Reviewer 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Comment Code** | **Location** | **Comment** | **Response** | **Action Taken** |
| R1C01 | All | My major concern is that the paper seems unnecessarily complex. | The number of procedures required to perform the method in R is relatively few, and easy to generalize. However we agree fully if the suggestion is that the paper makes the approach seem more complex than it is in practice. Tighter, clearer descriptions of the processes and a clearer motivation and illustrative example will be provided to try to address this. |  |
| R1C02 | All | For example, I see no need for simulation at all. If uj is the mean utility for respondents occupying state j of the mRS, and pj is the probability that a randomly selected patient reports state j of the mRS, then the mean utility for, e.g. , independent stroke is p\*’U\*, where p\* is a three dimensional vector indicating the probability that a respondent having had an independent stroke reports mRS equal to 0, 1, or 2, and U\* is the three dimensional vector giving the mean utilities for these three mRS states.  Then the variance of the mean utility for independent stroke is var(p\*’u\*) = E(var(p\*’ u\* | p\*) + var(E(p\*’u\* | p\*) = p\*’var(u\*)p\* + u\*’var(p\*) u\* (by the law of iterated expectation) Var(u\*) should be available from the data that collected the EQ-5D utilities, and var(p\*) can be derived by noting that if your prior distribution for p\* is Dirichlet(a), then given the data the posterior distribution for p\* is Dirichlet(n\* + a), where n\* is the 3-dimensioanl vector giving the numbers of people with mRS states equal to 0, 1, or 2. The variance-covariance matrix of the Dirichlet distribution (needed to calculate var(p\*)) is then a simple function of n\* + a. | We thank the reviewer for a proposed closed form solution which provides estimated means and variances. |  |
| R1C03 | All | As a general point, simulation should not be used when closed-form estimators like this are available. | We have intended to approach to be illustrative of a broader family of simulation based approaches, including some where a closed form solution is not obvious. The simulation based approach is also effective in generating sample estimates for use in PSA, which was the practical motivation for us, as the paper stems for a health technology assessment where PSA is mandatory. |  |
| R1C04 | Introduction | The introduction seems a bit light. Could you say something to set the problem within the wider context – what kinds of scenarios is this useful for, what methods are currently, what are the problems with them etc? | We would be happy to do so. We will discuss more about the specific application this approach was developed for. |  |
| R1C05 | P05  L29 | I wasn’t sure why OLS was used. Were the estimates adjusted for anything? If you’re just calculating mean utility for each state, why just not use the usual mean and its standard error? |  |  |
| R1C06 | P06  S 1.2.4 | I had reservations about assuming perfect correspondence between GOS and the mRS – this seems like a strong assumption. Could you cite some evidence that this correspondence exists? | We provide the verbal descriptions of the categories, both ranked in order of severity. Our aim in the paper is mainly to demonstrate an approach which makes the implications, to estimated model outputs, of making a set of statistical and mapping assumptions clearer, rather than to justify the assumptions per se. |  |
| R1C07 | P08  S 1.2.12 | I was a little confused by this section. In statistical terminology, expected value is the same as the mean. I think I know what you’re doing, but using language more precisely would be helpful. | We apologise for inprecision in the terminology here and elsewhere.  We believe in this case there is some ambiguity within usual statistical terminology, which may have made the imprecision here worse: If mu\_pop is the population mean, and {X} = {x1, x2, …, xn} a collection of observations from which to estimate mu\_pop, then E({X}) ~ N(mu\_sample, var). In this context the expected value E({X}) is a distribution of values, which is distinct from the mean mu\_sample (a point) or its expectation E(E({X})) (also a point, and in this case equal to mu\_sample). |  |
| R1C08 | P11  L06 | Re : the normal distribution, I think you only need the means to have a normal distribution, not individual utilities, right? If this is the case, then you could appeal to the Central Limit Theorem |  |  |
| R1C09 | All | I’m curious as to why you chose a multiplicative approach rather than an additive approach. |  | ACTION FOR MATT |

Reviewer 2

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Comment Code** | **Location** | **Comment** | **Response** | **Action Taken** |
| R2C01 | All | This paper presents a simulation approach to developing uncertainty for the conversion of scores on one scale into those on another. The method is useful but the paper presents it very poorly. |  |  |
| R2C02 | All | The approach needs better motivation – why was the mapping required? Why couldn’t the mRS be used directly? I presume it is because the trial didn’t assess mRS but that needs to be said explicitly up front. |  |  |
| R2C03 | All | Once you make an assumption about how one scale maps to another, the scoring is simple arithmetic. The paper is really about trying to incorporate uncertainty, but this is not clearly stated. |  |  |
| R2C04 | All | There is also insufficient explanation for many of the methodological statements. |  |  |
| R2C05 | All | The authors need to decide if this paper is about the methods or about the specific example they present – this is not clear. |  |  |
| R2C06 | All | I suspect the underlying model is a cohort Markov – never stated – but if so, why bother when you have created an individual simulation to derive the uncertainty? Why not stick with the individual simulation throughout? |  |  |
| R2C07 | Abstract | Abstract is poorly written and doesn’t reflect the paper. The results are improvements in stroke care but the paper is about converting one scale to another. |  |  |
| R2C08 | P03  L11 | What does reported using different systems mean? |  |  |
| R2C09 | P04  L26 | References 3 and 4 are to echocardiography papers – how do they related to the topic of this one? |  |  |
| R2C10 | P04  L27 | Unclear how this improves ‘consistency’ and ‘validity’. |  |  |
| R2C11 | P05  L02 | What do mean utilities have to do with mapping one scale into another? Wouldn’t any valuation do? Costs for example? |  |  |
| R2C12 | P05  L07 | Isn’t this assumption the crux of the matter? And once you make it, the method described is just about deriving uncertainty, not really about mapping. |  |  |
| R2C13 | P05  L14 | Relevance to this paper if inter-rater reliability of the mRS? |  |  |
| R2C14 | P05  L26 | “For simplicity” of what? How does this relate to the paper? |  |  |
| R2C15 | P06  L15 | Unclear what “method” is being applied in the figure |  |  |
| R2C16 | P06  L20 | Unclear what process is repeated 10,000 times.  Why 10,000? Why not 1,000? How was this implemented? |  |  |
| R2C17 | P07  L03 | Why normal? |  |  |
| R2C18 | P07  L14 | It is likely that other determinants differ between patients who had a stroke with mRS 0 and other levels so it is not obvious that it is valid to take that as the background utility |  |  |
| R2C19 | P07  L21 | Why multiplicative? Explain the rationale |  |  |
| R2C20 | P07  L32 | Unclear what estimates you refer to |  |  |
| R2C21 | P07  L32 | “For this reason” does not follow from preceding sentence |  |  |
| R2C22 | P08  L11 | Unclear how the “uncertainty in the true proportion of each component state” was taken into account |  |  |
| R2C23 | P09  L19 | How does this mean that estimates are “replicable”? |  |  |
| R2C24 | P09  L20 | How does this minimize number and influence of assumptions? |  |  |
| R2C25 | P09  L21 | Weights are not necessarily “sample size” |  |  |
| R2C26 | P09  L23 | Unclear how more recent data are to be used |  |  |
| R2C27 | P09  L24 | Claim of usefulness only holds if the mapping assumption is reasonable |  |  |
| R2C28 | P10  L02 | Unclear how “Greater consistency” is attained |  |  |
| R2C29 | P10  L03 | What would constitute “other similar situations”? |  |  |
| R2C30 | P10  L03 | What is the “right form”? |  |  |
| R2C31 | P10  L06 | How does this “show the implications of making simple and standard assumptions”? |  |  |
| R2C32 | P10  L06 | What are “standard assumptions”? by whose standards? |  |  |
| R2C33 | P10  L08 | Why are “mean utility scores … normally distributed”? |  |  |
| R2C34 | P10  L12 | What “intermediate state” is referred to? |  |  |
| R2C35 | P10  L14 | In what sense is this about “reliability”? |  |  |
| R2C36 | P10  L15 | If it is difficult to assess appropriateness, what are we to make of this whole thing? |  |  |
| R2C37 | P10  L31 | This is the first time in the paper that the rationale is provided. This should be in the introduction. |  |  |
| R2C38 | P11  L07 | Unclear what “general limitation” is referred to or what justifies labeling a poor choice of distribution this way |  |  |
| R2C39 | P11  L12 | The structural sensitivity analyses should be performed as part of this work. Not left to readers. |  |  |
| R2C40 | P11  L17 | Perfect mapping is very unlikely |  |  |
| R2C41 | P11  L23 | Unclear how these sensitivity analyses should be done |  |  |
| R2C42 | P11  L29 | Assumption about deaths is irrelevant to this paper |  |  |
| R2C43 | P12  L01 | “potential issues”? meaning? |  |  |
| R2C44 | P12  L03 | What cost data are you referring to? |  |  |
| R2C45 | P12  L09 | Unclear what research is recommended. As long as analysts are willing to make the mapping assumptions, the approach cannot fail. It simply provides uncertainty estimates |  |  |
| R2C46 | P12  L14 | “Clear interoperability between costs and utility summaries” – huh?! |  |  |
| R2C47 | P13  L05 | What “alternatives”? |  |  |
| R2C48 | P18  T3 | Why are credibility intervals here but not for T1? |  |  |
| R2C49 | P18  T3 | Why are there decimal places in the distribution values? |  |  |
| R2C50 | P19  T4 | Why are Bayesian intervals used? What is their basis? |  |  |
| R2C51 | P21  F1 | Fonts in the figures are so small they are illegible |  |  |
| R2C52 | P24  L14 | What does PSA have to do with this paper? |  |  |
| R2C53 | P03  L08 | Delete “different ant” – if they are incompatible they must be different. |  |  |
| R2C54 | P03  L10 | Not all consequences are “health-related QOL’ – delete parenthetical phrase |  |  |
| R2C55 | All | Unnecessary to number the entire manuscript thus making all the sections secondary at best. Delete 1. Manuscript |  |  |
| R2C56 | P04  L04 | “where the frequency of patients in each outcome” is unclear. |  |  |
| R2C57 | P04  L14 | “on the for”? |  |  |
| R2C58 | P04  L15 | “In an MDM paper” – replace by the citation |  |  |
| R2C59 | P07  L04 | Delete “from” – sampled by itself is sufficient |  |  |
| R2C60 | P10  L06 | ‘show’ should be ‘shows’ |  |  |
| R2C61 | P10  L12 | Delete ‘hypothetical’ – simulations are always hypothetical |  |  |
| R2C62 | P11  L18 | ‘bijection’? |  |  |
| R2C63 | P12  L12 | Incomplete sentence |  |  |
| R2C64 | P14 | Citation 11 is a repeat of 2 |  |  |
| R2C65 | P24 | Why is appendix labeled 1.1 |  |  |
| R2C66 | P24  L27 | ‘Report’ should be ‘reported’ |  |  |

Reviewer 3

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Comment Code** | **Location** | **Comment** | **Response** | **Action Taken** |
| R3C01 |  | The introduction needs to make it clear from the outset that the paper is about modelling uncertainty correctly (which it appears to be). Just stating that the new approach has different steps isn't helpful.  At the end of the introduction it is mentioned that the technique was developed in undertaking what appears to be a cost-effectiveness analysis. The last sentence states that' to be able to derive estimates of the consequences...' Estimates could have been produced without this new approach, so it is 'more precise estimates' or 'estimates with accurately estimated uncertainty' that seems to be the point. |  |  |
| R3C02 |  | It is also slightly disappointing, though not essential to the paper, that the effect on the cost-effectiveness results (the mean and uncertainty) are not investigated. |  |  |
| R3C03 |  | Is there a closed-form solution? I think the authors do need to address this point. |  |  |
| R3C04 |  | Since the paper appears to be about appropriately incorporating uncertainty, it seems to me that it would have been useful to compare the results using the new approach to those without capturing all the uncertainty in the results section: the impact on the estimates of utility multipliers and the uncertainty around those estimates. I.e. what has been gained from the extra complexity? As it stands, it just illustrates the implementation of the approach. |  |  |
|  |  | Much of the methods are clearly described but this highlights areas which are not, such as section 1.2.7. Be explicit about why you're producing a large number of simulated distributions, and is it sampling with replacement? |  |  |
|  |  | There is an unfinished sentence in the Further research section. |  |  |
|  |  | Conclusions: did the method 'appear' to provide updated estimates or did it in fact do that? Conclusions are drawn about the appropriateness of grouping mRS categories into GOS categories and yet this conclusion seems quite divorced from the thrust of the paper. A clearer introduction and argument would be help here. |  |  |