

Review

Quality review of a proposed EQ-5D-5L value set for England (Hernández-Alava et al.)

- Review critically assessed the data and statistical methods that were used to generate the EQ-5D-5L value set for England (Devlin et al. 2017)
- Below there is an overview of comments on the review by Hernández-Alava et al.

Main comments:

1. EuroQol-Valuation Technique (EQ-VT) protocol

- Changes of protocol: The English EQ-5D-5L valuation study applied the first version of the EQ-VT. Hernández-Alava et al. correctly state that the “international experience underlines the problems with EQ-VT 1.0 and suggests that major improvements are achievable.” (p. 17). Moreover, Hernández-Alava et al. indicate that a difficulty of the EQ-VT protocol lies in the fact that the protocol has changed over time and they define those changed as “[...] superseded successively [...]” (p. 17). This interpretation must be corrected as the valuation tasks have remained the same across the different version of the protocol. However, later protocol versions pay more attention to the implementation of the valuation tasks with the introduction of a quality control procedure (since EQ-VT 1.1) and a feedback module allowing participants to review their time trade-off (TTO) responses (in EQ-VT 2.0) (Ludwig et al. 2018).
- Experimental design: The coverage of possible health states is criticized by Hernández-Alava et al. However, the reasons behind the selected approach by Oppe & van Hout (2017) should be clarified: a certain number of observations per health state is needed to estimate a robust model. Therefore, there is a trade-off between coverage and statistical considerations. The proposed coverage by Hernández-Alava et al. based on severity level is not an appropriate alternative as it has other important limitations: severity is not a good proxy for health state selection as quite different health states can have the same severity level (e.g. 22222 versus 11152).

2. Data quality of EQ-5D-5L valuation study in England

- TTO data: Hernández-Alava et al. analyzed the data quality of the TTO data in detail. Table 2.5 summarizes the proportions of individual participants displaying potentially problematic response behavior. Five of the mentioned criteria are well-chosen and comprehensible (i.e. 1, 3, 5, 10a, 10b). However, the other criteria might also be the (partly) respondent's preferences: According to the preference of respondents an individual threshold can be achieved if a certain dimensional level is reached and an additional increase in severity will result in the same value (ties) (2). There might be people who cannot distinct between more than five values but this can also be low task engagement by the respondent and the interviewer (4, 6). Criteria 7, 8 and 9 can be simply the preference of respondent and therefore conclusions must be made with carefulness. Even though the degree of inconsistencies on a TTO response level (table 2.8) relativizes the extent of inconsistencies in the TTO data, the level of inconsistencies and the clustering of values in the data are concerning. Post hoc-analysis of the TTO data of two other EQ-5D-5L valuation studies in Spain and the Netherlands using EQ-VT 1.0 have shown that a high proportions of inconsistencies and clustering per respondent can be related to the interviewer performance (issues of protocol non-compliance and low engagement with the task) (Ramos-Goni et al. 2016).
- Discrete Choice (DC) data: Hernández-Alava et al. correctly indicate that no assessment of the inconsistency level of the DC data are possible due to the experimental design (i.e. no dominant pairs included). Possible data quality checks could include response pattern across all DC tasks (e.g. AAAAAAA,BBBBBBB, ABABABA, BABABAB) and analyzing if the proportion of the choice of A or B was correlated to the difference in the severity level between the health states.

3. Specification and estimation of the EQ-5D-5L value set for England

- Hernández-Alava et al. judge the modeling choices to be non-transparent. However, the authors of the English EQ-5D-5L value set published an extra paper on the modeling where the main choices are described (Feng et al. 2016). The modeling approaches to the EQ-5D-5L TTO data are innovative in that they account for truncating, censoring, heteroskedasticity, and preference heterogeneity. The modeling advances were driven by considerations obtained from carefully investigating the

TTO task and the resulting data, and by matching these to the assumptions underlying the regression models.

- Hernández-Alava et al. criticize that too many TTO data with quality issues are included unmodified in the modeling. However, the level of modification of preference values should be carefully traded-off as the resulting value set should still represent the opinion of the general population.
- Even though modeling choices always have a “normative nature”, in the following there are comments of three modeling choices made by the English team: (a) Even though the modeling choice for truncating at 1 is well described by the authors of the English value set, the criticism by Hernández-Alava et al. is justified as the maximum TTO value is theoretically bounded at 1 and a correction of the error term at 1 is by definition not inevitably needed. (b) In contrast to the opinion of Hernández-Alava et al., the correction for existing heteroskedasticity of the TTO data is strongly needed to be prevent biased model parameters. (c) The forced consistency criticized Hernández-Alava et al. is a comprehensible point of criticism. Other national EQ-5D-5L value sets, like the Netherlands, included collapsed level in the final value set as this was the preference of the respondents (Versteegh et al. 2016).
- Hernández-Alava et al. criticize combining TTO and DC data in hybrid model. In accordance to the other mentioned modeling choices, there are advantages and disadvantages for combining the (complementary) information in one value set. However, the authors of the English value set well-justified the inclusion of both data types in a joint model.

Recommendation of reviewer:

- Introduction of EQ-5D-5L for economic evaluations that inform UK health policy (reason: improved descriptive system: increased sensitivity, increased discriminability to detect differences on an individual level as well as between groups [Janssen et al. 2018]) .
- Hernández-Alava et al. come to the conclusion that a new data collection is recommended. With regard to the high proportions of inconsistencies per respondent and the strong clustering of values compared to previous EQ-5D-5L valuation studies, the recommendation is comprehensible. As there is strong evidence that the refined valuation protocol with its quality control process appears to be a solid basis for

estimating national EQ-5D-5L value sets (e.g. Ludwig et al. 2018; Hobbins et al. 2018), the new EQ-5D-5L valuation study in England should apply the improved EuroQol valuation protocol 2.0 to benefit from the evidence-based EQ-VT and to promote international comparability. Thereby, the modeling choices of the authors of the English EQ-5D-5L value set should be not simply replicated but the applicability to the new data should be scrutinized.

- In general, the value set should be updated on a regular basis to account for changes of preferences (response shift).

References:

Devlin NJ, Shah KK, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: an EQ-5D-5L value set for England. *Health Econ.* 2017. <https://doi.org/10.1002/hec.3564>.

Feng Y, Devlin NJ, Shah KK, Mulhern B, van Hout B. New methods for modelling EQ-5D-5L value sets: an application to English data. *Health Econ.* 2017. <https://doi.org/10.1002/hec.3560>.

Janssen MF, Bonsel GJ, Luo N. Is EQ-5D-5L better than EQ-5D-3L? A head-to-head comparison of descriptive systems and value sets from seven countries. *PharmacoEconomics* 2018;36:675-97

Hobbins A, Barry L, Kelleher D, Shah K, Devlin N, Ramos-Goni JM, O'Neill C. Utility Values for Health States in Ireland: A Value Set for the EQ-5D-5L. *PharmacoEconomics*, doi: 0.1007/s40273-018-0690-x.

Ludwig, K., Schulenburg, J-M. Graf v. d., Greiner, W. (2018): German Value Set for the EQ-5D-5L. *PharmacoEconomics*, 36(6): 663-674.

Oppe M, van Hout B. The ‘‘power’’ of eliciting EQ-5D-5L values: the experimental design of the EQ-VT. EuroQol working paper series. 2017;17003. <http://euroqol.org/wp-content/uploads/2016/10/EuroQol-Working-Paper-Series-Manuscript-17003-Mark-Oppe.pdf>.

Ramos-Goni JM, Oppe M, Slaap B, Busschbach JJV, Stolk E. Quality control process for EQ-5D-5L valuation studies. *Value Health.* 2016;20:466–73.

Versteegh MM, Vermeulen KM, Evers SMAA, de Wit GA, Prenger R, Stolk EA. Dutch tariff for the five-level version of EQ-5D. *Value Health.* 2016;19:343–52.