional EQ-VT design, for the method TTO, 10 blocks of health states are used, which account for 86 different health states. These health states are selected using DCE technique, combining orthogonality with priors. Each block includes one most severe health state (55555) as anchor state, and one of the five very mild health states (which demonstrates slight problem in any one of the five dimensions, i.e, 11112, 11121, 11211, 12111 and 21111). The remaining eight unique health states in each block (in total 80 health states in 10 blocks) are selected using Monte Carlo simulations to predict the prior values obtained from the multinational pilot study. This set of 80 states is selected on the mean squared error (MSE) between the prior parameters and estimated parameters from a main effects model, and level balance, but without making orthogonality an explicit criterion.”

Reviewer 3- Specific comment 4: Page 7, line 54. Not all readers will be familiar with these acronyms.

Authors’ response: Complying with the concern raised by the reviewer, we have provided the full-forms of the acronyms GSDP (Gross State Domestic Product) and IMR (Infant Mortality Rate).

Reviewer 3- Specific comment 5: Page 9, lines 49-57 (and elsewhere), ‘better/worse than death’ should be ‘better/worse than dead’.

Authors’ response: We are thankful to the reviewer for pointing-out this aspect. Complying with the suggestion of the reviewer, we have corrected this throughout the manuscript.
Reviewer 3- Specific comment 6: Page 9, lines 49-52. Lead time is added to both Life A (which involves time in full health) and Life B (which involves time in the health state under evaluation). This is not the same as what is described by the authors here.

Authors’ response: We thank the reviewer for highlighting this error. We have corrected it in the manuscript, and the revised text reads as follows:

Lines 205-206 on page 12: “The c-TTO involves adding healthy life years (‘lead time’) before both the alternatives (life A and life B) being compared.”

We would like to mention that the description of Life A and Life B has already been provided in the preceding paragraph of the revised manuscript.

Reviewer 3- Specific comment 7: Pages 9-10. Will be ordering of TTO and DCE tasks be randomised?

Authors’ response: We would like to mention that although the blocks used for the interview will be randomly assigned to a respondent, however, within a specific block (of 10 TTO health states/ 7 DCE pairs), the ordering of TTO and DCE tasks is not randomized. There are only two possible ways in which the health states within a particular block are read, first- left to right, and second- right to left. The assignment of these alternatives to a respondent will be random.

Reviewer 3- Specific comment 8: Page 10, lines 3-6. The health state value formulae do not make sense without defining x and t. The TTO protocol paper in Pharmacoeconomics by Oppe et al. (2016) may be worth citing to support the description of the TTO task.

Authors’ response: In compliance with the comment of the reviewer, we have defined ‘x’ and ‘t’. The suggested literature has also been cited (here as well as while describing other parts of the TTO task). The revised text reads as follows:

Lines 209-211 on page 12: “The value of health will be calculated as x/t for better than dead health states and (x-10)/t for worse than dead health states, where ‘x’ is the time remaining in life A at the point of indifference, and ‘t’ is the time offered in life B, i.e., 10 years.”

Reviewer 3- Specific comment 9: Page 12, lines 12-13. The text here refers to 150 TTO health states, yet information on the extended design / increased number of health states does not appear until page 13.

Authors’ response: We thank the reviewer for this comment. We would like to mention that the description of the extended design (150 TTO health states) has been provided in detail in the subsequent paragraphs of the manuscript, as it classifies to be a separate section. We completely agree with the reviewer that an abrupt mention of 150 health state may create confusion in the mind of readers. Therefore, in the revised manuscript, we have provided a mention of this regard in the last part of the introduction. This text reads as follows:
Lines 110-117 on pages 6-7: "Therefore, in the current design, only 86 health states are valued directly using a 10 blocks design, and the values of other 3039 health states are predicted using statistical modelling. However, it has not been established if the currently used number of health states (86) is enough to generate valid value- sets. Using the extended design with a richer number of health states (150), this study will not only give an idea about the methodological robustness of current health state valuation studies, but also propose a sound and empirically tested methodology for undertaking health state valuations in health technology assessments."

Reviewer 3- Specific comment 10: Page 13, lines 19-22. The explanation of how self-reported EQ-5D profiles are converted into values could be improved.

Authors’ response: We thank the reviewer for seeking clarication on this part of the analysis. Complying with the suggestion of the reviewer, we have elaborated this aspect in detail. The revised section is being provided herewith for your kind reference:

Lines 227-277 on pages 13-15: "TTO data will be modelled using the response values as dependent and the health states as explanatory variables. A main effects model will be employed that will include a constant and 5 main effects derived from the EQ-5D-5L descriptive system, using generalized least squares (GLS) and tobit models. The constant will reflect the utility decrement associated with any deviation from full health. Random effects will be included to account for the panel structure in the data. The basic equation for the random-effects GLS regression with random intercept will be as follows:

\[
Y_{it} = \beta_{0i} + \beta_{MO}M_{Ot} + \beta_{SC}SC_{it} + \beta_{UA}UA_{it} + \beta_{PD}PD_{it} + \beta_{AD}AD_{it} + \varepsilon_{it} + \mu_{0i},
\]

where \(Y_{it}\) refers to the TTO values dependent variable, \(\mu_{0i}\) will be the individual specific error component and \(\varepsilon_{it}\) refers to the combined time series and cross-section error component, \(i\) indicating the respondent, and \(t\) accounting for the panel structure of the dataset (because there are 10 cTTO questions per respondent). The terms MO, SC, UA, PD and AD refer to five dummy-coded regressors for mobility, self-care, usual activities, pain/discomfort and anxiety/depression, each representing the five levels of the EQ-5D-5L. So in the equation 1, each dimension has four coefficient with first level as baseline is

\[
\beta_{MO}M_{Ot} = \beta_{M1}M_{O1} + \beta_{M2}M_{O2} + \beta_{M3}M_{O3} + \beta_{M4}M_{O4} + \beta_{M5}M_{O5},
\]

which is similar for SC, UA, PD and AD, leading to a total of 20 regressors plus the constant. The tobit model will assume a latent variable \(Y_{it}^{*}\) underlying the observed \(Y_{it}\) cTTO values. This will match with the censored cTTO data, which by nature of the applied cTTO task will be left- censored at -1. The tobit model will account for this censored nature of the data by estimating the latent variable \(Y_{it}^{*}\), which can take on predicted preference values extrapolated beyond the range of the observed values. A likelihood function will be used to adjust the parameter estimates for the probability of \(Y_{it}\) being above the
censoring value. Hence, in the tobit model, the observed value \( Y_{it} \) will have the following properties when the censoring value is -1:

\[
Y_{it} = \begin{cases} 
Y^*_{it} & \text{if } Y^*_{it} > -1 \\
-1 & \text{if } Y^*_{it} \leq -1
\end{cases}
\]

The equation for \( Y^*_{it} \) will be linear. The DCE data will be modeled under random utility using the conditional logit model. The model will include the same 5 parameters as the cTTO model, reflecting utility decrements associated with levels 1, 2, 3, 4 and 5 for each of the five domains: MO, SC, UA, PD and AD. This model will have same structure as equation 1 regarding the parameters for the level-attribute combinations, so it will be a 20 parameter model as well. The regression equation is given below.

\[
U_{js} = \beta_1 MO_{js} + \beta_2 SC_{js} + \beta_3 UA_{js} + \beta_4 PD_{js} + \beta_5 AD_{js} + \epsilon_{js}, \quad \ldots(2)
\]

where \( js \) will be the choice alternative in the choice sets.

As both TTO and DCE data provide information about the values of health states, we will also implement a hybrid modelling approach that will make use of both c-TTO and DCE datasets to estimate the potential value sets. This approach has been used in several national EQ-5D-5L valuation studies. The hybrid model will combine the likelihood functions of a linear model for the c-TTO data and a logit model for the DCE data. As the coefficients will be estimated from a conditional logit and expressed on a latent arbitrary utility scale, we will use a rescaled parameter \( \theta \), which will assume that the c-TTO model coefficients are proportional to DCE model coefficients. This method will combine the utility values elicited in the c-TTO for the 150 health states with utility values elicited in the DCE experiment for 196 pairs of states. We will use cluster estimation to acknowledge that for each participant included in the models, 10 c-TTO and 7 DCE responses are available. We will also estimate adjusted hybrid model which adjust the social demographical variables like age, sex etc.

**Reviewer 3- Specific comment 11:** Page 13, line 26. The statement about undertaking modelling in STATA should appear elsewhere (perhaps in the modelling section). This particular sub-section does not refer to any modelling work.

**Authors’ response:** We thank the reviewer for highlighting it. In compliance with the suggestion of the reviewer, we have moved this line to the starting of the modeling section (line 227 on page 13).

**Reviewer 3- Specific comment 12:** Page 17, lines 15-19. The ‘Patient and Public Involvement’ section contains no mention of any public involvement (or of the lack of public involvement).

**Authors’ response:** We thank the reviewer for seeking more clarification on this aspect. We would like to mention that no explicit public or patient involvement has been there in the designing of this protocol, therefore we have not mentioned anything in this section. As per the journal guidelines, the patient and public involvement is defined as involvement from patients or public in the design, conduct, reporting and dissemination of the research, which is distinct from patients and public being the participants in
the research. Public involvement in during the conduct and dissemination of the study will be strictly as per the defined ethical standards for research, and a mention of such involvement will be there in the manuscript discussing the outcomes of the study. Nevertheless, the ethical considerations which will be taken under the account during the conduct of the study, where the members of general public will be involved as participants, have been mentioned under the heading of 'Ethical Approval and Dissemination'.

In order to explicitly state the extent of involvement of public and patients in the protocol, we have added the following line:

Lines 437-439 on page 22: “No patient or public involvement was there in the designing of this research protocol. Public involvement during the conduct and dissemination of the study will strictly be as per the established standards of ethics in research.”

VERSION 2 – REVIEW

REVIEWER | Bram Roudijk  
| EuroQol Research Foundation  
| Netherlands  

REVIEW RETURNED | 23-Sep-2020

GENERAL COMMENTS | I would like to thank the authors for the revisions to this manuscript. In my opinion, the manuscript has been greatly improved and reads well. I think the manuscript would be nearly acceptable for publication. Before it can be accepted, I have some very minor comments that I would like to see addressed, which are aimed at improving the accuracy of the description of the research protocol.

Methods, subsection valuation methods:
“As a rational choice, the respondents would prefer good health (life A) over the state of inferior health (life B) when the time available in both the alternatives is equal (10 years).” This may be true, but we know that respondents do not always make rational choices. Furthermore, it can be that respondents are not willing to trade off any life years (for various reasons) and therefore do not prefer life A in this situation, but may consider the health states to be roughly equal. Because of this, I would replace this sentence with a sentence that focuses a bit more on the technical aspects of the cTTO, such as “If the respondent prefers life A, he or she is presented with the next question, in which he or she has to choose between dying immediately or living for 10 years in life B”. Or anything similar to that sentence. Then explain what happens if you pick B (move to 5 years in full health vs 10 years in state B), after which the rest of the current text of the paper can be placed.

“Nevertheless, there are certain health states, for which the respondent prefers to die immediately rather than living in that health state. These health states are known as worse than dead (WTD), and their valuation will be done with the help of c-TTO.” I would state that there are health states that respondents may prefer to die immediately than living in that health state, rather
than the current phrasing. We know that quite a large share of the respondents will not prefer death over any health state. Technically speaking, the cTTO is the combination of the traditional TTO (tTTO), which is used for better than dead health states, and lead time TTO (LTTO) for states worse than dead, where we add 10 years in full health to both alternatives, which will precede the 10 years in State B for alternative B. Therefore I would write that we use lead time TTO for states that respondents consider worse than dead.

Modelling:

"where refers to the \( Y_{it} \) TTO values dependent variable, \( \mu_0i \) will be the individual specific error component and \( \varepsilon_{it} \) refers to the combined time series and cross-section error component, i indicating the respondent, and t accounting for the panel structure of the dataset (because there are 10 cTTO questions per respondent)."

I would state that the Mu parameter is a respondent specific error component as you said, but would refer to the epsilon as the response-related error term rather than time series. These type of models are frequently used in time series analysis, but in our case, the sequence of presentation of the health states is irrelevant and not taken into account, we just take into account the clustered nature of the data.

Quality control:

"Interviews are flagged as non-compliant if the explanations for the first two c-TTO exercises last for less than 3 min,"

I think it would be valuable information to specify that interviews are flagged if the time taken to complete the two wheelchair example exercises is less than 3 minutes.

"if the duration of c-TTO tasks is less than 5 min,"

I think it is clearer if it specified that these are the 10 "real" cTTO questions. Else it may be unclear whether the practice exercises are included or not.

"or if the value given to the worse health state is not the lowest and at least 0.5 higher than"

I would include that state 55555 is always the worst state presented to every respondent.

"During this process, it will be observed if the interviewer is influencing respondent by its interviewing style, shortening the task due to laziness, or facing difficulty in explaining WTD element of the c-TTO task."

I think it will suffice to state that you will investigate whether the interviewer's behaviour is influencing the responses of the respondent, whether the interviewer explains the task well, and is not short-cutting. I would remove the word lazy from this sentence, as the interviewer may not be lazy but still shortcut the task if he or she does not understand well how such an interview should be performed.

**VERSION 2 – AUTHOR RESPONSE**

Comments by Reviewer-1:

Comments to the Author: I would like to thank the authors for the revisions to this manuscript. In my opinion, the manuscript has been greatly improved and reads well. I think the manuscript would be nearly acceptable for publication. Before it can be accepted, I have some very minor comments that I
would like to see addressed, which are aimed at improving the accuracy of the description of the research protocol.

**Reviewer 1 - Comment 1:** Methods, subsection valuation methods: “As a rational choice, the respondents would prefer good health (life A) over the state of inferior health (life B) when the time available in both the alternatives is equal (10 years).” This may be true, but we know that respondents do not always make rational choices. Furthermore, it can be that respondents are not willing to trade off any life years (for various reasons) and therefore do not prefer life A in this situation, but may consider the health states to be roughly equal. Because of this, I would replace this sentence with a sentence that focuses a bit more on the technical aspects of the cTTO, such as “If the respondent prefers life A, he or she is presented with the next question, in which he or she has to choose between dying immediately or living for 10 years in life B”. Or anything similar to that sentence. Then explain what happens if you pick B (move to 5 years in full health vs 10 years in state B), after which the rest of the current text of the paper can be placed.

**Authors’ response:** We thank the reviewer for suggesting this change to make the explanation comprehensive. Complying with the suggestion of the reviewer, we have rewritten these lines. The revised text is read as follows:

Lines 185-192 on pages 11: “If the respondent prefers life A, he or she will be presented with the next question, in which he or she will have to choose between dying immediately or living for 10 years in life B. If the respondent prefers living in an inferior health state (as described in life B) over the immediate death (life A), he or she will be presented with the next question, wherein 5 years in full health will be provided in life A, and 10 years in some inferior health state will be provided in life B. Hence, the time available in life B will be kept constant at 10 years, while the time available in life A will be changed sequentially, and the respondent will be asked to select the better alternative between life A and life B.”

**Reviewer 1 - Comment 2:** “Nevertheless, there are certain health states, for which the respondent prefers to die immediately rather than living in that health state. These health states are known as worse than dead (WTD), and their valuation will be done with the help of c-TTO.” I would state that there are health states that respondents may prefer to die immediately than living in that health state, rather than the current phrasing. We know that quite a large share of the respondents will not prefer death over any health state. Technically speaking, the cTTO is the combination of the traditional TTO (tTTO), which is used for better than dead health states, and lead time TTO (LTTO) for states worse than dead, where we add 10 years in full health to both alternatives, which will precede the 10 years in State B for alternative B. Therefore, I would write that we use lead time TTO for states that respondents consider worse than dead.

**Authors’ response:** We completely agree with the reviewer on this aspect, and revised the text in the manuscript. The revised text reads as follows:

Lines 204-209 on page 12: “Nevertheless, there are certain health states, for which the respondent may prefer to die immediately rather than living in that health state. These health states are known as worse
than dead (WTD). We will use lead time TTO (L-TTO) for health states that respondents consider worse than dead. The c-TTO approach is a combination of the conventional TTO (which is used for better than dead health states) and L-TTO (which is used for states worse than dead).

**Reviewer 1 - Comment 3:** Modelling: “where refers to the $Y_{it}$ TTO values dependent variables $\mu_{0i}$ will be the individual specific error component and $\varepsilon_{it}$ refers to the combined time series and cross-section error component, indicating the respondent, and $t$ accounting for the panel structure of the dataset (because there are 10 cTTO questions per respondent).” I would state that the Mu parameter is a respondent specific error component as you said, but would refer to the epsilon as the response-related error term rather than time series. These types of models are frequently used in time series analysis, but in our case, the sequence of presentation of the health states is irrelevant and not taken into account, we just take into account the clustered nature of the data.

**Authors’ response:** We thank the reviewer for highlighting this point. Complying with the suggestion of the reviewer, we have edited the text. The revised text is as follows:

Lines 240-243 on page 14: “where $Y_{it}$ refers to the TTO values dependent variable, $\mu_{0i}$ will be the respondent specific error component and $\varepsilon_{it}$ refers to the response-related error term, $i$ indicating the respondent, and $t$ accounting for the panel structure of the dataset (because there are 10 cTTO questions per respondent).”

**Reviewer 1 - Comment 4:** Quality control: “Interviews are flagged as non-compliant if the explanations for the first two c-TTO exercises last for less than 3 min,”

I think it would be valuable information to specify that interviews are flagged if the time taken to complete the two wheelchair example exercises is less than 3 minutes.

“If the duration of c-TTO tasks is less than 5 min,” I think it is clearer if it specified that these are the 10 “real” cTTO questions. Else it may be unclear whether the practice exercises are included or not.

“or if the value given to the worse health state is not the lowest and at least 0.5 higher than” I would include that state 55555 is always the worst state presented to every respondent.

**Authors’ response:** We thank the reviewer for mentioning these additions, as it would increase the accuracy of the description. In compliance with the suggestion of the reviewer, we have edited these sentences. The revised sentences are as follows:

Lines 317-322 on page 17: “Interviews are flagged as non-compliant if the explanations for the two wheelchair example exercises last for less than 3 minutes, if the worse than dead element is not shown in the examples, if the duration of the 10 real c-TTO tasks is less than 5 min, or if the value given to the worst health state (health state ‘55555’, which is always the worst state presented to every respondent) is not the lowest and at least 0.5 higher than that of the state with the lowest value.”
Reviewer 1- Comment 5: “During this process, it will be observed if the interviewer is influencing respondent by its interviewing style, shortening the task due to laziness, or facing difficulty in explaining WTD element of the c- TTO task.” I think it will suffice to state that you will investigate whether the interviewer’s behavior is influencing the responses of the respondent, whether the interviewer explains the task well, and is not shortcutting. I would remove the word lazy from this sentence, as the interviewer may not be lazy but still shortcut the task if he or she does not understand well how such an interview should be performed.

Authors’ response: We thank the reviewer for this comment. As per the suggestion, we have revised the text in the manuscript. The revised lines are as follows:

Lines 332-334 on page 18: “During this process, we will investigate whether the interviewer’s behavior is influencing the responses of the respondent, whether the interviewer explains the task well, and the interviewer is not shortcutting the task.”

VERSION 3 – REVIEW

<table>
<thead>
<tr>
<th>REVIEWER</th>
<th>Bram Roudijk</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EuroQol Research Foundation</td>
</tr>
<tr>
<td></td>
<td>The Netherlands</td>
</tr>
<tr>
<td>REVIEW RETURNED</td>
<td>27-Oct-2020</td>
</tr>
</tbody>
</table>

| GENERAL COMMENTS | I would like to thank the authors for their revisions. I think the manuscript is of very high quality and I have no further comments. In my opinion, the manuscript is ready to be accepted for publication. |