Author's response to reviews

Title: Misspecification of the D1 EQ-5D regression model

Authors:

Kim Rand-Hendriksen (kim.rand-hendriksen@medisin.uio.no)
Liv Ariane Augestad (liv.ariane@gmail.com)
Fredrik A. Dahl (f.a.dahl@ahus.no)

Version: 2 Date: 21 November 2011

Author's response to reviews: see over
Dear Editor,

We appreciate the opportunity to submit a revised version of our paper “Misspecification of the D1 regression model”. The reviewers’ comments were helpful, and we have tried to accommodate their suggestions. Below, we have listed the reviewers’ comments one by one. Our responses are italicized. We hope you will consider the revised text for publication in Population Health Metrics.

Sincerely,

Kim Rand-Hendriksen

Responses to the reviewers’ comments

Reviewer 1, Richard Norman.

1. The authors show that the D1 approach can be converted into something closer to the more usual algorithm involving a constant (Table 2). Does the same argument hold for I3 and the level 3 terms? If the 0.122 term was added to the M3, S3, U3, P3, A3 terms, and an N1 term introduced, does this improve specification (and reduce collinearity)? If so, this would change the author’s conclusion regarding the relative sizes of the disutility associated with the level 2 and level 3 terms, as the level 3 coefficients would be deflated in the same way as the level 2 coefficients were). If it doesn’t hold, the authors should explain why.

The same argument holds for I3 and the level 3 dummies. Substituting I3 for N3 gives identical predicted values for all EQ-5D health states. Furthermore, N3 would have the same directionality as the level 3 coefficients, substantially reducing their magnitude. We have included a paragraph describing this, at the end of the discussion, right before the conclusions. We have chosen not to mention this at earlier points in the paper because we feel that it complicates the presentation, and switching between I3 and N3 doesn’t change the statistical significance of any variables. Since the level 3 coefficients would be deflated using the N3, the magnitude of the difference between the level 2 and level 3 coefficients would be smaller using using the N3. However, they do not approach zero in the way the level 2 coefficients do when using the constant term.

2. Variable Inflation Factors. I think this is an interesting way of looking at the issue, but want more detail. The authors argue that the use of D1 makes the independent variables in the regression much more collinear than under the standard constant approach. This seems sensible given that the D1 term has a strong relationship with almost every other parameter in the regression. I would like more opinion regarding whether this makes the standard errors (and hence the statistical significance) of the other terms in the regression unreliable. They have stated that the statistical significance of some of the parameters drops out when
the D1 term is replaced in the D1c model; is this partially because the standard errors are artificially high? If this doesn’t matter, or can’t be interpreted in this way, I would like the author’s views on why.

*We have recalculated the D1c standard errors. Previous tests indicated that SE values for the two models were identical, which surprised us. Performing the tests from scratch resulted in different SE measures, while the coefficients were unchanged. This is in line with what one would expect given different levels of collinearity, and suggests that we must have done something wrong in the previous tests. The standard errors in both models are inflated due to multicollinearity, but are smaller using the D1c specification. Correcting for this, only two of the level 2 coefficients are definitely nonsignificant. This has been changed throughout.*


*Done.*

4. p.7 “242 health states”. This is correct, but the idea behind omitting 11111 has not yet been introduced. Just needs a note saying ‘as discussed below’ or stating the reason why 11111 is not included.

*Changed to “243 health states” in accordance with the suggestion by reviewer 2. After the explanation, 242 is used.*

5. p.5 Issue of cost consideration dictating that only a subset of health states be valued. This is certainly one of the reasons why the subset was valued. However, it is highly questionable that it is a valid issue. The sample sizes of the UK and US studies (but also the smaller ones) would allow a good coverage of all 243 health states for exactly the same cost. For example, if you have 1,000 people each valuing 12 health states, that means you have 12,000 observations. If these are spread over the 242 non-11111 health states, you have almost 50 observations per health state.

*We agree. The sentence has been altered to read “Since directly valuing all EQ-5D health states was considered impractical, EQ-5D valuation studies have typically elicited values for a subset of the EQ-5D health states...”.*

6. p.10 The ease of utility calculation. This is true, but probably not a massive point. I don’t think the D1 algorithm is particularly complicated and, given that the ease of utility calculation can be easier or harder under their proposed specification makes the point minor in my opinion. The point is much less important than the finding that the level 2 coefficients
become not statistically significant.

*We agree. These analyses were included because the authors of the US valuation study used simplicity as an argument for including the D1 variable. We now point out that this is a less important point in the methods section.*
Reviewer 2, Mark Oppe:

1. Which improvements were made in the US study compared to the UK?

   *We have amended the sentence to point to improvements in sampling methods and advanced methods for modeling population representativeness using complex sampling. Whether these improvements should be considered as “distinct” is open to discussion. One objective of including the sentence is trying to sandwich our rather critical appraisal of the regression model.*

2. The US study is not from 2001, but from 2003 (published in 2005)

   *Fixed.*

3. I would remove this paragraph [the one stating that EQ-5D-3L tariffs cannot be directly used on EQ-5L] because it is possible to use 3L value sets for the EQ-5D-5L: A crossover study between the 3L and the 5L has been undertaken by the EuroQol Group for just that purpose. This allows users of the EQ-5D to calculate utilities for the 5L based on 3L value sets. These could be used until the 5L valuation studies are completed.

   *We have removed the sentence stating that the 3L tariffs cannot be used on 5L. The point we wanted to get across, however, remains: the introduction of the 5L version is likely to lead to a number of new valuation studies being conducted.*

4. Even though you only need to predict 242 of the possible 243 health states (since state 11111 is 1 by definition) I suggest you replace 242 by 243 in order to avoid confusion for the reader as to the number of EQ-5D states. The description and explanation on page 8 are good and should remain.

   *We have altered 242 to 243 in the instance that precedes the explanation.*

5. From the text it seems that anchoring on 11111=1 and death = 0 is related to TTO. It should be made clear that it is a requirement of the QALY model, not of TTO (i.e. in standard gamble, DCE and VAS the same anchor points are used).

   *The sentence now reads: “In order to be in line with the quality adjusted life-year (QALY) model, health state 11111 was given a fixed value of 1, and death was given a fixed value of 0.”*
6. Page 9, should “D1” be “D1o”.

Yes.

7. How do you calculate ViF’s? (I for one, never heard of them) include a short description +
references of the ViF methodology.

We have included an explanation of how VIFs are calculated and a reference to a paper
discussing when high levels of multicollinearity is a problem.

8. If parameters are not significant, it means that they are not different from zero. If they are
zero they won’t have an impact on the model, so why would you not exclude them?

There are different schools of thought on whether or not to include non-significant variables.
That is not really the issue in this case. Our objective here was to point out an inconsistency:
the authors threw out the I2-term because it wasn’t statistically significant. But instead of
throwing out the non-significant level 2 terms, they altered the model in such a way as to
shift them away from zero, thereby making them seem significant. Consistency would require
either excluding the non-significant level 2 terms, or including the I2.

9. If the model results (i.e. the parameter estimates + SEs) are equivalent, why is it bad to have
high values for ViF? the presence of multicollinearity does not seem to effect the model
estimation, which is why you would want to avoid it.

Three points. First, the SEs were incorrectly reported as identical in the two models. We have
amended this. However, even if the SEs were identical, the D1o model would result in less
accurate predicted values for health states with impairments on several dimensions, since the
uncertainty surrounding the D1 would be multiplied by up to 4. Second, the primary issue is
the fact that the D1o model artificially makes the level 2 terms pass the test of statistical
significant, but the way the model is built is also likely to lead to incorrect interpretations of
the general population’s preferences for hypothetical health states. The analyses of VIFs were
included to illustrate fact that the D1 variable doesn’t add new information, it merely
obscures information already present in the D1c.

10. Delete “increases”.

Duplicate “increases” deleted.

11. Elaborate what you mean by redundancy in this situation. "normally" this means that if you
have two highly correlated parameters in your model you’re effectively putting in the same
variable twice, which makes one of them redundant. However in this case, the number of
parameters between the models are the same, suggesting there is no redundancy.
The D1 variable is redundant in that it all the information contained in it is already present in the primary dimension dummies. Including the D1 requires the removal of one variable, or the removal of the constant term. Whichever variable is removed, will be picked up by the D1. Since the constant term was removed, the D1 became an incredibly inefficient way of describing the constant. The VIF is, in a sense, a measure of redundancy, since it reflects the degree to which a variable is predicted by other variables in the model. In the case of the D1, 98.5% of the variance could be explained by the other variables. The remaining 1.5% represents the constant. To us, this suggests redundancy.