Kate

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Number | Location | Comment | Response | Action |
| K1 | Email | Some of the U\_1 and U\_2 and (U\_0) in the main paper is in maths format (italics). Some of them are not, please check and make them consistent. |  | Check and make consistent [TO DO] |
| K2 | Email | In Legends for figures.docx, Us were written as U1 and U2, so please make a change. |  | Change and make consistent [TO DO] |
| K3 | Email | Title, including abbreviated title: methods for simulating monotonically related variables in health economic models when performing PSA. (Not sure about abbreviated title, do we need one?) | Abbreviated title required | Change main title to suggested [TO DO] |
| K4 | Email | Keywords  Probabilistic sensitivity analysis  Health economic models  Simulation  Monotonicity |  | Add this as keywords [TO DO]  Produce an additional document including  Keywords  Title  Abbreviated title  Etc [TO DO] |
| K5 | Email | Whether you think the big table describing all the methods should be an appendix rather than part of the main document – may be in the appendix since the main paper is already quite length |  | Moved to appendix |
| K6 | Ab  P1:8 | Spell out standard error |  |  |
| K7 | Abs  P1:12  KR5 | A supplementary analysis of what is performed? |  |  |
| K8 | Abs  P1:14 | Typo: Add ‘not |  |  |
| K9 | Abs  P1:21  KR6 | Shall we add that we have only looked at how to sample two monotonically related parameters? |  |  |
| K10 | Main  P4:1  KR8 | Can we say what the 10 methods are somewhere in the introduction?  There is a subsection called The Ten Methods. Will it make more sense to have it in the introduction section? |  |  |
| K11 | Main  P4:9 | Is HRQoL more commonly used? |  |  |
| K12 | Main:  P4:15 | I think it’s better to delete this. Unless you want to explain what x, y and 2 are |  |  |
| K13 | Main  P6:14 | Subscript on U1 and U2 |  |  |
| K14 | Main:  P10:12  KR29 | Any reason for not just using =? |  |  |
| K15 | Main  P11:17 | Why not just \delta? | Make clear this is a distribution of estimates |  |
| K16 | Main  P11:18 | Be more specific, e.g. in section XXX |  |  |
| K17 | Main  P17:1 | Subscripting on U1 and U2 |  |  |
| K18 | Main  P18:14 | Subscripting on U1 and U2 |  |  |
| K19 | Main  P19:13 | Be consistent about whether HRQoL or HRQL used |  |  |

Matt

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Number | Location | Comment | Response | Action |
| M1 | Email | Short title should be Handling monotonicity in PSA |  |  |
| M2 | Email | Keywords: add probabilistic and move monotonicity to the top |  |  |
| M3 | Abs  P1:3-4  M2 | Remove ‘and ultimately suboptimal decisions;  “Invites criticism and may not be true. The adoption decision may be the same for all methods. |  |  |
| M4 | Abs  P1:12 | Edit: “further analysis using a separate dataset were performed” |  |  |
| M5 | Abs  P1:21:  M7 | If we can do under the word limit then I agree |  |  |
| M6 | Abs  P2: 3-4 | Slight edits to text |  |  |
| M7 | Main  P4:1  M9 | I think it is better in methods, but moved earlier as highlighted |  |  |
| M8 | Main  P4:9  M12 | Yes, change all |  |  |
| M9 | Main  P4:15  M14 | Agreed |  |  |
| M10 | Main  P4:18 | Change of text |  |  |
| M11 | Main  P5:1 | This needs amending throughout. Are we saying less severe, thus U1 > U2, or are we saying less severe or equivalent to? |  |  |
| M12 | Main  P5:8-21  M19 | I still have concerns other this similar to those raised by Sophie. The analyst will not have the IPD, so the bootstrapping is an unknown parameter. If different data sets have the same summary statistics, but different bootstrapped values, then similarly to our chosen gold standard is a weak criteria. In the extreme the same random number could be correct.  I would keep this in, but think we need to downplay this much more. The reviewers may ask for a hypothetical data set that differs, but with the same summary value, but I would wait for them to ask, unless we have time to do this in advance and discuss the conclusions  Given this, it might be worth adding in text regarding the validation of the estimates with experts in the field. |  |  |
| M13 | Main  P 6:9-10  M21 | I think the text made it hard to understand as we haven’t seen table 1 yet. |  |  |
| M14 | Main  P7:6  M22 | This should go before the summary states and the bootstrapped estimates |  |  |
| M15 | Main  P7:23 | Don’t want to get into the probability with two wide distributions with close means the prob could approach 0.5 |  |  |
| M16 | Main  P8:5 | Add “(methods three to size)” |  |  |
| M17 | Main  P8:12  M27 | This should come after the more detailed descriptions of the methods. |  |  |
| M18 | Main  P9:2-4 | Add text “Which ensures monotonicity when the overlap between the distributions is confined to the upper tail of one distribution and the lower tail of the second distribution. |  |  |
| M19 | Main  P9:20 | Add ‘in this paper’ |  |  |
| M20 | Main  P10:20  M30 | “Too vague. Did we use criteria for increasing the CV parameter. If not, can we make it look like we did?” |  |  |
| M21 | Main  P10:21 | Change ‘either if’ to ‘if either’ |  |  |
| M22 | Main  P11:6 | Edit: add “methods seven, eight and nine” |  |  |
| M23 | Main  P11:18  M33 | Agree with Kate. But it may be that we want to have this section immediately following? |  |  |
| M24 | Main  P12:11  M34 | Play this down! |  |  |
| M25 | Main  P12:22-23 | Slight re-editing |  |  |
| M26 | Main  P13:6-14 | Slight re-editing |  |  |
| M27 | Main  P14:5-6  M39 | This is contradicted with discussion on Method 2 |  |  |
| M28 | Main  P14:14-15 | Deletion |  |  |
| M29 | Main  P15:7  M40 | Linked ot the main point, this bootstrapped mean is not known. I would compare to the known mean. |  |  |
| M30 | Main  P15:15-7 | Slight editing |  |  |
| M31 | Main  P15:25 | Slight editing |  |  |
| M32 | Main  P16:3-4 | What is the point being made here? You’ve written it as though it is important, but I need it spelt out a bit more, with relation to the comment on unknown bootstrapped data. |  |  |
| M33 | Main  P17:9-17 | Various edits. Comments’ Repitition’ |  |  |
| M34 | Main  P17:19-25 | I think this is two points and the advantages of one and two are the same. |  |  |
| M35 | Main  P18:5 | Provide URL |  |  |
| M36 | Main  P18:7-21 | Various edits. |  |  |
| M37 | Main  P19:4-18 | Various edits. |  |  |
| M38 | Main | Change ‘is available from the authors’ to ‘is available online’. |  |  |

Sophie

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Number | Location | Comment | Response | Action |
| S1 | Email | Any to (or more) parameters could be monotonically related in a model. Although HRQoL provides a convenient and common example I think this paper should relate to any to parameters so be good to make definitions and text more general. Other examples could be costs or transition probabilities so would be good to mention such examples too |  |  |
| S2 | Email | I think testing methods with additional data sets would be worthwhile and perhaps wouldn’t take long now you have the code? |  |  |
| S3 | Email | I think the source of the methods should be made clearer - with examples of where the methods have been used previously within HTA |  |  |
| S4 | Email | I like Kate’s title |  |  |
| S5 | Abs, p 1,  U1 | What is the title? |  |  |
| S6 | Abs,  P1,  U3 | To compare the appropriateness of ten methods which could be used – “ Vague – what do we really mean here? I think it is important that this is clearly defined. |  |  |
| S7 | Abs  P1,  U4 | The ten methods compared … “Say how the ten methods were identified” |  |  |
| S8 | Abs  P2,  l.4-5 | Suggestion: It does not appear discontinuities or biases, it produces consistent result and it appears to implement |  |  |
| S9 | Main  P4:6 | Are known.. to be monotonically related. |  | Remove? |
| S10 | Main  P4:7-8 | I don’t think this is a good example. There are probabily diseases for which more severe is associated with less aggressive treatment so less AUs and higher QoL. Perhaps change to two levels of symptoms? |  |  |
| S11 | Main  P4:16  U15 | Briefly mention other examples of parameters that could be monotonically related e.g. costs or transition probabilities |  |  |
| S12 | Main  P4:21  U16 | I would defined independent sampling in background here |  |  |
| S13 | Main  P4:22  U17 | It seems to want to be a general definition. If so it should not relate to the HRQL example. More generalizable and better if we say something like this: we have two parameters in the PSa and we know they are monotonicallty related U1>=U2.. |  |  |
| S14 | Main  P5:9-21 | Similarly to definition above I think this need to be more general and not specifically relate to health states. |  |  |
| S15 | Main  P7:8 | Need references and examples of use here.  Also were there any methods identified that were not tested here? |  |  |
| S16 | Main  P7:6 | Only really 9 as number one does not really consider monotonicity at all |  |  |
| S17 | Main  P7:10-21 | I would delete this as quite repetitive to have methods listed here and below and in table. |  |  |
| S18 | Main  p.8:20 | Change header ‘simple and naïve methods’  To ‘method one, independent sampling’ |  |  |
| S19 | Main  P9:1 | Add header ‘method two quantile matching’ |  |  |
| S20 | Main  P9  U28 | I would include the names of the methods in the figures |  |  |
| S21 | Main  p9:6 | Add header ‘methods three (upward replacement) and method four (downward replacement’ |  |  |
| S22 | Main  P9:10 | Add header: method 5 upward resampling and method 6 downwards resampling |  |  |
| S23 | Main  PP12-13:22-3 | I would replace with: Comparison of the methods via visual inspection of scatterplots and violin plots is presented. For the three most promising data sets comparison using the second hypothetical data set is also presented |  |  |
| S24 | Main  P13:5-20 | Is this section important or should it be in an appendix? |  |  |
| S25 | Main  P13:20 | Several of the figures would benefit from having a commentary in addition to a title and key |  |  |
| S25 | Main  P13:21 | Remove header: ‘First hypothetical dataset, all ten methods’ |  |  |
| S26 | Main  P14:2-3 | Move following to methods: ‘A good method should be able to produce a similar pattern of scatter given the aggregate data as the bootstrapped method is able to produce using the IPD. | Need to restate what counts as good here? |  |
| S27 | Main  P14:7-17 | Remove subfigure refs Fig 4dm e, etc |  |  |
| S28 | Main  P15:1-3 | Rewrite: Figure 5 present the distributions of U1, U2 and U1-U2 |  |  |
| S29 | Main  P15:9 | Remove: The bottom of the three subfigures, Figure 5c, shows… |  |  |
| S30 | Main  P15:19  U41 | Figure 5 needs axis labelling and a key adding  I would also suggest some commentary with figure.  E.g. methods 8 and 10 match the bootstrapped data the most closely |  |  |
| S31 | Main  P15:19  U41 | Why are diagrams all double? i.e. symmetrical around y axis – seems unnecessary but maybe it is convention or something? |  |  |
| S32 | Main  P15:24-5  U42 | The analysis (Second hypothetical dataset, three best methods) needs to be described in the methods section |  |  |
| S33 | Main  P16:20-21  U44 | This is true but I think it is known and obvious and should be in background and not in findings |  |  |
| S34 | Main  P16:21-22  U45 | This statement is too strong. If means were not close and sdeviations small then independent sampling would be OK I think? |  |  |
| S35 | Main  P17:1-2 | Rewrite: ‘methods using resampling or replacement (3-6)’ |  |  |
| S36 | Main  P17:5-8 | Rewrite lines about uncertainty |  |  |
| S37 | Main:  P18:5 | Provide in online appendix too? |  |  |
| S38 | Main  P18:9  U50 | I would include a section which looks at the feasibility of extension to three or more states. Either in methods or discussion. |  |  |
| S39 | Main  P18:12  U51 | If I was reviewing I would ask for more hypothetical data sets to be tested. Surely this would be quite quick to do now you have the code? |  |  |
| S40 | Main  P19:10  U52 | Good section but would also like it to relate to general monotonic parameters .e.g costs or transition probability examples |  |  |
| S41 | Main:  P19:16-18 | To me it is obvious that the choice could impact on the results and further empirical research is not needed to verify this. |  |  |