This paper presents a simulation approach to developing uncertainty estimates for the conversion of scores on one scale into those on another. The method is useful but the paper presents it very poorly. 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. 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. There is also insufficient explanation for many of the methodological statements. The authors need to decide if this paper is about the methods or about the specific example they present – this is not clear.

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?

Major

Page	Line	Comment
3		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.
	11	What does "reported using different systems" mean?
4	26	References 3 and 4 are to echocardiography papers – how do they relate to the topic of this one?
	27	Unclear how this improves "consistency" and "validity"
5	2	What do mean utilities" have to do with mapping one scale into another? Wouldn't any valuation do? Costs for example?
	7	Isn't this assumption the crux of the matter? And once you make it, the method described is just about deriving uncertainy, not really about mapping.
	14	Relevance to this paper of inter-rater reliability of the mRS?
	26	"For simplicity" of what? How does this relate to this paper?
6	15	Unclear what "method" is being applied in the Figure
	20	Unclear what process is repeated 10,000 times. Why 10,000? Why not 1,000? How was this implemented?
7	3	Why normal?
	14	It is likely that other determinants differ between patients who had a stroke with mRS 0 and the other levels so it is not obvious that it is valid to take that as the background utility
	21	Why multiplicative? Explain the rationale
	32	Unclear what estimates you refer to
	32	"For this reason" does not follow from preceding sentence
8	11	Unclear how the "uncertainty in the true proportion of each component state" was taken into account
9	19	How does this mean that estimates are "replicable"?
	20	How does this minimize number and influence of assumptions?
	21	Weights are required not necessarily "sample size"

	23	Unclear how more recent data are to be used
	24	Claim of usefulness only holds if the mapping assumption is reasonable
10	2	Unclear how "greater consistency" is attained
	3	What would constitute "other similar situations"?
	3	What is the "right form"?
	6	How does this "show the implications of making simple and standard assumptions"?
	6	What are "standard assumptions"? By whose standards?
	8	Why are "mean utility scores…normally distributed"?
	12	What "intermediate stage" is referred to?
	14	In what sense is this about "reliability"?
	15	If it is difficult to assess appropriateness, what are we to make of this whole thing?
	31	This is the first time in the paper that the rationale is provided. This should be in the introduction.
11	7	Unclear what "general limitation" is referred to or what justifies labelling a poor choice of distribution this way
	12	The structural sensitivity analyses should be performed as part of this work, not left to readers
	17	Perfect mapping is very unlikely
	23	Unclear how these sensitivity analyses should be done
	29	Assumption about deaths is irrelevant to this paper
12	1	"potential issues"? meaning?
	3	What cost data are you referring to?
	9	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.
	14	"clear interoperability between costs and utility summaries" – huh?!
13	5	What "alternatives"?
18	T3	Why are there credibility intervals here but not for T1?
		Why are there decimal places in the distribution values?
19	T4	Why are Bayesian intervals used? What is their basis?
21	F1	Fonts in the figures are so small, they are illegible
24	14	What does PSA have to do with this paper?

Minor

Page	Line	Comment
3	8	Delete "different and" – if they are incompatible they must be different
	10	Not all consequences are "health-related qol" – delete parenthetical phrase
4	1	Unnecessary to number the entire paper as 1. thus making all the sections secondary at best. Delete 1. Manuscript
	4	"where the frequency of patients in each outcome" is unclear
	14	"on the for"?
	15	"in an MDM paper" – replace by the citation
7	4	Delete "from" – sampled by itself is sufficient
10	6	"show**s**"

	12	Delete “hypothetical” – simulations are always hypothetical
11	18	“bijection”?
12	12	Incomplete sentence
14		Citation 11 is a repeat of 2
24		Why is Appendix labelled 1.1.?
	27	“report**ed**”