Article:

Tsuchiya, A. orcid.org/0000-0003-4245-5399, Bansback, N., Hole, A. et al. (1 more author) (2019) Manipulating the 5 dimensions of the EuroQol instrument: the effects on self-reporting actual health and valuing hypothetical health states. Medical Decision Making, 39 (4). pp. 380-392. ISSN 0272-989X

https://doi.org/10.1177/0272989X19851049

MDM 18 176 R1 for SAGE

Manipulating the five dimensions of the EuroQol instrument: the effects on self-reporting actual health and valuing hypothetical health states

Running head: Manipulating the EuroQol Instrument

Aki Tsuchiya, PhD[1,2], Nick Bansback, PhD[3], Arne Risa Hole, PhD[2], Brendan Mulhern, MRes[1,4]

1 School of Health and Related Research, University of Sheffield, UK
2 Department of Economics, University of Sheffield, UK
3 School of Population and Public Health, University of British Columbia, Canada
4 Centre for Health Economics Research and Evaluation, University of Technology Sydney, Australia

Corresponding author:

Aki Tsuchiya, School of Health and Related Research, Regent Court, 30 Regent Street, and Department of Economics, University of Sheffield. E-mail: a.tsuchiya@sheffield.ac.uk

Abstract:

Background The EQ-5D instrument has five dimensions. This paper reports on the effects of manipulating: (a) the order in which the five dimensions are presented (appearing first vs last); (b) splitting of the composite dimensions (“Pain or Discomfort”; and “Anxiety or Depression”); and (c) removing or ‘bolting off’ one of the five EQ-5D dimensions at a time. The effects were examined in two contexts: (1) self-reporting health, and (2) health state valuations.

Methods Three different Types of discrete choice experiments (DCE) including a duration attribute were designed. An online survey with 12 sub-Types, each with 10 DCE tasks, was designed, and completed by 2,494 members of the UK general public.

Results Of the three manipulations in the self-reporting context, only (b) splitting Anxiety or Depression had a significant effect. In the health state valuation context, (b) splitting level 5 Pain or Discomfort (relative to Pain) and splitting level 5 Anxiety or Depression (relative to Anxiety) had significant effects, as did (c) bolting off dimensions.

Conclusions We find that the values given to certain health dimensions are sensitive to the way in which it is described, and the other health dimensions presented. Of particular interest is the effect of splitting composite dimensions: a given EQ-5D(-5L) profile may mean different things depending on whether the profile is used to self-report one’s health or to value hypothetical states, so that the health state values EQ-5D(-5L) in population tariffs may not correspond to the states that patients self-report themselves in.

[246wds]

Keywords
EQ-5D; self-reported health; health state valuation; dimension order; composite dimensions; bolt-offs

Introduction

The EQ-5D(-5L) is a widely used generic instrument consisting of a health state classification system and preference-based weights [1,2]. The preference-based weights are derived using health state valuation methods such as the Time Trade-Off (TTO; [3]) or Discrete Choice Experiments (DCE [4]), and the classification system is used to assign these weights to different health states that patients classify themselves in. The five dimensions (Mobility (M); Self-Care (SC); Usual Activities (UA); Pain or Discomfort (PD); and Anxiety or Depression (AD)) are presented in a set order, and the latter two are composite dimensions of two closely related but different aspects of health. This paper explores three issues related to the EQ-5D classification system, across two contexts: one for self-reporting health and the other for valuing hypothetical health states.

The first issue concerns the order in which the dimensions are presented. The psychology literature has discussed ‘recency’ and ‘primacy’ biases, broadly described as serial position effects [5-7]. A recency bias represents a situation where a respondent remembers and processes more recent information more efficiently than earlier information [8]. On the other hand, a primacy bias occurs where a respondent might recall and process the information listed first rather than last [9]. In the context of self-reporting health, serial position effects may concern the *levels* of the dimensions: a primacy effect would suggest that respondents focus on the “no problems” level of a given health dimension because it is mentioned first; or a recency effect would make them more likely to focus on “extreme problems” because it appears last. As each dimension is presented separately as a set of five levels for which one is chosen, the overall order of the *dimensions* may not be explained by a serial position effect. On the other hand, in the context of health state valuations, where a health state is presented as a list of items, the order in which the dimensions of health are described could impact the way in which respondents process the information in line with recency or primacy effects [10]. Previous studies have found mixed effects. Mulhern et al [11] applied three different EQ-5D-5L dimension orders in TTO and DCE (without duration) and found that magnitude of the dimension coefficients varies across the different dimension orderings but without a clear pattern. Mulhern et al [12] used online DCE_{TTO} (with duration; see below) to compare the standard and systematically manipulated orderings of EQ-5D-5L dimensions at the between- and within-subject level and found little effect. Similarly, Norman et al [13] found that varying the dimension order in a valuation of the cancer specific EORTC QLU-C10D had little effect on level coefficients. In a study for monetary valuation, Kjaer et al [14] found that if the price attribute was presented last respondents provided lower willingness to pay – in DCE_{TTO}, this would correspond to the position of the duration attribute relative to the health state.

Another psychological effect of importance in the completion of questionnaires and DCE tasks is linked to the way in which information is attended to and cognitively processed. This is also affected by the way in which the information is presented, and the amount of information included. For example, Hensher [15] found that the amount and structure of the information provided in a DCE impacted the way in which that information was processed and responded to. This effect is tested in two further manipulations of the EQ-5D as described below.

The second issue concerns dimensions including more than one concept – composite dimensions. The PD and AD dimensions are each effectively a combination of two separate but related items which may lead to ambiguity. Furthermore, such composite dimensions have different meanings in the self-reporting and health state valuation contexts [16]. For example, it is entirely logical to use “moderate Pain or Discomfort” to self-report moderate Pain with no Discomfort, no Pain with moderate Discomfort, or moderate Pain alongside moderate Discomfort. However, where a respondent is asked to value “moderate Pain or Discomfort”, it would be incorrect to imagine moderate Pain alongside moderate Discomfort. This would suggest a mismatch between the state that people self-report and the state that a value is predicted for. There has only been limited empirical examination of this issue [17].

The third issue regards removing, or ‘bolting off’ dimensions. This is motivated by research on bolt-on dimensions to the EQ-5D(-5L) (e.g. [18]), which addresses the fact that the EQ-5D(-5L) only covers a limited range of health related quality of life (HRQL) dimensions and the concern that there may be contexts where information on other dimensions of HRQL is important not only for health care resource allocation decisions across clinical and public health, but also for health outcomes research more widely. In the context of self-report, if respondents feel that important dimensions are missing, they may report this information in their response to an existing dimension, so that the response to the original dimension depends on the other items included.[19] In the context of health state valuation, firstly, there may be circumstances where there is more than one “missing” dimension that needs adding to EQ-5D(-5L), but there is a limit to the number of dimensions that can reasonably be included in health state valuation studies. One way to deal with this in valuation studies is by introducing ‘overlap’ across dimensions so that certain attributes have the same level within choice sets, thus in effect, reducing the number of attributes that respondents need to consider [20]. However doing this reduces the efficiency of the study design, and another way to circumvent this could be by bolting *off* one or more of the existing EQ-5D(-5L) dimensions that may be less relevant to particular conditions. Secondly, the bolt-on literature has found that providing information that there is no problem in some additional dimension can make the state significantly better than the original EQ-5D state without this information [18,19,21]. Bolting off will allow a test of whether the same is applicable to the existing dimensions of the EQ-5D(-5L). We are aware of no other bolt-off studies.

Thus, this study aims to examine the effects of: the order in which the EQ-5D dimensions are presented (aim a), splitting up the two composite EQ-5D dimensions (PD and AD) (aim b), and bolting off one EQ-5D dimension at a time (aim c), on people’s self-reporting of actual health and the valuation of hypothetical states. There are six corresponding null hypotheses. First, regarding self-reported health,

- a1. the proportion of people reporting level 1 in a given dimension is not affected by whether the dimension appears first or last;
- b1. the proportion of people who self-report level 1 in a composite dimension is no different from the proportion of people who self-report level 1 in both sub-dimensions when the dimension is decomposed; and

c1. the proportion of people reporting problems in a dimension is unaffected by another dimension being bolted off.

Second, regarding health state valuation,

a2. the disutility of a dimension is unaffected by whether it is presented first or last;

b2. the disutility associated with a composite dimension is no larger than the disutility associated with either sub-dimension at the same level; and

c2. the disutility of a dimension is unaffected by another dimension being bolted off.

An online survey of the UK general public using DCE_{TTO} was conducted to address these.

Methods

Survey design, recruitment and the sample

Respondents were recruited from a commercial internet panel (IPSOS Observer). Quota sampling was used to ensure respondents were representative of the UK general population for age (across five age groupings from 18 to 65) and gender.

First, potential respondents accessed the survey webpage, read detailed project information and consented to take part. Those consenting to participate were then randomized to one survey variant and completed demographic and self-reported health status, ONS wellbeing questions and *the relevant variant* of the EuroQol Instrument (with five levels) for their own health based on the variant they were randomized to. They were then presented with information about the DCE_{TTO} tasks. These are DCE with duration as one of the attributes [22], and the naming reflects the fact that each choice requires respondents to trade-off between the health-related quality of life and the length of survival – as in a TTO. The DCE_{TTO} included details about the relevant EQ-5D-3/5L health dimensions (see below), and instructions to imagine: that they would experience each health state for the period shown without relief or treatment; that death would be very swift and completely painless; and that they would have no other health problems besides what was indicated. This was followed by ten DCE_{TTO} tasks. Respondents were screened out if they completed the survey in less than the minimum completion time of 2 minutes, which was set based on judgment of the research team following a soft launch phase. No maximum limit was set.

Experimental design: Overall

The DCE_{TTO} questions were based on "EQ-5D-**3/5L**" (with adaptations to address the relevant manipulations, as detailed below). The 3/5L indicates that it uses three of the five levels (1, 3, and 5) of EQ-5D-5L. The full EQ-5D-5L was not used for the valuation tasks in order to reduce the number of possible states to be valued.

The DCE_{TTO} scenarios consisted of "you" living in a particular state for one of three levels of Duration (6, 8, and 10 years) followed by death. The three levels of Duration were selected to include 10 years (a common value used in health state valuation research), and with narrower gaps than in previous DCE_{TTO} studies (see for example [22], which used 1, 5, and 10 years). The narrower gap was chosen, because a 10-fold difference in duration across a choice pair would mean that the scenario with the longer duration would be selected almost regardless of the state (provided both states are better than being dead) – narrower gaps would allow more trade-offs between the state and duration.

The analysis of DCE_{TTO} data involves modeling the pairwise choice data in terms of interactions between the health state and Duration [22].   The main DCE_{TTO} design is called Type III. (Types I and II addressed unrelated research questions and are reported in [23].) Type III involves 11 parameters (interactions between each of the EQ-5D-3/5L dummies and continuous Duration 5x(3-1)x1=10, plus continuous Duration). However, to allow for possible further analysis including EQ-5D-3/5L main effects and quadratic Duration, the design had 32 parameters (the above 11, plus main effects for EQ-5D-3/5L dummies 5x(3-1), interactions between these and Duration squared, and Duration squared). Sixty choice sets were selected based on a D-efficient design with zero prior values using Ngene [24], and allocated to six blocks of ten tasks. Target sample size for Type III was set at 700.

Experimental design for aim (a)

For aim (a), four sub-Types IIIa to IIId were created by varying the order in which the DCE_{TTO} attributes were presented. Even if the position of Duration is restricted to either precede or follow the health state dimensions, there are still 240 possible orderings (2x5x4x3x2x1). From these, the following particular four combinations were chosen: Type IIIa moved Mobility to the last of the EQ-5D dimensions, maintained the ordering of the remaining four dimensions, and kept Duration last (to understand how important the first dimension is); Type IIIb treated the first three dimensions (that are more functioning-based) and the last two dimensions (the are more symptoms-based) as blocks and swapped them round, and kept Duration last; Type IIIc reversed the ordering of the five EQ-5D dimensions (to test the impact of overall order), but kept Duration last; and Type IIId kept the EQ-5D ordering, but placed Duration first (to test the importance of the position of duration on the magnitude of its coefficient). Target sample size for the sub-Types IIIa to IIId was set at 150 each.

Experimental design for aim (b)

For aim (b), the composite dimensions were split to form EQ-**6D**-3/5L. One sub-variant of this (IVa) split Pain/Discomfort into a Pain dimension and a Discomfort dimension; the other sub-variant (IVb) split Anxiety/Depression into an Anxiety dimension and a Depression dimension – all else remained unchanged. This involves 13 parameters (interactions between each of the EQ-6D-3/5L level dummies and continuous Duration 6x(3-1)x1=12, plus continuous Duration). To allow further analyses in line with Type III, the Type IV design had 38 parameters (12x3+2). For each sub-Type, IVa and IVb, 60 choice sets were selected based on a D-efficient design with zero prior values, and allocated to six blocks of ten DCE_{TTO} tasks. Target sample size for the sub-Types IVa to IVb was set at 225 each.

Experimental design for aim (c)

For aim (c), one dimension was bolted off to form EQ-**4D**-3/5L. This has five sub-variants: one that dropped Mobility; another that dropped Self-Care; and so on. Duration always appeared last. This involves nine parameters (interactions between each of the EQ-4D-3/5L level dummies and continuous Duration 4x(3-1)x1=8, plus continuous Duration). To allow further analyses, the Type V design had 26 parameters (8x3+2). For each sub-Type, Va to Ve, 60 choice sets were selected based on a D-efficient design with zero prior values, and allocated to six blocks of ten DCE_{TTO} tasks. Target sample size for the sub-Types Va to Ve was set at 150 each.

Analysis

Self-reported health by sub-Type were summarised as bar charts, and differences in proportions reporting any problems in a given dimension were compared relative to Type III. In addition, for Type IV, cross-tables were used to examine the distribution of responses across the composite and corresponding split dimensions.

Throughout, DCE_{TTO} data were analysed as has been done previously [22], using conditional logit models with continuous Duration and interactions between the EQ-5D level dummies and Duration. Since the estimated coefficients are on a latent scale, they are "unanchored" and not directly comparable across models. For this, "anchored" coefficients representing decrements from full health on a scale with 1 for full health and 0 for dead are necessary. The main results are reported in terms of the anchored coefficients, and the unanchored coefficient are reported in the Appendices.

To investigate the effects of changing the ordering of dimensions (aim a), we first examined hypothesis a1 through the distribution of self-reported EQ-5D-5L health across the samples for Type III and Types IIIa to IIId. Regarding health state values (hypothesis a2), we compared the anchored coefficients between Types estimated from the separate models. Next, we replicated the analysis by Kjaer et al [14] by pooling the data across all the Types and incorporating interaction variables with the explanatory variables of the unanchored model multiplied by a dummy variable indicating the different Types. Given the number of variables, we focussed on the level-5 dimensions and the Duration attribute only. If the interaction variable with Type is statistically significant it will show that the different design has influenced that parameter estimate.

To address the effects of splitting composite dimensions (aim b), self-reported health was compared between Type III and Types IVa and IVb, to examine the distributions of self-reported PD in Type III alongside self-reported Pain and Discomfort in Type IVa; and similarly for AD (hypothesis b1). Furthermore, to test hypothesis b2, DCE_{TTO} data were modelled by sub-Type and compared to the Type III model. In particular, the size of the split coefficients (anchored) was compared with the corresponding composite coefficients.

To examine the effects of bolting off EQ-5D dimensions (aim c) to test hypotheses c1 and c2, a procedure similar to the one described above for hypotheses a1 and a2 was followed.

Results

Response rate and demographics

The analysis uses data from 2,494 respondents who completed the survey. Of these, 700 answered the baseline Type III DCE_{TTO} survey; 600 answered one of four sub-Types (IIIa to IIId) addressing aim (a); 450 answered one of two sub-Types (IVa and IVb) addressing aim (b); and just under 750 answered one of five sub-Types (Va to Ve) addressing aim (c). For details of the sample characteristics, see Table 1.

Results of aim (a)

Figure 1 illustrates the distribution of self-reported EQ-5D, by the samples for Types III and IIIa to IIId. The three panels on the left are for M, SC and UA, and the bar charts indicate the proportion of respondents at levels 1 to 5. The charts show that there are some deviations from Type III. However, across the four different orderings tested, none of the variations observed are explained by appearing first or last, and thus hypothesis a1 cannot be rejected.

Table 2 reports anchored coefficients of the DCE_{TTO} models. Looking along the rows, the largest decrement for four coefficients (M3, M5, SC3, UA3 and AD3) fall on IIId where Duration appears first, while it never falls on a dimension appearing first. Neither does the smallest decrement fall where the dimension appears first. The largest decrement appears last (just before Duration) only in III (AD3, but not AD5). The consistently smaller standard error when Duration is presented first (IIId) suggests that rearranging the EQ-5D dimensions can have larger impacts than the placing of Duration in the choice task.

Table 3 summarises the results of the Kjaer et al analysis incorporating interactions with the design variables (e.g. IIIa x M5xD), and using Type III as the baseline. All the data are pooled. The Duration dimension does not have significant design-interactions (other than IIIa x D), suggesting robustness amongst the four different orderings tested here. The M5 coefficient is the most vulnerable to ordering rearrangements amongst the level-5 coefficients, and its decrement inflated, except for when Duration appears first (IIId). Regarding hypothesis a2, this cannot be rejected. There is no clear pattern to be seen that can be explained with reference to serial position effects.

Results of aim (b)

Table 4 cross tabulates self-reported Pain, Discomfort, Anxiety and Depression using data from the samples for Types IVa, IVb and III. Since there are no statistically significant differences in self-assessed health across the samples and the other background characteristics are similar (cf. Table 1), it is reasonable to assume that underlying health across the samples is also similar. Taking the EQ-5D-5L wording at face value, only those with *no* Pain *or* Discomfort should report level 1 PD. From Type IVa in Table 4(a), this proportion is 51.6% (cell highlighted in dark grey), while the proportion observed in Type III in Table 4(c) is 56.9% (cell in dark grey; $p = 0.148$, *z*-test). Similarly, only those with *no* Anxiety *or* Depression should report level 1 AD. From Type IVb in Table 4(b), this proportion is 44.9% (cell in light grey), while the proportion observed in Type III in Table 4(c) is 57.3% (cell in light grey; $p = 0.001$). Thus, hypothesis b1 cannot be rejected for PD, but is rejected for AD.

Table 5 presents the anchored coefficient. Regarding PD, the level 3 and level 5 composite coefficients in Type III are larger in magnitude than the corresponding sub-domain coefficients in Type IVa for Pain (one-sided $p = 0.307$ for level 3, $p = 0.139$ for level 5); and similarly for Discomfort ($p = 0.307$ for level 3, $p < 0.001$ for level 5). Regarding AD, the composite coefficients in Type III are larger in magnitude than the corresponding coefficients in Type IVb for Anxiety ($p = 0.141$ for level 3, $p < 0.001$ for level 5); but smaller for Depression ($p = 0.093$ for level 3, $p = 0.119$ for level 5). Thus, hypothesis b2 is not rejected for any of the level 3 coefficients; for the level 5 coefficients, it is rejected for PD (relative to Pain) and AD (relative to Anxiety).

Results of aim (c)

Figure 2 illustrates the distribution of self-reported levels of each of the (relevant) EQ-5D dimension, by the samples for Types III and Va to Ve. The EQ-5D-5L dimensions bolted off appear as blank spaces. The charts show that there are variations in self-reported EQ-5D-5L problems across the bolt-off Types, but none of these are statistically significant from Type III (*z*-test at 5%). Thus, hypothesis c1 cannot be rejected.

Table 6 presents the anchored coefficients. Of the 36 significant anchored coefficients amongst the five Type V models, 33 of them are larger in the bolt-off model than the corresponding coefficient in Type III. Type Vd (bolting off PD) is the exception and has two level-3 coefficients that are non-significant (M3, SC3) and two coefficients that are smaller than the corresponding Type III coefficient (M5, UA5). Table 7 summarises the results of the Kjaer et al analysis incorporating interactions with the design variables (e.g. Vb x M5xD), using Type III as the baseline and therefore omitted; also omitted are the coefficients for the dimension that is bolted off in each Type (e.g. interaction of Va and M5xD is omitted). The results show that first, all the significant coefficients are negative; second, the M5 coefficient is not significantly affected by other dimensions being bolted off, but the other coefficients are; third, SC5, US5 and PD5 are affected more than AD5; and finally, Duration is not affected. Thus, hypothesis c2 is rejected: with the exception of Mobility, the disutilities of the dimensions are affected by another dimension being bolted off.

Discussion

This paper examines the effects of: (a) varying the order in which EQ-5D(-5L) dimensions are presented; (b) splitting up the two composite EQ-5D dimensions (PD and AD); and (c) bolting *off* one EQ-5D dimension at a time. The effects were examined in two contexts: (1) self-reporting own health; and (2) health state valuation using DCE_{TTO}. An online survey using a commercial internet panel was conducted in the UK with 2,494 respondents.

Regarding hypothesis a1 (the proportion of people reporting level 1 in a given dimension is not affected by whether the dimension appears first or last), and hypothesis a2 (the disutility of a dimension is unaffected by whether it is presented first or last), neither of these can be rejected by out data. While both self-reported health and health state values varied significantly by the order in which the EQ-5D-5L dimensions are presented, these cannot be explained with respect to the serial position effects. The study has only examined four orderings out of the possible 240 permutations (two Duration dimension positions, with 120 EQ-5D-5L dimension orderings), and the effect of the position may have interacted with the dimension itself. For health state valuation, the significant variations appear to contradict Mulhern et al [12] which randomised all 120 possible permutations of the five dimensions both within and between respondents (with Duration always last) and found no significant effect of dimension ordering in a DCE_{TTO}. However, Mulhern et al [12] examined the effect of the *position* of a dimension (independently of the other four dimensions), whereas this study examined the effect of the *ordering* (permutations) of the dimensions. (Mulhern et al [12] did not analyse self-reported health.)

Hypothesis b1 (the proportion of people who self-report level 1 in a composite dimension is no different from the proportion of people who self-report level 1 in both sub-dimensions when the dimension is decomposed) cannot be rejected for Pain or Discomfort, but is rejected for Anxiety or Depression. In other words, the PD dimension appears to be interpreted more literally than the AD dimension. It would be interesting to see if this holds for non-English versions of EQ-5D.

Hypothesis b2 (the disutility associated with a composite dimension is no larger than the disutility associated with either sub-dimension at the same level) is not rejected for any of the level 3 coefficients, but is rejected for the level 5 coefficients (PD relative to Pain, and AD relative to Anxiety). Specifically, what respondents have in mind when valuing "extreme Pain or Discomfort" is significantly worse than extreme Discomfort on its own, and what respondents have in mind when valuing "extreme Anxiety or Depression" is significantly worse than extreme Anxiety on its own.

Imagine a patient with extreme problems in all the dimensions: Table 5 reports the predicted values in the last row. Compared to Type III (EQ-5D-5L), Type IVa (splitting PD) results in a health state value that is milder by 0.099, while Type IVb (splitting AD) results in a value that is more severe by 0.118. These are substantial differences. To give some context, a recent study has calculated the minimally important difference for EQ-5D-5L across six countries and found them in the region of 0.037 to 0.069 [25], while the differences observed here are an order of magnitude larger.

This has major implications for EQ-5D-5L, and beyond. In effect, a given EQ-5D-5L profile may mean different things depending on whether the profile is used to self-report one's health or to value hypothetical states, so that the health state values EQ-5D-5L in population tariffs may not correspond to the states that patients self-report themselves in. Furthermore, since our study has used an adapted version of EQ-5D-5L with levels 1, 3, and 5 only, the three-level version of EQ-5D is highly likely to be susceptible to the same effect. Taken at face value, our results would suggest economic evaluation that use EQ-5D(-5L) are systematically biased. Establishing the robustness of the findings using other valuation methods, to gauge the size and direction of the error in terms of health gains (as opposed to values for health states) is a research priority. In addition, a composite dimension brings together items that are thought to be similar or related to each other, and if so, splitting them would, in effect, create two dimensions that violate independence. Given this, simply splitting the composite dimension(s) is unlikely to be the best solution. Further research on splitting-and-dropping items may be informative.

Regarding hypothesis c1 (the proportion of people reporting problems in a dimension is unaffected by another dimension being bolted off), this cannot be rejected. The effect of dropping an EQ-5D dimension on the other dimensions seemed to be relatively small, and none were significant. However, hypothesis c2 (the disutility of a dimension is unaffected by another dimension being bolted off) is rejected: the absence of a dimension was found to affect the values of the remaining dimensions other than Mobility, while the value of survival in full health remained unaffected.

To place this in context, the lower rows of Table 6 reports predicted values. With the exception of Type Vd (drop PD), the worst states in the bolt-off instrument (with four extreme dimensions) resulted in more severe predicted values than the corresponding EQ-5D-5L (with four extreme problems) – and the differences between corresponding states are an order of magnitude larger than the minimum importance difference referenced above – the implication is that when respondents value a health state, they do not assume that everything else not mentioned by the descriptive system is fine. Indeed (with the exception of Type Vd), the worst states in the bolt-off instrument (with four extreme dimensions) had predicted values that were worse than 55555 with five extreme dimensions (-0.774, Type III).

While similar findings have been observed in the bolt-*on* literature, this is inconsistent with the practice in typical health state valuation exercises (including the present study) instructing respondents to imagine that there are no further health problems beyond what is explicitly mentioned in the hypothetical health state to be valued. The results reinforce the possibility that respondents make their own inferences about problems in the unmentioned dimensions of health. This adds to the growing evidence on the non-independence of bolt-on items, and more generally on how respondents to health state valuation tasks supplement the minimal description of the hypothetical health states with their own concepts. These call for better ways of informing the respondents about the hypothetical health states they are valuing, and perhaps probing the respondents about those states afterwards, rather than simply presenting them with short abstract health state descriptions.

The study has a few limitations. The first concerns the use of DCE_{TTO}, which is not the valuation method used in the recommended protocol for valuing EQ-5D-5L [26]. Because this is an ordinal method that relies heavily on econometric modeling and its assumptions, there may be concerns over the validity of the anchored values produced. However, the aim of this study is entirely methodological and does not aim to produce a population tariff. Furthermore, the experimental versions of EQ-5D are compared against the reference that is also valued within the same study using DCE_{TTO}, without involving cross-method or cross-study comparisons. Of course, the design does not rule out the possibility that DCE_{TTO} is particularly vulnerable to, for example, splitting composite dimensions. And therefore it would be of major interest to see whether the findings hold for the composite Time Trade-Off method, and in face-to-face interviews [26]. The second would be the use of an online survey recruiting respondents from a commercial internet panel. The pros and cons of online surveys relative to face-to-face interviews in the context of health state valuations have been discussed elsewhere [27], and include the substantially lower costs, the speed of data collection, the absence of interviewer effects, alongside the exclusion of certain populations and possible lack of respondent engagement. The same study has, however, found that in terms of binary choice behaviour, online surveys and interview surveys do not differ in terms of binary choice behaviour of the kind used in DCE_{TTO}. And finally, the study was conducted in the UK using the English version of the EQ-5D-5L – the results may or may not hold in different language versions or populations beyond the UK.

To conclude, the paper reports on a study that examined effect of manipulating the EQ-5D health state classification system in three different ways (change dimension ordering; split composite dimensions; bolt-off one dimension at a time) across two contexts (self-reporting health and health state valuation). The values given to certain health dimensions are sensitive to the way in which it is described, and the other health dimensions presented. Of particular interest concerns how the composite dimensions are interpreted differently across self-reporting and health state valuation – this raises questions about the validity of EQ-5D(-5L) in economic evaluation.

[4996wds main text]

Acknowledgements:

Financial support for this study ('Further Exploration of Discrete Choice Experiments with duration for EQ-5D-5L Valuation' – the FEDEV project) was provided entirely by a grant from the EuroQol Research Foundation. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report. AT, NB, and BM are members of the EuroQol Group. Research Ethics approval was given by the University of Sheffield Research Ethics Committee (ref. 0692/KW). An earlier version of this paper was presented at the EuroQol Group conference, 2014, Stockholm. We are grateful to Ben van Hout, Claire Gudex, the four anonymous reviewers and the editor for comments, and all the respondents who took part. The usual disclaimers apply.

References

[1] Brooks R (1996). EuroQol: the current state of play. *Health Policy*, 37: 53-72.

[2] Herdman M, Gudex C, Lloyd A, Janssen MF, Kind P, Parkin D, Bonsel G, Badia X (2011) Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research*. 20(10):1727–36.

[3] Torrance GW, Thomas WH, Sackett DL (1972). A utility maximization model for evaluation of health care programs, *Health services Research*, 7(2):118-133

[4] Stolk E, Oppe M, Scalone L, Krabbe P. Discrete choice modelling for the quantification of health states: The case of the EQ-5D. *Value Health*. 2010;13(8):1005-13.

[5] Ebbinghaus H (1913). *Memory: A contribution to experimental psychology*. Teachers college, Columbia University.

[6] Haugtvedt CP, Wegener DT (1994). Message Order Effects in Persuasion: An Attitude Strength Perspective. *Journal of Consumer Research*. 21(1):205–18.

[7] Shteingart H, Neiman T, Loewenstein Y (2013). The role of first impression in operant learning. *Journal of Experimental Psychology*. 142(2):476.

[8] Trotman K, Wright A (2000). Order effects and recency: Where do we go from here? *Accounting & Finance*. 40(2):169–82.

[9] Lohse GL (1997). Consumer eye movement patterns on yellow pages advertising. *Journal of Advertising*. 26(1):61–73.

[10] Schkade David A, Kleinmuntz Don N (1994). Information displays and choice processes: Differential effects of organization, form, and sequence. *Organizational Behavior and Human Decision Processes*. 57(3):319–37.

[11] Mulhern B, Shah K, Janssen MF, Longworth L, Ibbotson R (2016). Valuing Health Using Time Trade-Off and Discrete Choice Experiment Methods: Does Dimension Order Impact on Health State Values? *Value Health*. 19(2): 210-7.

[12] Mulhern B, Norman R, Lorgelly P, Lancsar E, Ratcliffe J, Brazier J, Viney R (2017a). Is dimension order important when valuing health state using discrete choice experiments including duration? *PharmacoEconomics*, 35(4): 439-451

[13] Norman R, Kemmler G, Viney R, Pickard AS, Gamper E, Holzner B, Nerich V, King M (2016). Order of Presentation of Dimensions Does Not Systematically Bias Utility Weights from a Discrete Choice Experiment. *Value Health*. 19(8): 1033-38.

[14] Kjær T, Bech M, Gyrd-Hansen D, Hart-Hansen K (2006). Ordering effect and price sensitivity in discrete choice experiments: need we worry? *Health Economics*. 15(11):1217–28.

[15] Hensher, D.A. (2006). How do respondents process stated choice experiments? Attribute consideration under varying information load. *Journal of Applied Econometrics*, Vol. 21 No. 6, pp. 861-78

[16] Janssen MF, Birnie E, Haagsma JA, Bonsel GJ. (2008). Comparing the standard EQ-5D three-level system with a five-level version. *Value in Health*, 11(2), 275-284.

[17] Bryan S, Jowett S, Hardyman W, Bentham P. (2004). Does the EQ-5D "anxiety/depression" item measure anxiety, depression, both or neither? *EuroQol Plenary Proceedings*.

[18] Longworth L, Yang Y, Young T, Mulhern B, Hernandez Alava M, Mukuria C, Rowen D, Tosh J, Tsuchiya A, Evans P, Keetharuth A (2014). Use of generic and condition-specific measures of health-related quality of life in NICE decision-making: Systematic review, statistical modelling and survey, *Health Technology Assessment*, 18(9)

[19] Yang Y, Rowen D, Brazier J, Tsuchiya A, Young T, Longworth L. An exploratory study to test the impact on three "bolt-on" items to the EQ-5D. *Value in Health*. 2015 Jan 1;18(1):52-60.

[20] King MT, Viney R, Pickard AS, Rowen D, Aaronson NK, Brazier JE, Cella D, Costa DS, Fayers PM, Kemmler G, McTaggart-Cowen H (2018). Australian Utility Weights for the EORTC QLU-C10D, a Multi-Attribute Utility Instrument Derived from the Cancer-Specific Quality of Life Questionnaire, EORTC QLQ-C30. *Pharmacoeconomics*. 36(2): 225-238.

[21] Yang Y, Brazier J, Tsuchiya A (2014). Effect of adding a sleep dimension to the EQ-5D descriptive system: A "bolt-on" experiment, *Medical Decision Making*. 34(1): 42-53

[22] Bansback N, Brazier J, Tsuchiya A, Anis A (2012). Using a discrete choice experiment to estimate societal health state utility values. *Journal of Health Economics*, 31(1):306-318.

[23] Mulhern B, Bansback N, Hole AR, Tsuchiya A (2017b). Using Discrete Choice Experiments with duration to model EQ-5D-5L health state preferences: Testing experimental design strategies. *Medical Decision Making*, 37(3), pp.285-297.

[24] Choice Metrics (2012). Ngene

[25] McClure NS, Al Sayah F, Xie F, Luo N, Johnson JA. (2017). Instrument-defined estimates of the minimally important difference for EQ-5D-5L index scores. *Value in Health*, 20(4), 644-650.

[26] Oppe M, Devlin NJ, van Hout B, Krabbe PF, de Charro, F. (2014). A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. *Value in Health*, 17, 445–453.

[27] Mulhern B, Longworth L, Brazier J, Rowen D, Bansback N, Devlin N, Tsuchiya A. (2013). Binary choice health state valuation and mode of administration: head-to-head comparison of online and CAPI. *Value in Health*, 16(1), 104-113.

Table 1: Completion and demographics for all Types

	Type III (n,%)	Type IIIa (n,%)	Type IIIb (n,%)	Type IIIc (n,%)	Type IIId (n,%)	Type IVa (n,%)	Type IVb (n,%)	Type Va (n,%)	Type Vb (n,%)	Type Vc (n,%)	Type Vd (n,%)	Type Ve (n,%)
Completion process												
Invited	5247	1583	1341	1220	1079	1597	1610	1147	1189	1247	1369	1270
Accessed[1]	1098 (20.9)	225 (14.2)	232 (17.2)	223 (18.3)	242 (22.4)	377 (23.6)	392 (24.3)	251 (21.7)	230 (19.3)	235 (18.8)	244 (17.8)	251 (19.8)
Not consent/pass info[2]	361 (32.9)	63 (28.0)	71 (5.3)	63 (28.2)	85 (35.1)	141 (37.4)	155 (39.5)	96 (8.4)	77 (33.5)	80 (34.00	84 (34.4)	96 (38.2)
Drop out during survey[2]	36 (3.3)	11 (4.8)	9 (3.9)	10 (4.4)	7 (2.9)	11 (2.9)	12 (3.1)	5 (2.0)	7 (3.0)	4 (1.7)	10 (4.0)	5 (2.0)
Complete < 2 mins[2]	1 (0.1)	0	2 (0.8)	0	0	0	0	2 (0.8)	0	1 (0.4)	0	0
Full completer[2]	700 (63.8)	150 (66.7)	150 (64.7)	150 (67.3)	150 (62.0)	225 (59.7)	225 (57.4)	148 (59.4)	146 (63.5)	150 (63.8)	150 (61.5)	150 (59.8)
Obs per block (6 in each Type)	103-132	16-34	22-27	14-31	23-28	29-50	29-48	21-35	20-28	16-38	16-38	18-33
Demographics												
Male	340 (48.6)	72 (48.0)	73 (48.7)	71 (47.3)	69 (46.0)	102 (45.3)	108 (48.0)	73 (49.3)	69 (47.2)	73 (48.7)	75 (50.0)	72 (48.0)
Age (m)	41.5	42.9	41.8	41.5	42.1	42.3	42.6	41.02	41.1	42.7	41.7	42.6
Age (range)	18-65	18-73	17-65	18-65	18-65	18-65	18-65	18-74	18-65	18-66	18-65	18-65
Married/partner	390 (55.8)	90 (60.0)	79 (52.7)	86 (56.9)	88 (58.7)	129 (57.3)	141 (62.7)	88 (59.4)	90 (61.6)	95 (63.3)	91 (60.7)	93 (62.0)
In employment	462 (66.1)	90 (60.0)	96 (64.0)	99 (66.0)	93 (62.0)	148 (65.8)	154 (68.4)	98 (65.3)	97 (66.4)	100 (66.6)	99 (66.0)	96 (64.0)
Education past min age	585 (83.6)	128 (85.3)	125 (83.3)	126 (84.0)	120 (80.0)	184 (81.8)	202 (89.8)	121 (81.8)	127 (87.0)	122 (81.3)	133 (88.7)	123 (82.0)
Degree	389 (56.3)	83 (55.3)	87 (58.0)	78 (52.0)	76 (50.7)	105 (46.7)	132 (58.7)	80 (54.1)	88 (60.3)	85 (56.7)	87 (58.0)	77 (51.3)
Self-assessed health												
Good/very good/excellent	570 (82.1)	126 (84.0)	115 (76.7)	120 (80.0)	121 (80.7)	187 (83.1)	182 (80.9)	123 (83.1)	121 (82.9)	125 (83.3)	119 (79.4)	117 (78.0)
Fair/Poor	128 (17.9)	24 (16.0)	35 (23.3)	30 (20.0)	29 (19.3)	38 (16.9)	43 (19.1)	25 (16.9)	25 (17.1)	25 (16.7)	31 (20.6)	33 (22.0)

1 – Percentage of those invited to take part; 2 – Percentage of those accessing the survey

Table 2: Type III and sub-variants with different ordering (anchored)

	III		IIIa (SC-UA-PD-AD-M)		IIIb (PD-AD-M-SC-UA)		IIIc (AD-PD-UA-SC-M)		IIId (Duration first)	
	Coef	SE	Coef	SE	Coef	SE	Coef	SE	Coef	SE
M3	-0.081	0.044	-0.130	0.102	-0.076	0.099	-0.079	0.101	-0.142	0.083
M5	-0.355	0.046	-0.303	0.104	-0.273	0.100	-0.295	0.104	-0.393	0.088
SC3	-0.076	0.042	-0.060	0.099	-0.022	0.095	-0.088	0.098	-0.119	0.083
SC5	-0.307	0.045	-0.332	0.111	-0.297	0.100	-0.337	0.104	-0.334	0.084
UA3	-0.091	0.042	-0.093	0.095	-0.092	0.094	-0.119	0.097	-0.122	0.086
UA5	-0.246	0.044	-0.269	0.101	-0.243	0.096	-0.224	0.100	-0.243	0.085
PD3	-0.091	0.045	-0.117	0.100	-0.051	0.100	-0.043	0.104	-0.087	0.082
PD5	-0.452	0.050	-0.497	0.115	-0.465	0.112	-0.397	0.116	-0.381	0.087
AD3	-0.104	0.043	-0.083	0.098	-0.103	0.096	-0.091	0.099	-0.117	0.080
AD5	-0.414	0.048	-0.464	0.115	-0.389	0.103	-0.419	0.111	-0.350	0.084

NB. Unanchored coefficients and SE in Table A1 in the Appendix.

Table 3: The effect of changing the ordering on specific DCE_{TTO} attributes

	Pooled	
	Coef.	P>z
IIIa x M5xD	**0.032**	0.013
IIIb x M5xD	**0.034**	0.007
IIIc x M5xD	**0.026**	0.033
IIId x M5xD	**-0.025**	0.042
IIIa x SC5xD	**-0.038**	0.002
IIIb x SC5xD	-0.021	0.082
IIIc x SC5xD	-0.023	0.056
IIId x SC5xD	-0.007	0.543
IIIa x UA5xD	**-0.033**	0.012
IIIb x UA5xD	-0.023	0.068
IIIc x UA5xD	-0.001	0.911
IIId x UA5xD	0.002	0.871
IIIa x PD5xD	**-0.028**	0.039
IIIb x PD5xD	**-0.030**	0.020
IIIc x PD5xD	0.007	0.574
IIId x PD5xD	**0.047**	0.000
IIIa x AD5xD	**-0.027**	0.046
IIIb x AD5xD	0.009	0.464
IIIc x AD5xD	0.012	0.335
IIId x AD5xD	**0.046**	0.000
IIIa x D	**0.033**	0.014
IIIb x D	0.013	0.300
IIIc x D	-0.008	0.513
IIId x D	-0.015	0.244
No of respondents	1300	
No of choices	13000	
Log Likelihood	-9120	

NB1. Coefficients with $p < 0.05$ in bold.
NB2. All models controls for Duration and for two-way interactions between Dimension-level and Duration. All controls are significant at the 0.001 level. Full results available from the authors on request.

Table 4: Distribution of self-reported Pain, Discomfort, Anxiety and Depression

(a) Pain and Discomfort, Type IVa sample (n=250)

Type IVa (% of total)		Pain		
		None	Any	Total
Discomfort	None	51.6	4.9	56.4
	Any	4.9	38.7	43.6
	Total	56.4	43.6	100.0

(b) Anxiety and Depression, Type IVb sample (n=250)

Type IVb (%of total)		Anxiety		
		None	Any	Total
Depression	None	44.9	23.1	68.0
	Any	2.2	29.8	32.0
	Total	47.1	52.9	100.0

(c) Pain/Discomfort and Anxiety/Depression, Type III sample (n=700)

Type III (% of total)		Pain/Discomfort		
		None	Any	Total
Anxiety/ Depression	None	40.3	17.0	57.3
	Any	16.6	26.1	42.7
	Total	56.9	43.1	100.0

NB. The null hypothesis implies that the cells with the same highlights have the same percentages.

Table 5: Type III and Type IV splitting the composite dimensions (anchored)

	III (Baseline)		IVa (PD split)		IVb (AD split)	
Parameter	**Coef**	**SE**	**Coef**	**SE**	**Coef**	**SE**
M3	-0.081	0.044	-0.128	0.026	-0.083	0.029
M5	-0.355	0.046	-0.287	0.029	-0.338	0.035
SC3	-0.076	0.042	-0.103	0.024	-0.057	0.026
SC5	-0.307	0.045	-0.321	0.030	-0.283	0.034
UA3	-0.091	0.042	-0.024	0.024	0.003	0.027
UA5	-0.246	0.044	-0.140	0.024	-0.188	0.029
PD3	-0.091	0.045			-0.060	0.024
PD5	-0.452	0.050			-0.401	0.038
AD3	-0.104	0.043	-0.121	0.024		
AD5	-0.414	0.048	-0.400	0.034		
Pa3			-0.066	0.021		
Pa5			-0.387	0.033		
Di3			-0.066	0.021		
Di5			-0.140	0.024		
An3					-0.051	0.024
An5					-0.192	0.028
De3					-0.172	0.028
De5					-0.490	0.043
33333/333333	0.557		0.492		0.580	
55555/555555	-0.774		-0.675		-0.892	

NB. Unanchored coefficients and SE in Table A2 in the Appendix.

Table 6: Type III and Type V bolting off different EQ dimensions (anchored)

Parameter	III (Baseline)		Va (drop M)		Vb (drop SC)		Vc (drop UA)		Vd (drop PD)		Ve (drop AD)	
	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE	Coeff	SE
M3	-0.081	0.044			-0.107	0.036	-0.082	0.039	-0.016	0.029	-0.142	0.035
M5	-0.355	0.046			-0.438	0.049	-0.388	0.050	-0.264	0.031	-0.459	0.053
SC3	-0.076	0.042	-0.110	0.034			-0.093	0.036	-0.040	0.028	-0.069	0.035
SC5	-0.307	0.045	-0.488	0.054			-0.434	0.057	-0.331	0.034	-0.367	0.046
UA3	-0.091	0.042	-0.146	0.036	-0.143	0.036			-0.100	0.028	-0.092	0.037
UA5	-0.246	0.044	-0.381	0.045	-0.369	0.045			-0.229	0.033	-0.330	0.047
PD3	-0.091	0.045	-0.133	0.039	0.006	0.036	-0.120	0.038			-0.111	0.034
PD5	-0.452	0.050	-0.520	0.057	-0.532	0.058	-0.516	0.065			-0.631	0.071
AD3	-0.104	0.043	-0.185	0.038	-0.229	0.036	-0.085	0.033	-0.189	0.029		
AD5	-0.414	0.048	-0.533	0.063	-0.587	0.063	-0.481	0.063	-0.516	0.047		
3333	–		0.426		0.527		0.620		0.655		0.586	
3333*	–		0.638		0.633		0.648		0.648		0.661	
5555	–		-0.922		-0.926		-0.819		-0.340		-0.787	
5555*	–		-0.419		-0.467		-0.528		-0.322		-0.360	

*Predicted value of the corresponding state in EQ-5D 3/5L using the Type III coefficients (e.g. 13333 for 3333 in Va; or 55515 for 5555 in Vd)

NB. Unanchored coefficients and SE in Table A3 in the Appendix.

Table 7: The effect of bolting off dimensions on specific DCE_{TTO} attributes

	Pooled	
	Coef.	P>z
III x M5xD	Omitted	
Va x M5xD	Omitted	
Vb x M5xD	-0.014	0.294
Vc x M5xD	-0.007	0.579
Vd x M5xD	0.008	0.517
Ve x M5xD	-0.021	0.138
III x SC5xD	Omitted	
Va x SC5xD	**-0.057**	0.000
Vb x SC5xD	Omitted	
Vc x SC5xD	**-0.039**	0.002
Vd x SC5xD	**-0.028**	0.023
Ve x SC5xD	-0.022	0.084
III x UA5xD	Omitted	
Va x UA5xD	**-0.036**	0.005
Vb x UA5xD	**-0.028**	0.031
Vc x UA5xD	Omitted	
Vd x UA5xD	-0.003	0.826
Ve x UA5xD	**-0.027**	0.037
III x PD5xD	Omitted	
Va x PD5xD	-0.007	0.594
Vb x PD5xD	**-0.028**	0.040
Vc x PD5xD	-0.014	0.323
Vd x PD5xD	Omitted	
Ve x PD5xD	**-0.053**	0.000
III x AD5xD	Omitted	
Va x AD5xD	-0.018	0.204
Vb x AD5xD	-0.027	0.063
Vc x AD5xD	-0.022	0.123
Vd x AD5xD	**-0.042**	0.002
Ve x AD5xD	Omitted	
III x D	Omitted	
Va x D	-0.053	0.187
Vb x D	-0.046	0.236
Vc x D	-0.021	0.656
Vd x D	0.038	0.350
Ve x D	-0.042	0.330
No of respondents	1,444	
No of choices	14,440	
Log Likelihood	-6,504	

NB1. Coefficients with $p < 0.05$ in bold.
NB2. All models controls for Duration and for two-way interactions between Dimension-level and Duration. All controls are significant at the 0.001 level. Full results available from the authors on request.

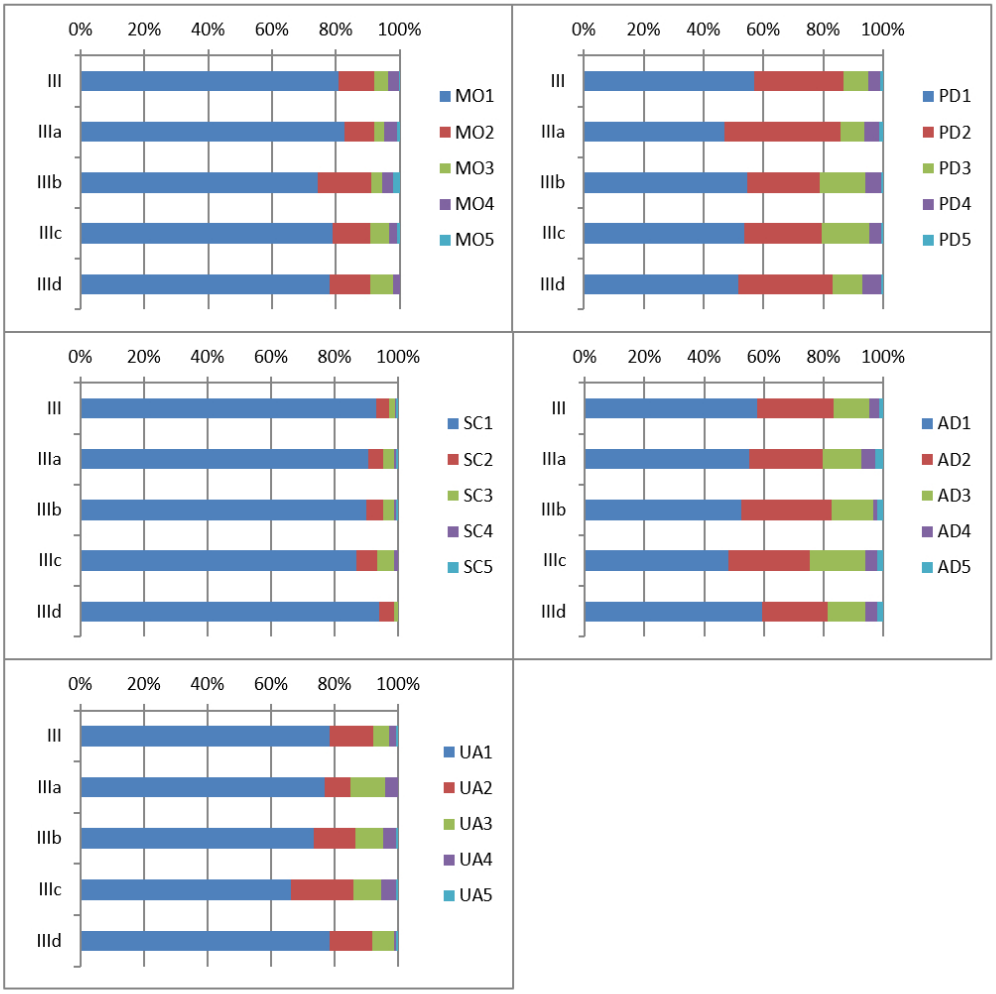

Figure 1: Distribution of self-reported health: Types III , IIIa, IIIb, IIIc, and IIId (ordering)

Type III: baseline
Type IIIa: (SC-UA-PD-AD-M-D)
Type IIIb: (PD-APD-M-SC-UA-D)
Type IIIc: (AD-PD-UA-SC-M-D)
Type IIId: (D-M-SC-US-PD-AD)

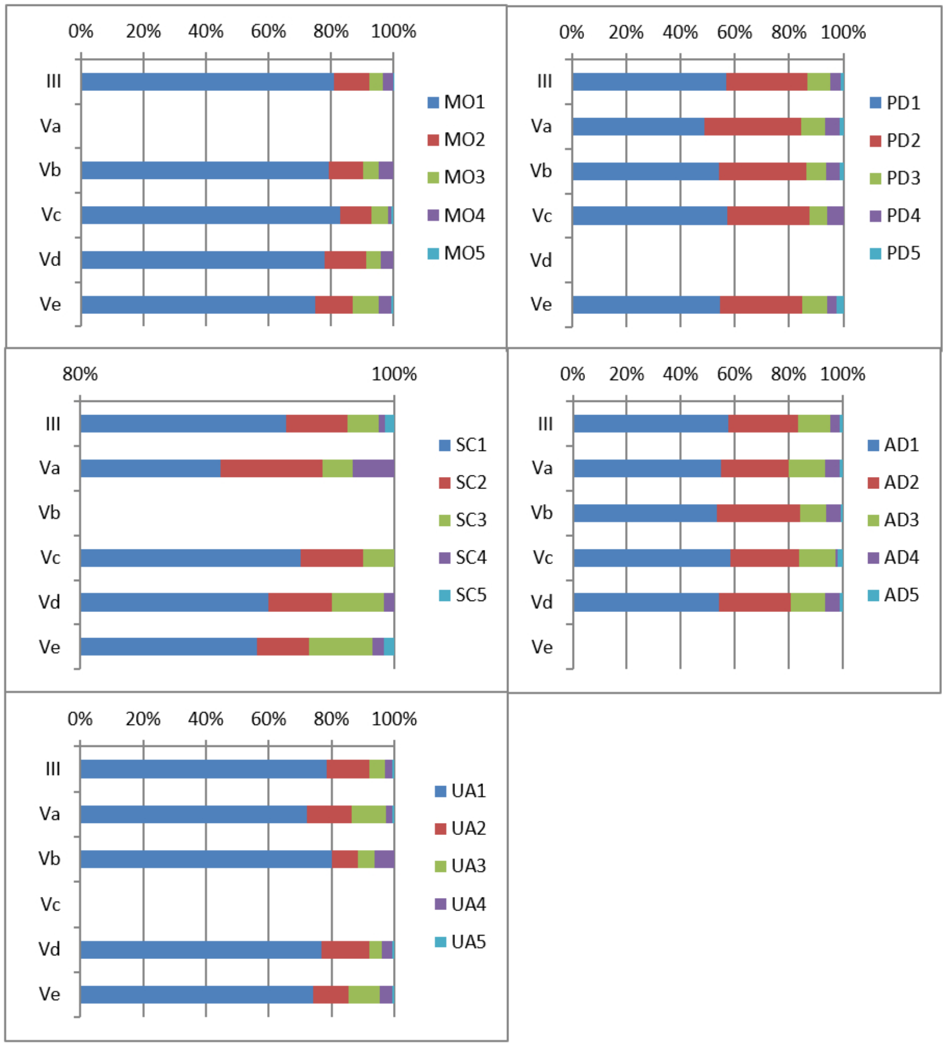

Figure 2: Distribution of self-reported health: Types III and V (bolt-off)

Type III: baseline
Type Va: drop Mobility
Type Vb: drop Self-Care
Type Vc: drop Usual Activities
Type Vd: drop Pain/Discomfort
Type Ve: drop Anxiety/Depression

APPENDICIES

Table A1: Type III and sub-variants with different ordering (unanchored)

	III			**IIIa (SC-UA-PD-AD-M)**			**IIIb (PD-AD-M-SC-UA)**			**IIIc (AD-PD-UA-SC-M)**			**IIId (Duration first)**		
Parameter	Coef	SE	P	Coef	SE	P	Coef	SE	P	Coef	SE	P	Coef	SE	P
M3 x D	**-0.032**	0.006	0.000	**-0.050**	0.014	0.000	**-0.028**	0.013	0.024	**-0.028**	0.012	0.023	**-0.085**	0.016	0.000
M5 x D	**-0.140**	0.006	0.000	**-0.117**	0.014	0.000	**-0.101**	0.013	0.000	**-0.104**	0.014	0.000	**-0.235**	0.017	0.000
SC3 x D	**-0.030**	0.006	0.000	-0.023	0.014	0.097	-0.008	0.013	0.514	**-0.031**	0.013	0.013	**-0.071**	0.016	0.000
SC5 x D	**-0.121**	0.006	0.000	**-0.128**	0.014	0.000	**-0.110**	0.013	0.000	**-0.119**	0.013	0.000	**-0.200**	0.016	0.000
UA3 x D	**-0.036**	0.006	0.000	**-0.036**	0.014	0.008	**-0.034**	0.013	0.008	**-0.042**	0.012	0.001	**-0.073**	0.019	0.000
UA5 x D	**-0.097**	0.006	0.000	**-0.104**	0.014	0.000	**-0.090**	0.013	0.000	**-0.079**	0.012	0.000	**-0.145**	0.016	0.000
PD3 x D	**-0.036**	0.006	0.000	**-0.045**	0.014	0.001	-0.019	0.013	0.162	-0.015	0.012	0.249	**-0.052**	0.018	0.000
PD5 x D	**-0.178**	0.006	0.000	**-0.192**	0.014	0.000	**-0.172**	0.013	0.000	**-0.140**	0.012	0.000	**-0.228**	0.015	0.000
AD3 x D	**-0.041**	0.006	0.000	**-0.032**	0.014	0.012	**-0.038**	0.013	0.002	**-0.032**	0.013	0.006	**-0.070**	0.017	0.000
AD5 x D	**-0.163**	0.006	0.000	**-0.179**	0.014	0.000	**-0.144**	0.013	0.000	**-0.148**	0.013	0.000	**-0.209**	0.016	0.000
D	**0.394**	0.019	0.000	**0.386**	0.042	0.000	**0.370**	0.041	0.000	**0.353**	0.038	0.000	**0.598**	0.048	0.000
No of respondents	700			150			150			150			150		
No of choices	7000			1500			1500			1500			1500		
Log Likelihood	-3353			-688			-734			-774			-653		

Coefficients with $p < 0.05$ in bold.

Table A2: Type III and Type IV splitting the composite dimensions (unanchored)

	III (Baseline)			IVa (PD split)			IVb (AD split)		
Parameter	**Coef**	**SE**	**P**	**Coef**	**SE**	**P**	**Coef**	**SE**	**P**
M3x D	**-0.032**	0.006	0.000	**-0.057**	0.011	0.000	**-0.033**	0.011	0.003
M5x D	**-0.140**	0.006	0.000	**-0.128**	0.011	0.000	**-0.134**	0.011	0.000
SC3x D	**-0.030**	0.006	0.000	**-0.046**	0.011	0.000	**-0.022**	0.011	0.034
SC5x D	**-0.121**	0.006	0.000	**-0.142**	0.012	0.000	**-0.112**	0.012	0.000
UA3x D	**-0.036**	0.006	0.000	-0.011	0.011	0.311	0.001	0.011	0.904
UA5x D	**-0.097**	0.006	0.000	**-0.062**	0.009	0.000	**-0.074**	0.010	0.000
PD3x D	**-0.036**	0.006	0.000				**-0.024**	0.010	0.018
PD5x D	**-0.178**	0.006	0.000				**-0.159**	0.011	0.000
AD3x D	**-0.041**	0.006	0.000	**-0.054**	0.010	0.000			
AD5x D	**-0.163**	0.006	0.000	**-0.177**	0.012	0.000			
Pa3x D				**-0.029**	0.010	0.003			
Pa5x D				**-0.171**	0.012	0.000			
Di3x D				**-0.029**	0.010	0.003			
Di5x D				**-0.062**	0.010	0.000			
An3x D							**-0.020**	0.010	0.035
An5x D							**-0.076**	0.011	0.000
De3x D							**-0.068**	0.010	0.000
De5x D							**-0.194**	0.012	0.000
T	**0.394**	0.019	0.000	**0.442**	0.035	0.000	**0.396**	0.035	0.000
No of respondents	700			225			225		
No of choices	7000			2250			2250		
Log Likelihood	-3353			-1147			-1110		

Coefficients with $p < 0.05$ in bold.

Table A3: Type III and Type V bolting off different EQ dimensions (unanchored)

	III (Baseline)			Va (drop M)			Vb (drop SC)			Vc (drop UA)			Vd (drop PD)			Ve (drop AD)		
Parameter	**Coef**	**SE**	**P**	**Coef**	**SE**	**P**	**Coef**	**SE**	**P**	**Coef**	**SE**	**P**	**Coef**	**SE**	**P**	**Coef**	**SE**	**P**
M3x D	**-0.032**	0.006	0.000				**-0.039**	0.013	0.003	**-0.031**	0.015	0.037	-0.007	0.013	0.574	**-0.054**	0.014	0.000
M5x D	**-0.140**	0.006	0.000				**-0.159**	0.014	0.000	**-0.146**	0.015	0.000	**-0.119**	0.013	0.000	**-0.174**	0.015	0.000
SC3x D	**-0.030**	0.006	0.000	**-0.042**	0.014	0.002				**-0.035**	0.014	0.013	-0.018	0.013	0.153	-0.026	0.014	0.055
SC5x D	**-0.121**	0.006	0.000	**-0.187**	0.014	0.000				**-0.163**	0.016	0.000	**-0.149**	0.013	0.000	**-0.139**	0.014	0.000
UA3x D	**-0.036**	0.006	0.000	**-0.056**	0.014	0.000	**-0.052**	0.013	0.000				**-0.045**	0.013	0.001	**-0.035**	0.013	0.009
UA5x D	**-0.097**	0.006	0.000	**-0.146**	0.014	0.000	**-0.134**	0.013	0.000				**-0.103**	0.013	0.000	**-0.125**	0.014	0.000
PD3x D	**-0.036**	0.006	0.000	**-0.051**	0.015	0.001	0.002	0.013	0.868	**-0.045**	0.014	0.001				**-0.042**	0.014	0.002
PD5x D	**-0.178**	0.006	0.000	**-0.199**	0.014	0.000	**-0.193**	0.014	0.000	**-0.194**	0.015	0.000				**-0.239**	0.016	0.000
AD3x D	**-0.041**	0.006	0.000	**-0.071**	0.014	0.000	**-0.083**	0.012	0.000	**-0.032**	0.013	0.012	**-0.085**	0.013	0.000			
AD5x D	**-0.163**	0.006	0.000	**-0.204**	0.016	0.000	**-0.213**	0.015	0.000	**-0.181**	0.014	0.000	**-0.232**	0.015	0.000			
T	**0.394**	0.019	0.000	**0.383**	0.042	0.000	**0.363**	0.038	0.000	**0.376**	0.047	0.000	**0.450**	0.041	0.000	**0.379**	0.016	0.000
No of respondents	700			148			146			150			150			150		
No of choices	7000			1480			1460			1500			1500			1500		
Log Likelihood	-3353			-647			-696			-658			-709			-672		

Coefficients with $p < 0.05$ in bold.