INTERACTION EFFECTS IN HEALTH STATE VALUATION STUDIES: AN OPTIMAL SCALING APPROACH

Marcel F. Jonker, PhD, Bas Donkers, PhD

PII: S1098-3015(22)04694-0
DOI: https://doi.org/10.1016/j.jval.2022.10.008
Reference: JVAL 3683

To appear in: Value in Health

Received Date: 18 February 2022
Revised Date: 28 September 2022
Accepted Date: 10 October 2022

Please cite this article as: Jonker MF, Donkers B, INTERACTION EFFECTS IN HEALTH STATE VALUATION STUDIES: AN OPTIMAL SCALING APPROACH, Value in Health (2022), doi: https://doi.org/10.1016/j.jval.2022.10.008.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Copyright © 2022, International Society for Pharmacoeconomics and Outcomes Research, Inc. Published by Elsevier Inc.
INTERACTION EFFECTS IN HEALTH STATE VALUATION STUDIES: AN OPTIMAL SCALING APPROACH

Marcel F. Jonker, PhD* 1,2 and Bas Donkers, PhD 1,3

1. Erasmus Choice Modelling Centre, Erasmus University Rotterdam, The Netherlands
2. Erasmus School of Health Policy & Management, Erasmus University Rotterdam, The Netherlands
3. Erasmus School of Economics, Erasmus University Rotterdam, The Netherlands

*corresponding author:

Dr. Marcel F. Jonker
Erasmus University Rotterdam
PO Box 1738
3000DR Rotterdam
The Netherlands
Email: marcel@mfjonker.com

Keywords:
health state valuations, interactions, optimal scaling

Author Contributions
❑ Concept and design: Jonker and Donkers
❑ Obtaining funding: Jonker
❑ Acquisition of data: Jonker
❑ Statistical analysis: Jonker and Donkers
❑ Drafting of the manuscript: Jonker and Donkers
❑ Critical revision of paper for important intellectual content: Jonker and Donkers

Conflict of Interest Disclosures

Dr. Jonker reported being a member of the EuroQol Group and the EuroQol Valuation Working Group. No other disclosures were reported.

Funding/Support
This work was supported by a grant (EQ Project 415-RA) from the EuroQol Research Foundation.

Role of Funder/Sponsor
The funder had no role in the design and conduct of the study; collection, management, analysis and interpretation of the data; preparation, review, or approval of the manuscript, and decision to submit the manuscript for publication.

Acknowledgments: The authors gratefully acknowledge financial support from the EuroQol Research Foundation and emphasize that the views as expressed in this article do not necessarily reflect those of the EuroQol Group.
OBJECTIVE: To introduce a parsimonious modelling approach that enables the estimation of interaction effects in health state valuation studies.

METHODS: Instead of supplementing a main effects model with interactions between each and every level, a more parsimonious optimal scaling approach is proposed. This approach is based on the mapping of health-state levels onto domain-specific continuous scales. The attractiveness of health states is then determined by the importance-weighted optimal scales (i.e. main effects) and the interactions between these domain-specific scales (i.e. interaction effects). The number of interaction terms only depends on the number of health domains. As a result, interactions between dimensions can be included with only a few additional parameters.

EMPIRICAL APPLICATIONS: The proposed models with and without interactions are fitted on three valuation datasets from two different countries, i.e. a Dutch latent-scale discrete choice experiment (DCE) dataset with N=3,699 respondents, an Australian time-trade-off (TTO) dataset with N=400 respondents, and a Dutch DCE with duration dataset with N=788 respondents.

RESULTS: Important interactions between health domains were found in all three applications. The results confirm that the accumulation of health problems within health states has a decreasing marginal effect on health state values. A similar effect is obtained when so-called N3 or N5 terms are included in the model specification, but the inclusion of two-way interactions provides superior model fits.

CONCLUSIONS: The proposed interaction model is parsimonious, produces estimates that are straightforward to interpret, and accommodates the estimation of interaction effects in health state valuation studies with realistic sample size requirements. Not accounting for interactions is shown to result in profoundly biased value sets, particularly in stand-alone DCE with duration studies.
INTRODUCTION

Generic measures of health-related quality of life (HRQoL), such as the EQ-5D and SF-6D, are commonly used in the economic evaluation of health care interventions.\[1\] These measures include a descriptive system and a corresponding utility value set for use in the estimation of quality adjusted life years (QALYs).\[2\] Traditionally, standard gamble (SG) and time trade-off (TTO) methods have been used to obtain these value sets, although more recently discrete choice experiments (DCEs) have increased in popularity.\[3\] Irrespective of the valuation method, however, currently available value sets rely on statistical estimations and are thereby prone to model misspecification.

A well-known form of model misspecification is the estimation of a main-effects model in the presence of interaction effects. The traditional approach to accommodating interaction effects is to extend the main effects specification with interactions based on the number and severity level of the health problems in each health state (e.g.\[4-6\]), including so-called N terms (e.g. N3 or N5) that provide an extra decrement when at least one of the attributes is presented at its most severe level. Alternatively, a full set of two-way interactions between each of the levels of the HRQoL instrument can be included in the model specification (e.g.\[7-9\]). Unfortunately, both approaches are problematic: the first fails to differentiate between the various dimensions of the instrument and introduces undesirable health-state ordering artifacts whereas the inclusion all two-way interactions is only feasible for relatively small instruments, such as the EQ-5D-3L, and produces estimates that are difficult to interpret. Moreover, for larger instruments, such as the EQ-5D-5L or SF-6D, the number of interaction terms becomes unwieldy and the results virtually impossible to interpret – even apart from the required increase in sample size requirements and difficulty of accounting for preference heterogeneity.

In this paper, we therefore propose a new modelling approach to including interactions in health state valuation studies. The proposed approach relies on an optimal scaling method to constructing a parsimonious set of interaction effects, see e.g.\[10-13\]. As will be shown, these interaction effects are not only parsimonious but also more intuitive to interpret than a full set of two-way interactions. More specifically, the interactions are included between the health domains instead of between the various levels of each health domain; as a result, the interactions directly capture what happens when more severe health problems in one domain are combined with more severe health problems in another domain. Apart from the more intuitive interpretation, the proposed modelling approach scales well for larger instruments. This means that it can be used for the EQ-5D-5L, SF-6D or even larger instruments if combined with an appropriately powered TTO or DCE experimental design.
In the remainder of this paper, our optimal scaling approach is first described and subsequently applied to two EQ-5D-3L and one EQ-5D-5L dataset. The presented applications highlight the advantages of our optimal scaling method in three different valuation formats, i.e. latent-scale DCE, TTO, and stand-alone DCE with duration, and serve as examples of how much misspecification bias can be introduced when incorrectly relying on a main-effects utility specification. Accordingly, future data collection efforts are strongly recommended to be adequately powered for the identification and inclusion of interaction effects, which would allow future value sets to avoid such biases and to provide a more accurate representation of respondents’ health state preferences.

METHODS

The default approach to allowing for interactions in health-state valuation models is to interact each of the levels of the HRQoL instrument. With D health dimensions with L levels each, and omitting the perfect health constant that is included in many health state valuation studies, this increases the number of parameters in the utility function from D(L-1) in the main-effects specification to D(L-1) main effects and ½D*(D-1)(L-1)^2 interactions when all levels are interacted. For example, this increases the number of parameters from 10 to 50 for the EQ-5D-3L, from 20 to 180 for the EQ-5D-5L, and from 36 to 612 for the EQ-HWB-S instrument.

This is a vast increase but might still seem manageable for the smaller EQ-5D-3L instrument. In practice, however, the additional 40 interaction parameters in the EQ-5D-3L are already notoriously difficult to interpret: even with large data samples, one generally obtains a mixture of significant and insignificant parameters, with both positive and negative signs,\(^{[7-9]}\) making a clear interpretation of the modelling results very difficult. Moreover, in random parameter models, the increased number of mean-parameters also increases the size of the covariance matrix that needs to be estimated. For example, the number of covariance-parameters in a MIXL model (that is typically used to analyze DCEs) increases from 55 to 1,275 for the EQ-5D-3L, from 210 to 16,290 for the EQ-5D-5L, and from 666 to 187,578 for the EQ-HWB when a full set of two-way interactions is included. Irrespective of the interpretation issue, the latter require sample sizes far beyond what is feasible in typical health-state valuation studies.
Including interactions using an optimal scaling approach

The proposed solution is to avoid interactions between each and every level of all domains and instead to include interactions between the health domains themselves. This can be achieved using an ‘optimal scaling’ approach, where each attribute level is transformed into a value on a domain-specific continuous scale.\textsuperscript{10-13} The value on this scale then affects the health-state value. Without interactions, this optimal scaling model is an exact re-parametrisation of the standard main-effects model (see Figure 1).

More formally, let $X_{tdk}$ denote whether level $k$ of health dimension $d$ is present in valuation task $t$. The standard dummy-coded health-state decrement specification is then given by

\[
HS_{\text{value}} = \sum_{d=1}^{D} \sum_{k=1}^{L_d} \beta_{d,k} \cdot X_{tdk}
\]

(1)

in which the beta-parameters measure the decrement of each non-base case level. In the optimal scaling approach, this specification is re-parametrized by first creating domain-specific optimal scales ($S_{t,d}$) that represent the value of the levels of each domain on a continuous scale:

\[
S_{t,d} = \sum_{k=1}^{L_d} \phi_{d,k} \cdot X_{tdk}
\]

(2)

and then multiplying the value of the optimal scales with domain-specific beta-parameters ($\beta_d$):

\[
HS_{\text{value}} = \sum_{d=1}^{D} \beta_d \cdot S_{t,d}
\]

(3)

To statistically identify this decomposition, it is necessary to impose a constraint on the optimal scales. One possibility is to impose a ‘zero mean/variance of one’ constraint, which is, for example, proposed by Van Rosmalen et al.\textsuperscript{13} However, in the case of health-state valuation instruments it makes more sense to restrict the length of the optimal scales, which would imply that the domain-specific $\beta_d$ parameters directly capture the domains’ relative importance. As shown in Figure 1 we propose to constrain the optimal scales to have unit length, ranging from 0 (no health problems) to 1 (the most severe problems for each dimension) with the intermediate levels positioned in between, i.e.
\[ \varphi_{d,1} = 0 , \]
\[ 0 \leq \varphi_{d,k} \leq 1 \quad \text{for} \quad k = 2, \ldots, L_d - 1 , \]
\[ \varphi_{d,L_d} \equiv 1 \]

for all health dimensions \( d = 1, \ldots, D \). Although equation 4 does not impose monotonically increasing severity levels within each dimension this can easily be included as an additional constraint in the construction of the optimal scales, for example by transforming an unrestricted set of parameters \( \varphi_{d,k}^* \) as follows:

\[ \varphi_{d,1} \equiv 0 , \]
\[ \varphi_{d,k} = \varphi_{d,k-1} + \exp(\varphi_{d,k-1}^*)/(1 + \sum_{m=1}^{L_d-2} \exp(\varphi_{d,m}^*)) \quad \text{for} \quad k = 2, \ldots, L_d - 1 , \]
\[ \varphi_{d,L_d} \equiv 1. \] (5)

Equation 5 implies \( \varphi \)-parameters that are monotonically increasing and bounded between 0 and 1. It is also the specification that is used in the three empirical applications of the optimal scaling method in this paper – although equations (4) and (5) of course only differ for health domains with 4 or more levels. Figure 2 provides a graphical representation of the optimal scaling model applied to the EQ-5D-3L instrument. As shown, similar to the default dummy-coded utility specification, 10 parameters are required: 5 phi-parameters to position the levels 2 on the optimal scales and 5 beta-parameters to capture each domain’s importance.

The optimal scaling decomposition so far only represents a reparameterization of the standard dummy-coded health-state decrements. The main advantage of the optimal scaling approach, however, is that each domain is now represented by a single, continuous scale. This allows for the convenient and parsimonious inclusion of interaction terms in the model specification:

\[ \text{HSvalue}_t = \sum_{d=1}^{D} \beta_d^{\text{main}} * S_{td} + \sum_{d=1}^{D-1} \sum_{e=d+1}^{D} \beta_d^{\text{interaction}} * S_{t,d} * S_{t,e} \] (6)

in which the \( \beta \)-main parameters are complemented by the set of \( D(D-1)/2 \) two-way interactions between the constructed optimal scales. Note that the number of interaction parameters in this specification is considerably smaller than the number of possible level-interactions. Moreover, the number of interaction parameters only depends on the number of health dimensions of the instrument.
and not on the total number of levels. Hence the number of interaction parameters does not increase when moving from, for example, the EQ-5D-3L to the EQ-5D-5L instrument, even though the latter would require many more interaction parameters if all levels were to be interacted.

**Empirical application #1.**

In the first empirical example, the proposed optimal scaling model was applied to the Australian EQ-5D-3L composite time trade-off (cTTO) dataset of Viney et al.[14] This dataset is unique in the sense that it comprises a large selection of health states (i.e. 198 out of 243 possible EQ-5D-3L health states) based on a design that was confirmed to be adequately powered for the recovery of two-way interaction effects based on a simulation study. 45 health states that were considered ‘implausible’ were excluded from the design, i.e. all health states that combined level 3 on mobility with either level 1 on usual activities or self-care. Each of the 417 participants in the valuation study completed 12 cTTO tasks under the supervision of a trained interviewer and received $60 for their participation.

Following state-of-the-art modelling of cTTO data, 17 participants with a positively sloped relationship between their cTTO valuations and the misery index of the health states were excluded from the analysis and the data of the remaining 400 respondents were analyzed using a heteroskedastic Tobit model with censoring at -1.[15] This model assumes the existence of a true but incompletely observable variable (cTTO*) that underlies the observed cTTO values. The likelihood then includes the probability of the cTTO* value being beyond the censored value for all observations with cTTO=-1. Accordingly, the observed cTTO values for respondent i in task t were censored as

$$cTTO_{it} = \begin{cases} cTTO^*_{it} & \text{if } cTTO^*_{it} > -1 \\ -1 & \text{if } cTTO^*_{it} \leq -1 \end{cases}$$  \hspace{1cm} (7)$$

with $cTTO^*_{it}$ assumed to be normally distributed

$$cTTO^*_{it} \sim Normal(\mu_{it}, \sigma_{it})$$  \hspace{1cm} (8)$$

with the mean ($\mu_{it}$) and standard deviation ($\sigma_{it}$) reflecting the average health state value and variation among respondents in their valuation of the health state presented in task t of respondent i, respectively.
In the standard dummy-coded specification the mean is defined as

\[
\mu_{it} = \beta_0 + \beta_1 \text{MO}_{2it} + \beta_2 \text{MO}_{3it} + \beta_3 \text{SC}_{2it} + \beta_4 \text{SC}_{3it} + \beta_5 \text{UA}_{2it} + \beta_6 \text{UA}_{3it} + \\
\beta_7 \text{PD}_{2it} + \beta_8 \text{PD}_{3it} + \beta_9 \text{AD}_{2it} + \beta_{10} \text{AD}_{3it} ,
\]

(9a)

which is subsequently reparametrized as

\[
\mu_{it} = \beta_0 + \beta_1 \text{MO}_{it} + \beta_2 \text{SC}_{it} + \beta_3 \text{UA}_{it} + \beta_4 \text{PD}_{it} + \beta_5 \text{AD}_{it}
\]

(9b)

for the optimal scaling model without interactions. The optimal scales MO, SC, UA, PD, and AD were defined conform equation (5) and Figure 2, that is, constrained to 0 for level 1, estimated as \(q_d\) for level 2 of domain d, and constrained to 1 for level 3.

Similar to Pickard et al. the standard deviation (\(\sigma_{it}\)) of the normal distribution was modelled as a 4\(^{th}\)-order polynomial of the health state values

\[
\sigma_{it} = \exp(\gamma_0 + \gamma_1 \mu_{it} + \gamma_2 \mu_{it}^2 + \gamma_3 \mu_{it}^3 + \gamma_4 \mu_{it}^4) ,
\]

(10)

which ensured that the variances of the predicted values can flexibly depend on the health state severity. Finally, the optimal scaling model was also extended with a full set of two-way interactions:

\[
\mu_{it} = \beta_0 + \beta_1 \text{MO}_{it} + \beta_2 \text{SC}_{it} + \beta_3 \text{UA}_{it} + \beta_4 \text{PD}_{it} + \beta_5 \text{AD}_{it} + \\
\beta_6 \text{MO}_{it}\text{SC}_{it} + \beta_7 \text{MO}_{it}\text{UA}_{it} + \beta_8 \text{MO}_{it}\text{PD}_{it} + \beta_9 \text{MO}_{it}\text{AD}_{it} + \beta_{10} \text{SC}_{it}\text{UA}_{it} + \\
\beta_{11} \text{SC}_{it}\text{PD}_{it} + \beta_{12} \text{SC}_{it}\text{AD}_{it} + \beta_{13} \text{UA}_{it}\text{PD}_{it} + \beta_{14} \text{UA}_{it}\text{AD}_{it} + \beta_{15} \text{AD}_{it}\text{PD}_{it}
\]

(9c)

as well as extended with a standard N3 term, which equals 1 if one or more attributes were presented at level 3 and equals 0 otherwise

\[
\mu_{it} = \beta_0 + \beta_1 \text{MO}_{it} + \beta_2 \text{SC}_{it} + \beta_3 \text{UA}_{it} + \beta_4 \text{PD}_{it} + \beta_5 \text{AD}_{it} + \beta_6 N3_{it} .
\]

(9d)
Empirical application #2.

In the second empirical example, the optimal scaling model was applied to the EQ-5D-3L DCE dataset of Nicolet et al.,[9] which comprises a nationally representative sample of 3,669 Dutch respondents. Each respondent completed 16 pairwise choice tasks based on a Bayesian D-efficient DCE design that contained 400 unique choice tasks split into 25 equally sized design blocks. Similar to the Australian cTTO study, the DCE design was specifically optimized to accommodate the estimation of all main and two-way interaction effects and implausible health states were not excluded from the DCE design. All respondents completed the DCE tasks as part of an online, unattended survey and received a small financial compensation from the survey sample provider.

Given the large sample size and number of questions per respondent, we were able to follow the state-of-the-art modelling approach for DCE data using a mixed logit (MIXL) model specification.[17] This means that all model parameters were assumed to be individual specific to account for heterogeneity in health state valuations across individuals. Conform standard MIXL assumptions, the mixing distribution of the model parameters, $\beta$ and $\phi$, was assumed to be multivariate normal and the observed choices

\[ Y_{itj} \in \{0,1\} , \tag{11} \]

which equal 1 if alternative j was chosen by individual i in choice task t and equals zero otherwise,

\[ (Y_{itj} = 1) \Rightarrow (Y_{itk} = 0, \forall k \neq j) \tag{12} \]

were assumed categorical distributed

\[ Y_{itj} \sim \text{categorical}(P_{itj}) \tag{13} \]

with the probability of choosing each choice option defined as

\[ P_{itj} = \frac{\exp(\text{HSvalue}_{itj})}{\sum_{m=1}^{J} \exp(\text{HSvalue}_{itm})} . \tag{14} \]

Here $\text{HSvalue}_{itj}$ denotes respondent i’s health state value for choice alternative j in choice task t. Because duration of life was not included in the DCE, it was not possible to anchor the DCE derived
Values onto the QALY scale. For the same reason, the health state valuation function does not require modelling of time preferences. As a result, the health state values for the standard MIXL model with dummy-coded health state levels:

$$\text{HSvalue}_{ij} = \beta_{i,1}MO_{itj} + \beta_{i,2}MO_{3itj} + \beta_{i,3}SC_{2itj} + \beta_{i,4}SC_{3itj} + \beta_{i,5}UA_{2itj} + \beta_{i,6}UA_{3itj} + \beta_{i,7}PD_{2itj} + \beta_{i,8}PD_{3itj} + \beta_{i,9}AD_{2itj} + \beta_{i,10}AD_{3itj}$$

(15a)

was reparametrized as:

$$\text{HSvalue}_{ij} = \beta_{i,1}MO_{itj} + \beta_{i,2}SC_{itj} + \beta_{i,3}UA_{itj} + \beta_{i,4}PD_{itj} + \beta_{i,5}AD_{itj}.$$  

(15b)

for the optimal scaling specification without interactions. The optimal scales MO, SC, UA, PD, and AD were again defined conform equation (5), that is, constrained to 0 for level 1, estimated as $\varphi_{l,d}$ for level 2, and constrained to 1 for level 3.

As before, the optimal scaling model without interactions was extended with a full set of two-way interactions

$$\text{HSvalue}_{ij} = \beta_{i,1}MO_{itj} + \beta_{i,2}SC_{itj} + \beta_{i,3}UA_{itj} + \beta_{i,4}PD_{itj} + \beta_{i,5}AD_{itj} + \beta_{i,6}MO_{itj}SC_{itj} + \beta_{i,7}MO_{itj}UA_{itj} + \beta_{i,8}MO_{itj}PD_{itj} + \beta_{i,9}MO_{itj}AD_{itj} + \beta_{i,10}SC_{itj}UA_{itj} + \beta_{i,11}SC_{itj}PD_{itj} + \beta_{i,12}SC_{itj}AD_{itj} + \beta_{i,13}UA_{itj}PD_{itj} + \beta_{i,14}UA_{itj}AD_{itj} + \beta_{i,15}AD_{itj}PD_{itj}$$

(15c)

and also extended with a standard N3 term

$$\text{HSValue}_{ij} = \beta_{i,1}MO_{itj} + \beta_{i,2}SC_{itj} + \beta_{i,3}UA_{itj} + \beta_{i,4}PD_{itj} + \beta_{i,5}AD_{itj} + \beta_{i,6}N3_{itj}$$

(15d)

As mentioned, conforming to standard MIXL model assumptions, the model parameters, $\beta$ and $\varphi^*$, were assumed normally distributed with population mean ($\mu$) and covariance matrix ($\Sigma$)

$$\beta_i \sim \text{MVN}(\mu_\beta, \Sigma_\beta)$$

(16)

$$\varphi_i^* \sim \text{MVN}(\mu_{\varphi^*}, \Sigma_{\varphi^*})$$

(17)
Empirical application #3.

In the third empirical application the optimal scaling model was applied to the EQ-5D-5L dataset of Jonker et al.,\cite{18} which comprises a nationally representative sample of 788 Dutch respondents. Each respondent completed 12 matched pairwise choice tasks that included duration of life based on a Bayesian heterogeneous D-efficient DCE design with multiple design versions. Unlike the previous empirical applications, the DCE design was not optimized for the measurement of interaction effects. In addition, the experimental design was neither optimized for the measurement of non-linear duration preferences nor optimized with a QALY-balanced selection of health states, which are more recent developments in the optimization of DCE with duration designs.\cite{19,20} On the other hand, the design was optimized using supercomputer facilities of the Dutch National Computing Facilities Foundation (NCF) and administered to a relatively large number of respondents. 430 respondents completed the DCE online being part of the scientific Longitudinal Internet Studies for the Social Sciences (LISS) panel\cite{21} and 358 respondents completed the DCE under the supervision of a trained interviewer as an addendum to the Dutch EQ-5D-5L valuation study.\cite{22} Respondents received a small financial compensation for participating in the scientific panel as well as for their participation in the Dutch valuation study, although respondents’ participation in the ex-post paper-and-pencil DCE of the Dutch valuation study was entirely voluntary and could be declined without effect on respondents’ payments.

Similar to the DCE without duration the choice data were analyzed using a MIXL model. Therefore the structure of the MIXL model (equations 11-14) as well as the construction of the optimal scales (equations 16-18) remained the same and only the specification of the health state valuation function was different. More specifically, the utility specification of Jonker et al.\cite{23} was used to accommodate for the inclusion of duration of life and potentially non-linear time preferences of the respondents. Accordingly, the standard MIXL model specification of Jonker et al.

$$
HS_{value_{ij}} = (\beta_{i,0} + \beta_{i,1} MO_{2_{itj}} + \beta_{i,2} MO_{3_{itj}} + \beta_{i,3} MO_{4_{itj}} + \beta_{i,4} MO_{5_{itj}} + \beta_{i,5} SC_{2_{itj}} + \beta_{i,6} SC_{3_{itj}} + \beta_{i,7} SC_{4_{itj}} + \beta_{i,8} SC_{5_{itj}} + \beta_{i,9} UA_{2_{itj}} + \beta_{i,10} UA_{3_{itj}} + \beta_{i,11} UA_{4_{itj}} + \beta_{i,12} UA_{5_{itj}} + \beta_{i,13} PD_{2_{itj}} + \beta_{i,14} PD_{3_{itj}} + \beta_{i,15} PD_{4_{itj}} + \beta_{i,16} PD_{5_{itj}} + \beta_{i,17} AD_{2_{itj}} + \beta_{i,18} AD_{3_{itj}} + \beta_{i,19} AD_{4_{itj}} + \beta_{i,20} AD_{5_{itj}} ) \times NPV_{itj}
$$

was re-parametrized as
HSvalue\(_{itj}\) = (\(\beta_{i,0} + \beta_{i,1}MO_{itj} + \beta_{i,2}SC_{itj} + \beta_{i,3}UA_{itj} + \beta_{i,4}PD_{itj} + \beta_{i,5}AD_{itj}\)) * \(NPV_{itj}\) \hspace{1cm} (18b)

for the optimal scaling model without interactions. Here \(NPV_{itj}\) denotes the net present value of the nominal number of life years \(T\), that is \(NPV_{itj} = \sum_{s=1}^{T_{itj}} PV_{(r,s)}\) with discount rate \(r\). One of the advantages of this specification is that different types of discounting functions can be used to define the present value (PV) of the health states. Here the most commonly used exponential discounting function was used, which is defined as \(PV_{(r,s)} = \exp(-r * s)\). After working out the summation of the PV’s combined with an exponential discount function, the NPV is the standard annuity with exponential discount rate \(r\):

\[
NPV_{itj} = \begin{cases} 
1 & \text{if } r = 0 \\
(1 - \exp(-rT_{itj}))/((\exp(r) - 1) & \text{if } r \neq 0
\end{cases}
\]

When the discount rate is equal to zero, all future life years receive the same unit weight and the NPV equals \(T\). When the discount rate is larger than zero, which is typically the case, future life years that are further in the future receive less weight - with the amount of discounting determined by the discount rate.

As with the other empirical applications, the optimal scaling model without interactions was also extended with a full set of two-way interactions:

\[
HSvalue_{itj} = (\beta_{i,0} + \beta_{i,1}MO_{itj} + \beta_{i,2}SC_{itj} + \beta_{i,3}UA_{itj} + \beta_{i,4}PD_{itj} + \beta_{i,5}AD_{itj} + \\
\beta_{i,6}MO_{itj}SC_{itj} + \beta_{i,7}MO_{itj}UA_{itj} + \beta_{i,8}MO_{itj}PD_{itj} + \beta_{i,9}MO_{itj}AD_{itj} + \beta_{i,10}SC_{itj}UA_{itj} + \\
\beta_{i,11}SC_{itj}PD_{itj} + \beta_{i,12}SC_{itj}AD_{itj} + \beta_{i,13}UA_{itj}PD_{itj} + \beta_{i,14}UA_{itj}AD_{itj} + \beta_{i,15}AD_{itj}PD_{itj}) * \hspace{1cm} (18c)
\]

and extended with a \(N45\) term, which equals 1 if one or more attributes are presented at levels 4 or 5 and equals 0 otherwise

\[
HSvalue_{itj} = (\beta_{i,0} + \beta_{i,1}MO_{itj} + \beta_{i,2}SC_{itj} + \beta_{i,3}UA_{itj} + \beta_{i,4}PD_{itj} + \beta_{i,5}AD_{itj} + \beta_{i,6}N45_{itj}) * \hspace{1cm} (18d)
\]

\(NPV_{itj}\).

The online supplemental materials provide similar results for the optimal scaling model extended
with an N3 term, which equals 1 if one or more attributes are presented at level 3 and equals zero otherwise; this specification had an inferior model fit compared to the model with the N45 term.

Model estimation

The optimal-scaling models can, similar to the standard dummy-coded heteroskedastic Tobit and MIXL models, be estimated using classical and Bayesian methods. In this paper, Bayesian Markov Chain Monte Carlo (MCMC) methods were used, which involves the selection of prior distributions for the unknown parameters and updating these via the likelihood of the observed data. For the heteroskedastic Tobit models, uninformative normal priors (i.e., with a mean of zero and standard deviation of 10) were assigned to the $\beta$, $\gamma$, and $\varphi^*$. For the MIXL models, the same uninformative normal priors were assigned to the population mean parameters ($\mu_\beta$, $\mu_\varphi$) and an uninformative Wishart prior with an identity scale matrix with degrees of freedom equal to the number of parameters to the inverse covariance matrix and a uniform (0,1) prior to the interest rate.

For the MIXL models, standard Gibbs update steps were used to update the population level mean and covariance parameters, slice update steps to update the discount rate, and a custom-implemented Metropolis-within-Gibbs algorithm with antithetic sampling to update the individual level MIXL and parameters in the cTTO model. All estimations were implemented in the BUGS language and fitted using OpenBUGS. The cTTO model estimates were based on 10,000 burn-in iterations to let three chains converge and a total of 30,000 MCMC iterations to reliably approximate the posterior distributions of the parameters. The MIXL model estimates used 25,000 burn-in iterations to let three chains converge and subsequently 75,000 MCMC iterations to reliably approximate the posterior distributions. To speed-up the MIXL computations, the normal likelihood (for the WAIC computations, see below), categorical probability (for the MIXL models), and ordered logistic normal computations (for the EQ-5D-5L $\varphi$-parameters) were implemented as user-written functions. Convergence was evaluated based on a visual inspection of the MCMC chains and the convergence diagnostics as implemented in the OpenBUGS package. Note that all BUGS model codes are included in the online supplemental and that a pre-compiled version of OpenBUGS with the user-written functions included is available upon request from the first author.

Model comparisons

The statistical fit of the optimal scaling models with and without interactions was compared based on the Watanabe-Akaike information criterion (WAIC). In addition, the impact of the
reparameterization of the standard heteroskedastic Tobit and MIXL model as well as the inclusion of interactions in the optimal scaling approach was evaluated using plots of the mean-posterior health-state values of all health states as defined by the EQ-5D instruments. For the cTTO and DCE with duration (i.e. empirical applications 1 and 3), the health state values are all anchored on the QALY scale and therefore directly comparable. For the latent-scale DCE without duration, all health-state values are rescaled from 0 (for the worst health state) to 1 (for the best health state) to allow for a comparison between the models with and without interactions.

RESULTS

Table 1 presents the heteroskedastic Tobit estimates for the optimal scaling specifications with and without interactions. Starting with the γ parameters, there is clear evidence that the heterogeneity of the respondents’ health state valuations depends on the health state severity and hence that the implemented Tobit model should indeed accommodate heteroskedasticity. Based on the estimated polynomials, the SD of the cTTO predictions reduce from approximately 0.55 to 0.25 for health states with predicted values smaller than 0.4 versus larger than 0.8, respectively. Turning to the φ-parameters, which indicate how the second level of each health state dimension relates to the best and worst level, these range from 0.12 to 0.61 and are largest for the self-care and anxiety/depression domains. Of course, the overall impact of the levels 2 depends not only on their position on the optimal scale (φ) but also on the relative importance of the optimal scales (β) themselves. Both the value and interpretation of the β-parameters differs between the model specifications. In the main-effects specification without interactions, the constant (β₀) is 0.86, which increases to 0.90 and 0.91 in the models with the N3 and full set of interactions, respectively. In the main-effects specification, the β₁−₅ parameters directly capture the effect of levels 3 in each health domain and thus immediately reflect the relative importance of the five EQ-5D domains. Based on the main-effects β-parameters, pain/discomfort and anxiety/depression are considered most important, closely followed by mobility and with self-care and usual activities being least important. The effect of level 2 in each health domain is obtained by multiplying the corresponding β₁−₅ and φ₁−₅-parameters; hence the overall importance of level 2 of a domain with a large φ-parameter (e.g. self-care) can be smaller than that of a domain with a relatively small φ combined with a larger β-parameter (e.g. pain/discomfort). In the specification with the N3 term, the main effect parameters are slightly smaller but this effect is compensated for via the negative coefficient of the N3 term. This model has an improved model fit in terms of WAIC, which can already be interpreted as evidence of the presence and relevance of interaction effects. In the specification with a full set of two-way interactions between the optimal
scales, the interpretation of the main effects $\beta$-parameters is slightly different. Here the $\beta_{1-5}$ parameters reflect the impact of levels 3 of each domain in the absence of any problems in the other domains, resulting in more negative main effects. When health problems across multiple dimensions are combined within a health state, the positive interaction parameters counteract these more negative main effects. As shown in Table 1, almost all interaction parameters are positive and half are significant (i.e. have 90% or 95% CI that do not comprise zero). Only one interaction, the interaction between self-care and usual activities, is negative (and significant). In terms of WAIC, the optimal scaling model with a full set of two-way interactions provides the best model fit, considerably better than the other specifications. Finally, as shown in Table 1, the inclusion of interaction effects increases the value of the worst possible health state from -0.39 in the main-effects model to -0.23 in the two-way interaction model.

Tables 2 and 3 provide the MIXL estimates of the optimal scaling models with and without interactions for the latent scale EQ-5D-3L and EQ-5D-5L with duration samples, respectively. Because the presented models are all on their own latent utility scales, the estimates cannot be compared directly. It is nonetheless possible to compare the estimates in terms of signs, relative magnitudes, and significance (i.e. whether the CI comprise zero or not). As can be seen, a major difference between the two DCEs is the relative importance of the mobility domain. In the EQ-5D-3L sample, mobility (with level 3 defined as being confined to bed) is the most important attribute whereas mobility is one of least important attributes in the EQ-5D-5L sample (with level 5 defined as having extreme problems with mobility). In all other aspects the estimates are very comparable and also very similar to those reported in the cTTO application. For example, the accommodation of interaction effects via the N3 and N45 terms improves the WAIC model fits. When N3 and N45 terms are included, the main-effects estimates become smaller, which is compensated for by the negative parameter estimates of the interaction terms. Furthermore, in both DCEs the optimal scaling model with all two-way interactions has the best WAIC model fit. In addition, all interaction parameter estimates are positive and significant, apart from the interactions between anxiety/depression and pain/discomfort in the 5L and anxiety/depression and usual activities in the 3L application. Similar to the cTTO results, the inclusion of two-way interactions in the DCE with duration MIXL model increases the QALY-anchored value of the worst possible health state (i.e. 5-5-5-5-5) from -1.34 in the main-effects specification to -0.10 in the interaction model. Of course, such a comparison cannot be made in the latent-scale DCE.

Figure 3 provides graphical comparisons between the main-effects optimal scaling model on the horizontal axis and the corresponding standard main-effects dummy-coded heteroskedastic Tobit and
MIXL models (top), optimal scaling models with N3/N45 terms (middle), and full set of two-way interactions (bottom) on the vertical axis. For the cTTO and DCE with duration applications, the scatterplots are based on EQ-5D health states anchored on the 0-1 QALY scale, whereas the latent-scale DCE values are rescaled between 0 and 1 for the best and worst possible health state. As shown in the top panels, all points are on or very close to the diagonal, which confirms that the main-effects optimal scaling model and standard dummy-coded models are re-parametrizations of one another and result in close to identical predicted health state values. The middle panels show that the inclusion of the N3 and N45 terms already capture some degree of decreasing marginal effects of the accumulation of health problems in health states on health state values. And as shown in the bottom panels, the inclusion of two-way interactions has a considerably stronger impact on the distribution of health state values than the inclusion of N3/N45 terms. The impact of the interactions is largest in the DCE with duration application, reflecting the fact that the inclusion of the interaction effects not only affects the ordering of the health states but also the model-based anchoring of the health states on the QALY scale.

DISCUSSION

This paper has introduced a parsimonious modelling approach that enables the estimation of interaction effects in health state valuation studies. Instead of supplementing a main effects model with interactions between each and every level, which produces estimates that are difficult to interpret and requires very large sample sizes, a more parsimonious optimal scaling approach is proposed. This approach is based on the mapping of the health levels in each health domain onto domain-specific continuous scales. The attractiveness of each health state is then determined as the sum of the relative importance of and interactions between the optimal scales, which implies that the number of model parameters only depends on the number of dimensions of the instrument.

Based on three different empirical applications, one based on cTTO, one on a latent scale DCE, and one on a DCE with duration, the inclusion of two-way interactions is shown to not only improve statistical model fits but also to affect the ordering of and relative distance between health states. This occurs due to the positive interaction effects and subsequent compression of values for health states that combine multiple health problems. The effect is particularly pronounced in the DCE with duration application where not only the relative importance of health states but also the anchoring of health states onto the QALY scale is affected by the model specification. cTTO-based valuations are comparatively robust to model misspecification because it is directly observed whether health states...
that are valued by respondents are better or worse than immediate death. DCE with duration, on the other hand, requires a model-based extrapolation to the point of zero duration to determine the anchor point of the QALY scale.

One of the major advantages of the optimal scaling model is that it produces estimates of main effects and interaction terms that are straight-forward to interpret. This is uniquely different from previous attempts, in which the mixture of positive and negative parameter estimates with varying sizes and statistical significance was difficult, or even impossible, to interpret. Based on the subset of the most significant interaction effects, previous investigations did find that the inclusion of interactions provides a more accurate description of respondents’ preferences. Moreover, the most significant interactions were “generally positive and therefore in the opposite direction to the main effects”, which concurs with our own findings. This suggests that the diminishing marginal effect of accumulated health problems on the health state values is relevant to many valuation studies, and also relevant for the valuation of bolt-on dimensions and/or the application of reduced parameter estimation models.

Another advantage of the optimal scaling model is that it is sufficiently parsimonious to be applicable to a wide range of HRQoL instruments. As shown, it can be used for EQ-5D-3L and EQ-5D-5L instruments and for cTTO, latent scale DCE, and stand-alone DCE applications alike. Also, in comparison to the conventional approach of including all two-way interactions between the levels of each of the domains, an enormous reduction in the number of model parameters is achieved. For example, the cTTO heteroskedastic optimal scaling Tobit model with fixed parameters and a full set of two-way interactions requires 54% and 81% fewer parameters for the EQ-5D-3L and EQ-5D-5L, respectively. For the MIXL models with random parameters, the difference is even more substantial with a corresponding reduction of 87% and 98% for the latent scale DCE and 88% and 99% for the DCE with duration model. Consequently, the optimal scaling model allows for the inclusion of interaction effects in health state valuation studies for which it was previously not feasible to include a full set of two-way interactions.

In addition, even without the inclusion of interactions in the model specification the optimal scaling model has several attractive properties. For example, the optimal scaling model conveniently allows the levels within each health dimension to be monotonically constrained (e.g. via a logistic-normal distribution as in equation 5). Of course, the model can also be estimated without imposing a logical ordering of the level decrements (cf. equation 4), but in health state valuation studies it often makes sense to preclude illogical preference reversals. In addition, unlike the standard MIXL model, the
optimal scaling model does not comprise a single variance-covariance matrix for all health-state decrements. Instead, the decomposition of the level decrements allows for separate decisions about whether to allow for respondent-specific parameters in the health-dimension importance (β) and within-optimal scale level severity (φ) parameters. For larger HRQoL instruments this provides additional flexibility to create more parsimonious models than the standard MIXL model. This may not be necessary for instruments such as the EQ-5D-3L and EQ-5D-5L but could be essential for the valuation of substantially larger instruments like the WOOP and EQ-HWB.[28-29]

In terms of data quality, the latent scale DCE dataset provided an ideal testing ground for the optimal scaling model. It was based on a Bayesian d-efficient DCE design that was optimized for the measurement of two-way interactions, included N=3,699 respondents, and there were no combinations of health state levels systematically blocked from the DCE design. In contrast, the DCE with duration sample was based on a DCE design that was not explicitly optimized for the identification of two-way interactions, did not comprise a QALY-balanced selection of health states, and was also not optimized for the measurement of non-linear duration. The latter are more recent design improvements that have been found to be particularly important to obtain unbiased DCE with duration results.[19-20] The sample of N=788 respondents still provided adequate statistical power to detect interaction effects, which is likely related to the efficient design optimization using supercomputer facilities. The Australian cTTO study was based on a large selection of health states (i.e. 198 or 81% of all 243 possible health states) and a simulation study to confirm statistical identification of two-way interaction effects. Unfortunately, all combinations of mobility level 3 with levels 1 on usual activities and self-care (i.e. 19% of all possible health states) were blocked from the DCE design. The unexpected negative and significant interaction effect between usual activities and self-care could be related to this. The most important limitation, however, was the relatively small sample size of N=400 respondents, which precluded the inclusion of random parameters in the model specification and resulted in half of the interaction parameters having 90% or 95% confidence intervals that comprised zero. A larger sample would clearly have been beneficial.

The presented modelling approach also has some limitations. For example, even though monotonicity of within-domain preferences can be imposed in the construction of the optimal scales, the combined impact of the interaction effects has not been restricted to be smaller than the main effect of each domain. Of course, such a constraint could be imposed when value sets are computed since it avoids illogical preference reversals in the interaction models. Moreover, the positioning of the intermediate levels on a single optimal scale implies that intermediate levels are
assumed to have a proportional impact on the main and interaction effects. On the one hand, we consider this a reasonable trade-off between model parsimony and flexibility. On the other hand, it can easily be relaxed by allowing for separate optimal scales for the main and interaction effects if the available sample size permits it. Appendix C in the online supplemental presents the model estimates of the optimal scaling model with separate optimal scales for the main and interaction effects in the latent scale DCE dataset. The results indicate that for the mobility, pain/discomfort, and anxiety/depression domains intermediate levels more strongly affect the interaction than main effects, resulting in an improved WAIC of 54,664.

A more practical limitation of the optimal scaling approach is that the inclusion of the interaction effects requires an experimental design that identifies the main effects and interactions and also requires a larger sample size compared to that of a main-effects only model specification. This is less of a problem for future health state valuation studies, for which the experimental design can be adjusted and sample sizes increased, but implies that the model will not be estimable for many existing studies that are optimized for a main-effects only specification. For example, the vast majority of EQ-5D-5L valuation studies only includes a small subset of 86 out of all 3,125 possible health states, which already results in problems with overfitting for main-effects models[30] and does not accommodate for reliable estimations of two-way interaction models. Conform the simulation study of Viney et al.[14] a large selection of health states is required to accommodate the estimation of two-way interaction effects. In this sense, our recommendation to accommodate for interaction effects in health state valuation studies clearly deviates from recent efforts to further reduce the number of health states and respondents in cTTO valuation studies.[31-32]

CONCLUSION

In conclusion, we have introduced an interaction model that is not only parsimonious but also produces estimates that are straight-forward to interpret and accommodates the estimation of interactions in health state valuation studies for which it was thus far not feasible to include interactions. Substantial evidence of decreasing marginal effects of the accumulation of health problems in health states on health state values was found. Therefore, we recommend future valuation studies to power their data collection for the inclusion of interaction effects and to include interactions based on the proposed optimal scaling approach in the statistical analyses.
REFERENCES

1. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. Oxford university press; 2015
2. Brazier J, Ratcliffe J, Saloman J, Tsuchiya A. Measuring and valuing health benefits for economic evaluation. OXFORD university press; 2017
3. Mulhern B, Norman R, Street DJ, Viney R. One method, many methodological choices: a structured review of discrete-choice experiments for health state valuation. Pharmacoeconomics. 2019;37(1):29-43
4. Augustovski FA, Irazola VE, Velazquez AP, Gibbons L, Craig BM. Argentine valuation of the EQ-5D health states. Value in health. 2009 Jun 1;12(4):587-96
5. Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. Medical care. 2005 Mar 1:203-20
6. Tongsiri S, Cairns J. Estimating population-based values for EQ-5D health states in Thailand. Value in Health. 2011 Dec 1;14(8):1142-5
7. Viney R, Norman R, King MT, Cronin P, Street DJ, Knox S, Ratcliffe J. Time trade-off derived EQ-5D weights for Australia. Value in Health. 2011 Sep 1;14(6):928-36
8. Kularatna S, Whitty JA, Johnson NW, Jayasinghe R, Scuffham PA. Valuing EQ-5D health states for Sri Lanka. Quality of Life Research. 2015 Jul;24(7):1785-93
9. Nicolet A, Groothuis-Oudshoorn CG, Krabbe PF. Does inclusion of interactions result in higher precision of estimated health state values?. Value in health. 2018 Dec 1;21(12):1437-44
10. Young FW, De Leeuw J, Takane Y. Regression with qualitative and quantitative variables: An alternating least squares method with optimal scaling features. Psychometrika. 1976 Dec;41(4):505-29
11. Gifi A. Nonlinear multivariate analysis. Wiley-Blackwell; 1990
12. Linting M, Meulman JJ, Groenen PJ, Van der Kooij AJ. Stability of nonlinear principal components analysis: An empirical study using the balanced bootstrap. Psychological methods. 2007 Sep;12(3):359
13. van Rosmalen J, Koning AJ, Groenen PJ. Optimal scaling of interaction effects in generalized linear models. Multivariate Behavioral Research. 2009 Feb 24;44(1):59-81
14. Viney R, Norman R, King MT, Cronin P, Street DJ, Knox S, Ratcliffe J. Time trade-off derived EQ-5D weights for Australia. Value in Health. 2011 Sep 1;14(6):928-36.
15. Rowen D, Mukuria C, McDool E. A Systematic Review of the Methodologies and Modelling Approaches Used to Generate International EQ-5D-5L Value Sets. PharmacoEconomics. 2022 Jul 13:1-20.
16. Pickard AS, Law EH, Jiang R, Fundinayegani E, Shaw JW, Xie F, Oppe M, Boye KS, Chapman RH, Gong CL, Balch A. United States valuation of EQ-5D-5L health states using an international protocol. Value in Health. 2019 Aug 1;22(8):931-41.

17. Soekhai V, de Bekker-Grob EW, Ellis AR, Vass CM. Discrete choice experiments in health economics: past, present and future. Pharmacoeconomics. 2019 Feb;37(2):201-26.

18. Jonker MF, Attema AE, Donkers B, Stolk EA, Versteegh MM. Are health state valuations from the general public biased? A test of health state reference dependency using self-assessed health and an efficient discrete choice experiment. Health economics. 2017 Dec;26(12):1534-47.

19. Lim S, Jonker MF, Oppe M, Donkers B, Stolk E. Severity-stratified discrete choice experiment designs for health state evaluations. Pharmacoeconomics. 2018 Nov;36(11):1377-89.

20. Jonker MF, Bliemer MC. On the optimization of Bayesian D-efficient discrete choice experiment designs for the estimation of QALY tariffs that are corrected for nonlinear time preferences. Value in Health. 2019 Oct 1;22(10):1162-9.

21. Scherpenzeel AC. “‘True” Longitudinal and Probability-Based Internet Panels: Evidence From the Netherlands. InSocial and behavioral research and the Internet 2018 Oct 24 (pp. 77-104). Routledge.

22. Versteegh MM, Vermeulen KM, Evers SM, De Wit GA, Prenger R, Stolk EA. Dutch tariff for the five-level version of EQ-5D. Value in health. 2016 Jun 1;19(4):343-52.

23. Jonker MF, Donkers B, de Bekker-Grob EW, Stolk EA. Advocating a paradigm shift in health-state valuations: the estimation of time-preference corrected QALY tariffs. Value in Health. 2018 Aug 1;21(8):993-1001.

24 Bédard M, Douc R, Moulines E. Scaling analysis of delayed rejection MCMC methods. Methodology and Computing in Applied Probability. 2014 Dec;16(4):811-38.

25. Gelman A, Hwang J, Vehtari A. Understanding predictive information criteria for Bayesian models. Statistics and computing. 2014 Nov;24(6):997-1016.

26. Stolk EA, Oppe M, Scalone L, Krabbe PF. Discrete choice modeling for the quantification of health states: the case of the EQ-5D. Value in health. 2010 Dec;13(8):1005-13.

27. Rand-Hendriksen K, Ramos-Goñi JM, Augestad LA, Luo N. Less is more: cross-validation testing of simplified nonlinear regression model specifications for EQ-5D-5L health state values. Value in Health. 2017 Jul 1;20(7):945-52

28. Hackert MQ, van Exel J, Brouwer WB. Well-being of Older People (WOOP): Quantitative validation of a new outcome measure for use in economic evaluations. Social Science & Medicine. 2020 Aug 1;259:113109.

29. Peasgood T, Mukuria C, Brazier J, Marten O, Kreimeier S, Luo N, Mulhern B, Greiner W, Pickard AS, Augustovski F, Engel L. Developing a New Generic Health and Wellbeing Measure:
30. Rand-Hendriksen K, Ramos-Goñi JM, Augestad LA, Luo N. Less is more: cross-validation testing of simplified nonlinear regression model specifications for EQ-5D-5L health state values. Value in Health. 2017 Jul 1;20(7):945-52.

31. Yang Z, Luo N, Oppe M, Bonsel G, Busschbach J, Stolk E. Toward a smaller design for EQ-5D-5L valuation studies. Value in Health. 2019 Nov 1;22(11):1295-302.

32. Oppe M, Norman R, Yang Z, van Hout B. Experimental Design for the Valuation of the EQ-5D-5L. InValue Sets for EQ-5D-5L 2022 (pp. 29-54). Springer, Cham.
Figure 1. Example comparison between standard (dummy-coded) health-state decrements and the optimal scaling re-parameterization *

A. Standard (dummy-coded) health-state decrements

| Level  | β₁ | β₂ | β₃ |
|--------|----|----|----|
| Level 1| 0  |    |    |
| Level 2|    | β₁=-0.10 |    |
| Level 3|    | β₂=-0.25 |    |
| Level 4|    | β₃=-0.50 |    |

B. Optimal scaling approach (worst level constrained to 1)

| Level  | φ₁ | φ₂ |
|--------|----|----|
| Level  | 0  |    |
| Level 2| φ₁=0.20 |    |
| Level 3| φ₂=0.50 |    |
| Level 4| 1.00 (worst level) |    |

β₁=-0.50 (health domain importance)

* For a health domain with four levels, three parameters are required to model the QALY or utility decrements: a) β₁-3 when standard dummy-coded health-state decrements are used (left), or b) φ₁-2 to construct the optimal scale plus β₁ to measure the domain’s importance under the optimal scaling approach (right).
Figure 2. Graphical representation of the optimal scaling model without interactions for the EQ-5D-3L instrument **

| Level | Mobility | Selfcare | Usual activities | Pain / discomfort | Anxiety / depression |
|-------|----------|----------|------------------|-------------------|---------------------|
| Level 1 (best level) | 0 | 0 | 0 | 0 | 0 |
| Level 2 | $\phi_1$ | $\phi_2$ | $\phi_3$ | $\phi_4$ | $\phi_5$ |
| Level 3 (worst level) | 1 | 1 | 1 | 1 | 1 |

attribute importances: $\beta_{1-5}$

* In the optimal scaling model without interactions, 5 phi-parameters ($\phi_{1-5}$) and 5 beta-parameters ($\beta_{1-5}$) are required to construct the optimal scales and to capture the 5 EQ domains’ importances, respectively. ** 10 additional phi-parameters would be required to construct the optimal scales for the EQ-5D-5L instrument. For both the EQ-5D-3L and 5L instruments, only 10 interaction parameters ($\beta_{6-15}$) are required to include a full set of two-way interactions between the optimal scales.
Table 1. Optimal scaling heteroskedastic Tobit model estimates for the CPTO application

|                  | 1. without interactions | 2. with N3 term | 3. with full set of interactions |
|------------------|-------------------------|-----------------|----------------------------------|
| $\beta_0$ intercept | 0.85 (0.01) ***          | 0.89 (0.01) *** | 0.90 (0.02) ***                  |
| $\beta_1$ mobility (MO) | -0.27 (0.02) ***         | -0.27 (0.02) *** | -0.45 (0.05) ***                |
| $\beta_2$ self-care (SC) | -0.20 (0.02) ***         | -0.15 (0.02) *** | -0.26 (0.03) ***                |
| $\beta_3$ usual activities (UA) | -0.12 (0.01) ***         | -0.05 (0.02) *** | -0.12 (0.03) ***                |
| $\beta_4$ pain/discomfort (PD) | -0.33 (0.02) ***         | -0.27 (0.02) *** | -0.51 (0.03) ***                |
| $\beta_5$ anxiety/depression (AD) | -0.32 (0.01) ***         | -0.25 (0.02) *** | -0.49 (0.03) ***                |
| $\beta_6$ MO x SC | n.a.                    | n.a.            | 0.12 (0.06) **                  |
| $\beta_7$ MO x UA | n.a.                    | n.a.            | 0.02 (0.05)                    |
| $\beta_8$ MO x PD | n.a.                    | n.a.            | 0.09 (0.05) **                  |
| $\beta_9$ MO x AD | n.a.                    | n.a.            | 0.02 (0.05)                    |
| $\beta_{10}$ SC x UA | n.a.                    | n.a.            | -0.10 (0.04) ***               |
| $\beta_{11}$ SC x PD | n.a.                    | n.a.            | 0.06 (0.05)                    |
| $\beta_{12}$ SC x AD | n.a.                    | n.a.            | 0.15 (0.05) ***                |
| $\beta_{13}$ UA x PD | n.a.                    | n.a.            | 0.04 (0.04)                    |
| $\beta_{14}$ UA x AD | n.a.                    | n.a.            | 0.06 (0.04)                    |
| $\beta_{15}$ PD x AD | n.a.                    | n.a.            | 0.26 (0.04) ***                |
| $\beta_{16}$ N3 | n.a.                    | -0.18 (0.02) *** | n.a.                            |
| $\varphi_1$ MO level 2 **** | 0.18 (0.04) ***         | 0.20 (0.04) *** | 0.15 (0.04) ***                |
| $\varphi_2$ SC level 2 **** | 0.43 (0.06) ***         | 0.63 (0.10) *** | 0.35 (0.05) ***                |
| $\varphi_3$ UA level 2 **** | 0.07 (0.09) ***         | 0.09 (0.21) *** | 0.26 (0.12) ***                |
| $\varphi_4$ PD level 2 **** | -0.13 (0.04) ***        | 0.22 (0.05) *** | 0.12 (0.03) ***                |
| $\varphi_5$ AD level 2 **** | 0.34 (0.03) ***         | 0.44 (0.05) *** | 0.31 (0.03) ***                |
| $\gamma_1$ intercept | -0.57 (0.03) ***        | -0.56 (0.02) *** | -0.58 (0.02) ***               |
| $\gamma_2$ hs-value | 0.48 (0.16) ***         | 0.38 (0.13) *** | 0.47 (0.09) ***                |
| $\gamma_3$ hs-value^2 | -0.78 (0.22) ***        | -1.53 (0.34) *** | -1.41 (0.38) ***               |
| $\gamma_4$ hs-value^3 | -2.33 (0.97) ***        | -0.93 (1.30) *** | -1.97 (0.99) **                |
| $\gamma_5$ hs-value^4 | 1.31 (1.03)             | 1.09 (1.13)     | 2.07 (0.79) ***                |
| 3-3-3-3-3 (pits) | -0.39                   | -0.29           | -0.23                          |
| WAIC (value, lower=better) | 15,078                   | 14,990          | 14,942                         |
| WAIC (rank) | 3                       | 2               | 1                              |

* Note: mean posterior estimates with standard deviations in parenthesis ** 90% CI does not contain zero *** 95% CI does not contain zero **** estimates on a scale from 0 (EQ-5D levels 1) to 1 (EQ-5D levels 3).
Table 2. Optimal scaling mixed logit model estimates for the latent scale DCE application *

|                | 1. without interactions | 2. with N3 term          | 3. with full set of interactions |
|----------------|--------------------------|--------------------------|---------------------------------|
|                | population mean | population SD  | population mean | population SD  | population mean | population SD  |
| \( \beta_1 \) mobility (MO)    & -2.68 (0.05)*** & 1.80 (0.05)***       & -2.62 (0.05)*** & 1.80 (0.05)***       & -3.32 (0.10)*** & 2.06 (0.11)***       |
| \( \beta_2 \) self-care (SC)     & -1.74 (0.04)*** & 1.16 (0.05)***       & -1.67 (0.04)*** & 1.16 (0.05)***       & -2.21 (0.09)*** & 1.26 (0.13)***       |
| \( \beta_3 \) usual activities (UA) & -1.73 (0.04)*** & 1.33 (0.04)***       & -1.68 (0.04)*** & 1.33 (0.04)***       & -2.17 (0.10)*** & 1.50 (0.13)***       |
| \( \beta_4 \) pain/discomfort (PD) & -2.54 (0.05)*** & 2.03 (0.05)***       & -2.48 (0.05)*** & 2.03 (0.05)***       & -3.05 (0.10)*** & 2.36 (0.12)***       |
| \( \beta_5 \) anxiety/depression (AD) & -2.20 (0.05)*** & 1.94 (0.05)***       & -2.15 (0.05)*** & 1.94 (0.05)***       & -2.63 (0.10)*** & 2.25 (0.10)***       |
| \( \beta_6 \) MO x SC             & n.a.        | n.a.                | n.a.        | n.a.                | 0.46 (0.07)*** & 0.50 (0.08)***       |
| \( \beta_7 \) MO x UA             & n.a.        | n.a.                | n.a.        | n.a.                | 0.32 (0.06)*** & 0.50 (0.08)***       |
| \( \beta_8 \) MO x PD             & n.a.        | n.a.                | n.a.        | n.a.                | 0.28 (0.07)*** & 0.43 (0.08)***       |
| \( \beta_9 \) MO x AD             & n.a.        | n.a.                | n.a.        | n.a.                | 0.27 (0.07)*** & 0.54 (0.08)***       |
| \( \beta_{10} \) SC x UA          & n.a.        | n.a.                | n.a.        | n.a.                | 0.19 (0.06)*** & 0.41 (0.07)***       |
| \( \beta_{11} \) SC x PD          & n.a.        | n.a.                | n.a.        | n.a.                | 0.18 (0.07)*** & 0.49 (0.08)***       |
| \( \beta_{12} \) SC x AD          & n.a.        | n.a.                | n.a.        | n.a.                | 0.20 (0.06)*** & 0.43 (0.07)***       |
| \( \beta_{13} \) UA x PD          & n.a.        | n.a.                | n.a.        | n.a.                | 0.29 (0.07)*** & 0.40 (0.08)***       |
| \( \beta_{14} \) UA x AD          & n.a.        | n.a.                | n.a.        | n.a.                | 0.08 (0.07)*** & 0.53 (0.09)***       |
| \( \beta_{15} \) PD x AD          & n.a.        | n.a.                | n.a.        | n.a.                | 0.31 (0.07)*** & 0.64 (0.08)***       |
| \( \beta_{16} \) N3                 & n.a.        | n.a.                | -0.28 (0.05)*** & 0.39 (0.05)***       & n.a.        | n.a.                |
| \( \phi_1 \) MO level 2 ***       & 0.18 (0.01)*** & 0.20 (0.01)***       & 0.19 (0.01)*** & 0.20 (0.01)***       & 0.20 (0.01)*** & 0.20 (0.01)***       |
| \( \phi_2 \) SC level 2 ***       & 0.35 (0.01)*** & 0.14 (0.02)***       & 0.37 (0.01)*** & 0.13 (0.02)***       & 0.37 (0.01)*** & 0.13 (0.02)***       |
| \( \phi_3 \) UA level 2 ***       & 0.24 (0.01)*** & 0.16 (0.03)***       & 0.25 (0.01)*** & 0.16 (0.02)***       & 0.25 (0.01)*** & 0.15 (0.02)***       |
| \( \phi_4 \) PD level 2 ***       & 0.22 (0.01)*** & 0.12 (0.02)***       & 0.23 (0.01)*** & 0.12 (0.01)***       & 0.24 (0.01)*** & 0.12 (0.01)***       |
| \( \phi_5 \) AD level 2 ***       & 0.30 (0.01)*** & 0.09 (0.01)***       & 0.32 (0.01)*** & 0.08 (0.01)***       & 0.31 (0.01)*** & 0.09 (0.01)***       |
| WAIC (value, lower=better)         & 55,390       | 55,387              & 55,001       |                    |
| WAIC (rank)                        & 3            | 2                    | 1             |                    |

* Note: mean posterior estimates with standard deviations in parenthesis ** 90% CI does not contain zero *** 95% CI does not contain zero **** estimates on a scale from 0 (EQ-5D level 1) to 1 (EQ-5D level 3).
Table 3. Optimal scaling mixed logit model estimates for the DCE with duration application *

| β1    | perfect health       | 1.86 (0.11)***  | 1.56 (0.13)***  | 1.90 (0.11)***  | 1.62 (0.12)***  | 1.74 (0.10)***  | 1.41 (0.12)***  |
|-------|----------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| β2    | mobility (MO)        | -0.66 (0.04)***  | 0.51 (0.04)***   | -0.53 (0.04)***  | 0.51 (0.04)***   | -0.75 (0.04)***  | 0.58 (0.04)***   |
| β3    | self-care (SC)       | -0.52 (0.03)***  | 0.38 (0.03)***   | -0.40 (0.03)***  | 0.40 (0.03)***   | -0.64 (0.03)***  | 0.45 (0.03)***   |
| β4    | usual activities (UA)| -0.70 (0.04)***  | 0.43 (0.03)***   | -0.60 (0.06)***  | 0.45 (0.03)***   | -0.81 (0.04)***  | 0.54 (0.04)***   |
| β5    | pain/discomfort (PD) | -1.18 (0.06)***  | 0.68 (0.05)***   | -1.15 (0.06)***  | 0.69 (0.05)***   | -1.20 (0.06)***  | 0.74 (0.05)***   |
| β6    | anxiety/depression (AD)| -1.27 (0.07)***| 0.78 (0.06)***   | -1.17 (0.02)***  | 0.79 (0.05)***   | -1.25 (0.06)***  | 0.79 (0.06)***   |
| β7    | MO x SC              | n.a.             | n.a.             | n.a.             | n.a.             | 0.35 (0.06)***   | 0.47 (0.07)***   |
| β8    | MO x UA              | n.a.             | n.a.             | n.a.             | n.a.             | n.a.             | n.a.             |
| β9    | MO x PD              | n.a.             | n.a.             | n.a.             | n.a.             | 0.32 (0.07)***   | 0.45 (0.07)***   |
| β10   | MO x AD              | n.a.             | n.a.             | n.a.             | n.a.             | 0.33 (0.07)***   | 0.53 (0.08)***   |
| β11   | SC x UA              | n.a.             | n.a.             | n.a.             | n.a.             | 0.43 (0.06)***   | 0.45 (0.06)***   |
| β12   | SC x PD              | n.a.             | n.a.             | n.a.             | n.a.             | 0.28 (0.05)***   | 0.46 (0.06)***   |
| β13   | SC x AD              | n.a.             | n.a.             | n.a.             | n.a.             | 0.26 (0.06)***   | 0.49 (0.07)***   |
| β14   | UA x PD              | n.a.             | n.a.             | n.a.             | n.a.             | 0.25 (0.07)***   | 0.61 (0.09)***   |
| β15   | UA x AD              | n.a.             | n.a.             | n.a.             | n.a.             | 0.21 (0.08)***   | 0.41 (0.07)***   |
| β16   | PD x AD              | n.a.             | n.a.             | n.a.             | n.a.             | 0.05 (0.08)      | 0.63 (0.08)***   |
| β17   | N45                  | n.a.             | n.a.             | n.a.             | n.a.             | 0.17 (0.02)***   | n.a.             |
|       |                      |                  |                  |                  |                  |                  |                  |
| φ1,1  | MO level 2           | 0.15 (0.01)***   | 0.10 (0.01)***   | 0.18 (0.01)***   | 0.10 (0.02)***   | 0.16 (0.01)***   | 0.12 (0.02)***   |
| φ1,2  | MO level 3           | 0.22 (0.01)***   | 0.13 (0.02)***   | 0.27 (0.01)***   | 0.14 (0.02)***   | 0.23 (0.01)***   | 0.14 (0.02)***   |
| φ1,3  | MO level 4           | 0.65 (0.02)***   | 0.10 (0.01)***   | 0.64 (0.02)***   | 0.10 (0.02)***   | 0.74 (0.03)***   | 0.09 (0.02)***   |
| φ2,1  | SC level 2           | 0.15 (0.01)***   | 0.07 (0.02)***   | 0.20 (0.02)***   | 0.09 (0.04)***   | 0.15 (0.02)***   | 0.10 (0.04)***   |
| φ2,2  | SC level 3           | 0.24 (0.02)***   | 0.12 (0.02)***   | 0.31 (0.02)***   | 0.13 (0.04)***   | 0.26 (0.02)***   | 0.14 (0.03)***   |
| φ2,3  | SC level 4           | 0.65 (0.02)***   | 0.09 (0.02)***   | 0.59 (0.02)***   | 0.11 (0.03)***   | 0.73 (0.02)***   | 0.09 (0.02)***   |
| φ3,1  | UA level 2           | 0.12 (0.01)***   | 0.07 (0.01)***   | 0.14 (0.01)***   | 0.08 (0.02)***   | 0.12 (0.01)***   | 0.07 (0.02)***   |
| φ3,2  | UA level 3           | 0.21 (0.01)***   | 0.09 (0.01)***   | 0.25 (0.02)***   | 0.10 (0.02)***   | 0.22 (0.01)***   | 0.10 (0.02)***   |
| φ3,3  | UA level 4           | 0.67 (0.01)***   | 0.11 (0.02)***   | 0.61 (0.02)***   | 0.13 (0.03)***   | 0.71 (0.02)***   | 0.12 (0.02)***   |
| φ4,1  | PD level 2           | 0.09 (0.01)***   | 0.06 (0.01)***   | 0.11 (0.01)***   | 0.07 (0.01)***   | 0.10 (0.01)***   | 0.07 (0.01)***   |
| φ4,2  | PD level 3           | 0.15 (0.01)***   | 0.09 (0.01)***   | 0.17 (0.01)***   | 0.12 (0.02)***   | 0.15 (0.01)***   | 0.11 (0.02)***   |

* Asterisks denote significance levels: *** p < 0.001, ** p < 0.01, * p < 0.05.
| \( \varphi_{4,3} \) PD level 4 | 0.62 (0.01) | 0.11 (0.02) | 0.60 (0.01) | 0.12 (0.02) | 0.63 (0.02) | 0.10 (0.02) |
| \( \varphi_{5,1} \) AD level 2 | 0.13 (0.01) | 0.08 (0.01) | 0.15 (0.01) | 0.09 (0.01) | 0.13 (0.01) | 0.08 (0.01) |
| \( \varphi_{5,2} \) AD level 3 | 0.22 (0.01) | 0.14 (0.01) | 0.25 (0.01) | 0.16 (0.01) | 0.23 (0.01) | 0.14 (0.01) |
| \( \varphi_{5,3} \) AD level 4 | 0.55 (0.01) | 0.11 (0.02) | 0.55 (0.01) | 0.12 (0.02) | 0.58 (0.02) | 0.11 (0.02) |

| \( r \) discount rate | 0.043 (0.007) | 0.047 (0.008) | 0.019 (0.008) |
| 5-5-5-5-5 (pits) | -1.34 | -1.06 | -0.10 |
| WAIC (value, lower=best) | 16,043 | 15,873 | 15,531 |
| WAIC (rank) | 3 | 2 | 1 |

*Note: mean posterior estimates with standard deviations in parenthesis  ** 90% CI does not contain zero  *** 95% CI does not contain zero  **** estimates on a scale from 0 (EQ-5D level 1) to 1 (EQ-5D level 5)
Figure 3. Comparison between the predicted values for all EQ-5D health states based on the optimal scaling model without interactions (on the horizontal axis) and a) the default main-effects models (without interactions), b) the optimal scaling model with N3/45 term, and c) the optimal scaling model with full set of two-way interactions (on the vertical axis).
Figure 3 (continued)
(Note: nine graphs on 1 page makes them too small)
Online Supplemental

Appendix A. Screenshots of valuation tasks

**Figure A1.** Example cTTO task

[Add screenshot]

*screenshot not yet received*

*Source: Viney et al. (2011)*
**Figure A2.** Example latent scale DCE choice task

Which is better, state A or state B?

- I have some problems with walking about
- I have no problems with self-care
- I have some problems with performing my usual activities
- I have extreme pain or discomfort
- I am moderately anxious or depressed

- I have no problems with walking about
- I have some problems with self-care
- I have no problems with performing my usual activities
- I have extreme pain or discomfort
- I am moderately anxious or depressed

*Source: Nicolet et al. (2018)*
Figure A3: Example DCE with duration ‘matched pairs’ choice task

A. Type-I choice task

Which health state do you prefer, A or B?

| A | B | C |
|---|---|---|
| **1. Mobility** | **10 years in this health state, followed by death** | **no problems in walking about** |
| | | **no problems washing or dressing** |
| | | **moderate problems doing usual activities** |
| | | **severe pain or discomfort** |
| | | **extremely anxious or depressed** |
| | **10 years in this health state, followed by death** | **no problems in walking about** |
| | | **no problems washing or dressing** |
| | | **moderate problems doing usual activities** |
| | | **extreme pain or discomfort** |
| | | **slightly anxious or depressed** |
| | **7 years in this health state, followed by death** | **no problems in walking about** |
| | | **no problems washing or dressing** |
| | | **no problems doing usual activities** |
| | | **no pain or discomfort** |
| | | **not anxious or depressed** |

B. Type-II choice task (in reversed color coding)

Which health state do you prefer, B or C?

| A | B | C |
|---|---|---|
| **1. Mobility** | **10 years in this health state, followed by death** | **no problems in walking about** |
| | | **no problems washing or dressing** |
| | | **moderate problems doing usual activities** |
| | | **severe pain or discomfort** |
| | | **extremely anxious or depressed** |
| | **10 years in this health state, followed by death** | **no problems in walking about** |
| | | **no problems washing or dressing** |
| | | **moderate problems doing usual activities** |
| | | **extreme pain or discomfort** |
| | | **slightly anxious or depressed** |
| | **7 years in this health state, followed by death** | **no problems in walking about** |
| | | **no problems washing or dressing** |
| | | **no problems doing usual activities** |
| | | **no pain or discomfort** |
| | | **not anxious or depressed** |

Source: Jonker et al. (2017)
### Table B1. DCE with duration optimal scaling mixed logit model estimates *

|          | 1. with N5 term | 2. with N45 term |
|----------|-----------------|-----------------|
|          | population mean | population SD   | population mean | population SD   |
| $\beta_1$ perfect health | 1.95 (0.12) *** | 1.69 (0.13) *** | 1.90 (0.11) *** | 1.62 (0.12) *** |
| $\beta_2$ mobility (MO) | -0.64 (0.14) *** | 0.50 (0.04) *** | -0.53 (0.04) *** | 0.51 (0.04) *** |
| $\beta_3$ self-care (SC) | -0.50 (0.12) *** | 0.37 (0.03) *** | -0.40 (0.03) *** | 0.40 (0.03) *** |
| $\beta_4$ usual activities (UA) | -0.70 (0.13) *** | 0.44 (0.03) *** | -0.60 (0.06) *** | 0.45 (0.03) *** |
| $\beta_5$ pain/discomfort (PD) | -1.20 (0.17) *** | 0.70 (0.05) *** | -1.15 (0.06) *** | 0.69 (0.05) *** |
| $\beta_6$ anxiety/depression (AD) | -1.29 (0.16) *** | 0.83 (0.06) *** | -1.17 (0.02) *** | 0.79 (0.05) *** |
| $\beta_7$ MO x SC | n.a. | n.a. | n.a. | n.a. |
| $\beta_8$ MO x UA | n.a. | n.a. | n.a. | n.a. |
| $\beta_9$ MO x PD | n.a. | n.a. | n.a. | n.a. |
| $\beta_{10}$ MO x AD | n.a. | n.a. | n.a. | n.a. |
| $\beta_{11}$ SC x UA | n.a. | n.a. | n.a. | n.a. |
| $\beta_{12}$ SC x PD | n.a. | n.a. | n.a. | n.a. |
| $\beta_{13}$ SC x AD | n.a. | n.a. | n.a. | n.a. |
| $\beta_{14}$ UA x PD | n.a. | n.a. | n.a. | n.a. |
| $\beta_{15}$ UA x AD | n.a. | n.a. | n.a. | n.a. |
| $\beta_{16}$ PD x AD | n.a. | n.a. | n.a. | n.a. |
| $\beta_{17}$ N45 | 0.02 (0.11) | 0.17 (0.02) *** | -0.17 (0.02) *** | 0.16 (0.01) *** |
| $\varphi_{1,1}$ MO level 2 | 0.16 (0.05) *** | 0.14 (0.09) ** | 0.18 (0.01) *** | 0.10 (0.02) *** |
| $\varphi_{1,2}$ MO level 3 | 0.25 (0.07) *** | 0.16 (0.08) *** | 0.27 (0.01) *** | 0.14 (0.02) *** |
| $\varphi_{1,3}$ MO level 4 | 0.71 (0.13) *** | 0.09 (0.02) *** | 0.64 (0.02) *** | 0.10 (0.02) *** |
| $\varphi_{2,1}$ SC level 2 | 0.17 (0.06) *** | 0.12 (0.10) *** | 0.20 (0.02) *** | 0.09 (0.04) *** |
| $\varphi_{2,2}$ SC level 3 | 0.27 (0.07) *** | 0.15 (0.09) *** | 0.31 (0.02) *** | 0.13 (0.04) *** |
| $\varphi_{2,3}$ SC level 4 | 0.73 (0.17) *** | 0.06 (0.02) *** | 0.59 (0.02) *** | 0.11 (0.03) *** |
| $\varphi_{3,1}$ UA level 2 | 0.14 (0.04) *** | 0.08 (0.03) *** | 0.14 (0.01) *** | 0.08 (0.02) *** |
| $\varphi_{3,2}$ UA level 3 | 0.22 (0.04) *** | 0.10 (0.01) *** | 0.25 (0.02) *** | 0.10 (0.02) *** |
| $\varphi_{3,3}$ UA level 4 | 0.72 (0.04) *** | 0.08 (0.02) *** | 0.61 (0.02) *** | 0.13 (0.03) *** |
| $\varphi_{4,1}$ PD level 2 | 0.10 (0.14) *** | 0.07 (0.03) *** | 0.11 (0.01) *** | 0.07 (0.01) *** |
| $\varphi_{4,2}$ PD level 3 | 0.15 (0.02) *** | 0.11 (0.06) *** | 0.17 (0.01) *** | 0.12 (0.02) *** |
| $\varphi_{4,3}$ PD level 4 | 0.65 (0.02) *** | 0.12 (0.06) *** | 0.60 (0.01) *** | 0.12 (0.02) *** |
| $\varphi_{5,1}$ AD level 2 | 0.13 (0.02) *** | 0.08 (0.02) *** | 0.15 (0.01) *** | 0.09 (0.01) *** |
| $\varphi_{5,2}$ AD level 3 | 0.22 (0.02) *** | 0.14 (0.02) *** | 0.25 (0.01) *** | 0.16 (0.01) *** |
| $\varphi_{5,3}$ AD level 4 | 0.38 (0.08) *** | 0.11 (0.04) *** | 0.55 (0.01) *** | 0.12 (0.02) *** |
| $r$ discount rate | 0.056 (0.019) *** | 0.047 (0.008) *** |
| 5-5-5-5-5 (pits) | -1.23 | -1.06 |
| WAIC (value, lower=better) | 16,275 | 15,873 |

* Note: mean posterior estimates with standard deviations in parenthesis ** 90% CI does not contain zero *** 95% CI does not contain zero **** estimates on a scale from 0 (EQ-5D level 1) to 1 (EQ-5D level 5)
Table C1. Latent scale DCE optimal scaling mixed logit model estimates *

|                  | 1. with single set of optimal scales | 2. with separate optimal scales |
|------------------|-------------------------------------|---------------------------------|
|                  | population mean | population SD | population mean | population SD |
| $\beta_1$ mobility (MO) | -3.32 (0.10) *** 2.06 (0.11) *** | -3.36 (0.11) *** 2.02 (0.09) *** |
| $\beta_2$ self-care (SC) | -2.21 (0.09) *** 1.26 (0.13) *** | -2.27 (0.08) *** 1.32 (0.11) *** |
| $\beta_3$ usual activities (UA) | -2.17 (0.10) *** 1.50 (0.13) *** | -2.19 (0.11) *** 1.45 (0.12) *** |
| $\beta_4$ pain/discomfort (PD) | -3.05 (0.10) *** 2.36 (0.12) *** | -3.09 (0.08) *** 2.20 (0.08) *** |
| $\beta_5$ anxiety/depression (AD) | -2.63 (0.10) *** 2.25 (0.10) *** | -2.57 (0.09) *** 2.06 (0.08) *** |
| $\beta_6$ MO x SC | 0.46 (0.07) *** 0.50 (0.08) *** | 0.49 (0.06) *** 0.41 (0.07) *** |
| $\beta_7$ MO x UA | 0.32 (0.06) *** 0.50 (0.08) *** | 0.28 (0.09) *** 0.45 (0.07) *** |
| $\beta_8$ MO x PD | 0.28 (0.07) *** 0.43 (0.08) *** | 0.29 (0.05) *** 0.48 (0.07) *** |
| $\beta_9$ MO x AD | 0.27 (0.07) *** 0.54 (0.08) *** | 0.17 (0.08) *** 0.47 (0.08) *** |
| $\beta_{10}$ SC x UA | 0.19 (0.06) *** 0.41 (0.07) *** | 0.18 (0.07) *** 0.41 (0.08) *** |
| $\beta_{11}$ SC x PD | 0.18 (0.07) *** 0.49 (0.08) *** | 0.14 (0.07) *** 0.45 (0.07) *** |
| $\beta_{12}$ SC x AD | 0.20 (0.06) *** 0.43 (0.07) *** | 0.14 (0.07) *** 0.39 (0.06) *** |
| $\beta_{13}$ UA x PD | 0.29 (0.07) *** 0.40 (0.08) *** | 0.32 (0.07) *** 0.40 (0.06) *** |
| $\beta_{14}$ UA x AD | 0.08 (0.07) *** 0.53 (0.09) *** | 0.04 (0.05) *** 0.49 (0.08) *** |
| $\beta_{15}$ PD x AD | 0.31 (0.07) *** 0.64 (0.08) *** | 0.27 (0.05) *** 0.52 (0.07) *** |
| $\varphi_1$ MO level 2 **** | 0.20 (0.01) *** 0.20 (0.01) *** | 0.25 (0.01) *** 0.16 (0.01) *** |
| $\varphi_2$ SC level 2 **** | 0.37 (0.01) *** 0.13 (0.02) *** | 0.39 (0.01) *** 0.10 (0.01) *** |
| $\varphi_3$ UA level 2 **** | 0.25 (0.01) *** 0.15 (0.02) *** | 0.25 (0.01) *** 0.13 (0.02) *** |
| $\varphi_4$ PD level 2 **** | 0.24 (0.01) *** 0.12 (0.01) *** | 0.27 (0.01) *** 0.09 (0.01) *** |
| $\varphi_5$ AD level 2 **** | 0.31 (0.01) *** 0.09 (0.01) *** | 0.34 (0.01) *** 0.08 (0.01) *** |
| $\Psi_1$ MO level 2 **** | n.a. | n.a. | 0.49 (0.10) *** 0.40 (0.02) *** |
| $\Psi_2$ SC level 2 **** | n.a. | n.a. | 0.42 (0.05) *** 0.24 (0.04) *** |
| $\Psi_3$ UA level 2 **** | n.a. | n.a. | 0.26 (0.05) *** 0.27 (0.04) *** |
| $\Psi_4$ PD level 2 **** | n.a. | n.a. | 0.53 (0.07) *** 0.34 (0.01) *** |
| $\Psi_5$ AD level 2 **** | n.a. | n.a. | 0.71 (0.11) *** 0.19 (0.05) *** |
| ($\Psi_1 - \varphi_2$) | n.a. | n.a. | 0.25 (0.09) *** n.a. |
| ($\Psi_2 - \varphi_3$) | n.a. | n.a. | 0.04 (0.04) n.a. |
| ($\Psi_3 - \varphi_3$) | n.a. | n.a. | 0.01 (0.05) n.a. |
| ($\Psi_4 - \varphi_4$) | n.a. | n.a. | 0.26 (0.06) *** n.a. |
| ($\Psi_5 - \varphi_5$) | n.a. | n.a. | 0.37 (0.10) *** n.a. |

WAIC (value, lower=better) 55,001 54,664

* Note: mean posterior estimates with standard deviations in parenthesis ** 90% CI does not contain zero *** 95% CI does not contain zero **** estimates on a scale from 0 (EQ-5D level 1) to 1 (EQ-5D level 3).
Appendix D. **BUGS** model codes

1. Heteroskedastic optimal scaling Tobit model with two-way interactions

```r
model {
  # N = number of respondents
  # T = number of cTTO evaluations per respondent

  for (i in 1:N){
    for (t in 1:T){

      # normal likelihood with censoring at -1
      Y[i,t] ~ dnorm(mu_y[i,t], prec_y[i,t])C(-1,)

      # standard deviation of normal distribution
      prec_y[i,t] <- 1/(sd_y[i,t]^2)
      sd_y[i,t] <- exp( gamma[1] +
        gamma[2] * value[i,t] +
        gamma[3] * value[i,t] * value[i,t] +
        gamma[4] * value[i,t] * value[i,t] * value[i,t] +
        gamma[5] * value[i,t] * value[i,t] * value[i,t] * value[i,t] )

      # mean of normal distribution
      mu_y[i,t] <- beta[16] + beta[1] * optimal_scale[i,t,1] +
        beta[2] * optimal_scale[i,t,2] +
        beta[3] * optimal_scale[i,t,3] +
        beta[4] * optimal_scale[i,t,4] +
        beta[5] * optimal_scale[i,t,5] +
        beta[6] * optimal_scale[i,t,1] * optimal_scale[i,t,2] +
        beta[7] * optimal_scale[i,t,1] * optimal_scale[i,t,3] +
        beta[8] * optimal_scale[i,t,1] * optimal_scale[i,t,4] +
        beta[9] * optimal_scale[i,t,1] * optimal_scale[i,t,5] +
        beta[10] * optimal_scale[i,t,2] * optimal_scale[i,t,3] +
        beta[11] * optimal_scale[i,t,2] * optimal_scale[i,t,4] +
        beta[12] * optimal_scale[i,t,2] * optimal_scale[i,t,5] +
        beta[13] * optimal_scale[i,t,3] * optimal_scale[i,t,4] +
        beta[14] * optimal_scale[i,t,3] * optimal_scale[i,t,5] +
        beta[15] * optimal_scale[i,t,4] * optimal_scale[i,t,5]

      # model for optimal scales
      for (dim in 1:5){
        # EQ-5D-3L level 1 => always 0.0
        # EQ-5D-3L level 2 => phi[d] * X
        # EQ-5D-3L level 3 => always 1.0 * X
        optimal_scale[i,t,dim] <- 0.0 + phi[dim]*X[i,t,dim,1] + X[i,t,dim,2]
      }
    }
  }

  # prior on betas
  beta[1:16] ~ dmnorm(b_zeros[], b_prec[,])
  for (b in 1:16){
    b_zeros[b] <- 0
    for (bb in 1:16){
      b_prec[b,bb] <- equals(b,bb)/100
    }
  }

  # prior on gammas
  gamma[1:5] ~ dmnorm(g_zeros[], g_prec[,])
  for (g in 1:5){
    g_zeros[g] <- 0
    for (gg in 1:5){
      g_prec[g,gg] <- equals(g,gg)/100
    }
  }

  # prior on phi_stars
  phi_star[1:5] ~ dmnorm( t_zeros[] , t_prec[,) }
  for (t in 1:5){
    t_zeros[t] <- 0
  }
}
```
for (tt in 1:5)
  t_prec[t,tt] <- equals(t,tt)/100
}

# implied prior on phis
for (t in 1:5)
  logit(phi[t]) <- phi_star[t]

# monitor individual-level likelihood and log-likelihoods (for WAIC computations)
for (i in 1:N)
  LL_resp[i] <- normalLogLikelihood(Y[i,], mu_y[i,], prec_y[i,])
  L_resp[i] <- exp(LL_resp[i])

2. Latent scale MIXL optimal scaling model with two-way interactions

model {
  # N = number of respondents
  # T = number of choice tasks per respondent
  # A = number of alternatives per choice task
  # V = number of betas (main-effects + two-way interactions)
  # D = number of attributes / dimensions

  for (n in 1:N)
    for (t in 1:T)
      # X-data are exported with chosen alternative in first position
      Y[n,t] <- 1

      # categorical likelihood
      Y[n,t] ~ dcat(prob[n, t, 1:A])
      prob[n,t,1:A] <- softmaxInteractions( optimal_scale[n,t,1,1:D],
                                            optimal_scale[n,t,2,1:D],
                                            beta[n,1:V] )

      # construction of optimal scales
      for (a in 1:A)
        for (dim in 1:D)
          # EQ-5D level 1 => always 0.0
          # EQ-5D level 2 => phi[d] * X
          # EQ-5D level 3 => always 1.0 * X
          optimal_scale[n,t,a,dim] <- 0.0 + phi[n,dim]*X[n,t,a,dim,1] + X[n,t,a,dim,2]

  # prior on betas
  for (n in 1:N)
    for (v in 1:V)
      mu_beta[n,1:V] ~ dmnorm(mu_beta[,], tau_beta[,])
    mu_beta[1:V] ~ dmnorm(hyper_mu_beta[,], hyper_tau_beta[,])
    tau_beta[1:V,1:V] ~ dwish(identity[,,] , V)

  for (v in 1:V)
    hyper_mu_beta[v] ~ 0
  for (vv in 1:V)
    identity[v,vv] <- equals(v,vv)
    hyper_tau_beta[v,vv] <- equals(v,vv)/100

  # prior on phi_stars
  for (n in 1:N)
    mu_phi_star[n,1:D] ~ dmnorm( mu_phi_star[,], tau_phi_star[,])
    mu_phi_star[1:D] ~ dmnorm(hyper_mu_phi_star[,], hyper_tau_phi_star[,])
    tau_phi_star[1:D,1:D] ~ dwish(identityMatrix[,,] , D)

  for (d in 1:D)
    hyper_mu_phi_star[d] <- 0
  for (dd in 1:D)
    identityMatrix[d,dd] <- equals(d,dd)
# implied prior on phis
for (n in 1:N){
  for (d in 1:D){
    logit(phi[n,d]) <- phi_star[n,d]
  }
}

# monitor sd of betas
covar_beta[1:V,1:V] <- inverse(tau_beta[,])
for (v in 1:V){ sd_beta[v] <- sqrt(covar_beta[v,v]) }

# monitor mean and sd of phis
for (d in 1:D){
  mu_phi[d] <- mean(phi[,d])
  sd_phi[d] <- sd(phi[,d])
}

# monitor individual-level likelihood and log-likelihoods (for WAIC computations)
for (n in 1:N){
  for (t in 1:T){ LL_task[n,t] <- log(prob[n,t,1]) }
  LL_resp[n] <- sum(LL_task[n,])
  L_resp[n] <- exp(LL_resp[n])
}

3. Stand-alone MIXL optimal scaling model with non-linear duration and two-way interactions

model {

  # N = number of respondents
  # T = number of choice tasks per respondent
  # A = number of alternatives per choice task (fixed at 2)
  # V = number of explanatory variables (including non-linear time preference)
  for (n in 1:N){
    for (t in 1:T){
      # X-data are exported with chosen alternative in first position
      Y[n,t] <- 1

      # categorical likelihood
      Y[n,t] ~ dcat(prob[n,t,1:A])
      prob[n,t,1:2] <- softmaxExp3Interactions( optim_scale[n,t,1,], Q[n,t,1],
                                                optim_scale[n,t,2,], Q[n,t,2],
                                                beta[n,],
                                                rate )

      # construction of optimal scales
      for (alt in 1:2){
        optim_scale[n,t,alt,1] <- inprod( phi_att1[n,1:4] , X_attr[n,t,alt,1:4] )
        optim_scale[n,t,alt,2] <- inprod( phi_att2[n,1:4] , X_attr[n,t,alt,2:1:4] )
        optim_scale[n,t,alt,3] <- inprod( phi_att3[n,1:4] , X_attr[n,t,alt,3:1:4] )
        optim_scale[n,t,alt,4] <- inprod( phi_att4[n,1:4] , X_attr[n,t,alt,4:1:4] )
        optim_scale[n,t,alt,5] <- inprod( phi_att5[n,1:4] , X_attr[n,t,alt,5:1:4] )
      }
    }
  }

  # prior on betas
  for (n in 1:N){ beta[n,1:16] ~ dmnorm(mu_beta[], prec_beta[],) } 
  mu_beta[1:16] ~ dmnorm(zero[], small_prec[],)
  prec_beta[1:16,1:16] ~ dwish(identity[],16)
for (v in 1:16){
  zero[v] <- 0
  for (vv in 1:16){
    identity[v,vv] <- equals(v,vv)
    small_prec[v,vv] <- equals(v,vv)/100
  }
}

# prior on discount rate
rate ~ dunif(0,1)

# prior on phi_stars
for (n in 1:N){ phi_star[n,1:15] ~ dmnorm( mu_phi_star[,] , prec_phi_star[,] ) } 
mu_phi_star[1:15] ~ dmnorm( zeros[,] , low_prec[,] )
pred_phi_star[1:15,1:15] ~ dwish( identityMatrix[,] , 15)

for (v in 1:15){
  zeros[v] <- 0
  for (vv in 1:15){
    identityMatrix[v,vv] <- equals(v,vv)
    low_prec[v,vv] <- equals(v,vv)/100
  }
}

# implied prior on phis
for (n in 1:N){
  phi_att1[n,1:4] <- orderedLogisticNormal( phi_star[n,1:3] ) 
  phi_att2[n,1:4] <- orderedLogisticNormal( phi_star[n,4:6] ) 
  phi_att3[n,1:4] <- orderedLogisticNormal( phi_star[n,7:9] ) 
  phi_att4[n,1:4] <- orderedLogisticNormal( phi_star[n,10:12] ) 
  phi_att5[n,1:4] <- orderedLogisticNormal( phi_star[n,13:15] )
}

# monitor mean and sd of phis
for (level in 1:3) {
  mu_phi[1,level] <- mean( phi_att1[,level] )
  mu_phi[2,level] <- mean( phi_att2[,level] )
  mu_phi[3,level] <- mean( phi_att3[,level] )
  mu_phi[4,level] <- mean( phi_att4[,level] )
  mu_phi[5,level] <- mean( phi_att5[,level] )

  sd_phi[1,level] <- sd( phi_att1[,level] )
  sd_phi[2,level] <- sd( phi_att2[,level] )
  sd_phi[3,level] <- sd( phi_att3[,level] )
  sd_phi[4,level] <- sd( phi_att4[,level] )
  sd_phi[5,level] <- sd( phi_att5[,level] )
}

# monitor sd of betas
covar_beta[1:16,1:16] <- inverse(prec_beta[,] )
for (a in 1:16) { sd_beta[a] <- sqrt(covar_beta[a,a] ) }

# monitor individual-level likelihood and log-likelihoods (for WAIC computations)
for (n in 1:N){
  for (t in 1:T) { LL_task[n,t] <- log( prob[n,t, Y[n,t] ] ) }
  LL_resp[n] <- sum(LL_task[n,])
  LResp[n] <- exp(LL_resp[n])
}

# monitor decrements on QALY scale
beta_QALY[1] <- 1
for (v in 2:16) { beta_QALY[v] <- mu_beta[v] / mu_beta[1] }
pits <- sum(beta_QALY[])
A. Standard (dummy-coded) health-state decrements

| Level | Value | \( \beta_1 = -0.10 \) | \( \beta_2 = -0.25 \) | \( \beta_3 = -0.50 \) |
|-------|-------|---------------------|---------------------|---------------------|
| Level 1 | 0 (base-case) | | | |
| Level 2 | | | | |
| Level 3 | | | | |
| Level 4 | | | | |

B. Optimal scaling approach (worst level constrained to 1)

| Level | Value | \( \phi_1 = 0.20 \) | \( \phi_2 = 0.50 \) | 1.00 (worst level) |
|-------|-------|---------------------|---------------------|---------------------|
| Level 1 | 0 (base-case) | | | |
| Level 2 | | | | |
| Level 3 | | | | |
| Level 4 | | | | |

\( \beta_1 = -0.50 \) (health domain importance)
Level 1 (best level)  

Level 2  

Level 3 (worst level)  

EQ-5D domain importances: $\beta_{5,5}$
A. Standard (dummy-coded) health-state decrements

Level 1 0 (base-case)

Level 2 \( \beta_1 = -0.10 \)

Level 3 \( \beta_2 = -0.25 \)

Level 4 \( \beta_3 = -0.50 \)

B. Optimal scaling approach (worst level constrained to 1)

Level 1 0 (base-case)

Level 2 \( \varphi_1 = 0.20 \)

Level 3 \( \varphi_2 = 0.50 \)

\( \beta_1 = -0.50 \) (health domain importance)

1.00 (worst level)
EQ-5D domain importances:

Level 1 (best level):
- Mobility: $\phi_1$
- Selfcare: $\phi_2$
- Usual activities: $\phi_3$
- Pain/discomfort: $\phi_4$
- Anxiety/depression: $\phi_5$

EQ-5D domain importances: $\beta_{1-5}$