Master's Degree Dissertation

# Valuing health related quality of life in adolescents from their own perspective

## A potential EQ-5D-Y-3L Value Set for Spain


**Anabel Estévez-Carrillo**

Master's Degree in Master in Health Economics and Pharmacoeconomics

UPF Barcelona School of Management


**Academic Year 2019 – 2021**

**Juan Manuel Ramos-Goñi**

BARCELONA SCHOOL OF MANAGEMENT

Project performed within the framework of the Master in Health Economics and Pharmacoeconomics program taught by Barcelona School of Management, a centre

# Resumen


Introducción: Existe un interés creciente en medir las preferencias de los jóvenes para la toma de decisiones. El EQ-5D-Y-3L es un instrumento genérico desarrollado por Grupo EuroQol basado en las preferencias de niños y adolescentes. Recientemente se ha demostrado la necesidad de conjuntos de valores específicos para el EQ-5D-Y-3L. Nuestro estudio analiza la validez de las respuestas DCE de los adolescentes y explora una metodología potencial para obtener un conjunto de valores para los adolescentes en España desde su propia perspectiva.

Métodos: Realizamos un DCE con control de calidad en una muestra de adolescentes. Como la heterogeneidad de preferencias está presente en los datos de DCE, probamos modelos de clases latentes y logit mixtos de coeficientes aleatorios. Específicamente, para los modelos de clases latentes seguimos un enfoque iterativo, aumentando el número de clases hasta que no se obtuvieron mejoras en las predicciones del modelo y la bondad de ajuste (BIC). Una vez que se seleccionó el mejor modelo, anclamos las utilidades de la escala latente del DCE en la escala de AVAC utilizando el enfoque de “anclaje en el peor estado”.

Resultados: El 38% de los participantes que completaron el DCE no cumplieron con las reglas de control de calidad. Todos los modelos reportaron resultados consistentes, sin embargo, el modelo de clases latentes con 4 clases tuvo el mejor desempeño.

Discusión/conclusión: La validez de las respuestas DCE en adolescentes es menor que en adultos. Sin embargo, la muestra después del procedimiento de control de calidad es válida y es factible obtener utilidades de escala latente DCE. El enfoque de “anclaje en el peor estado” aboga por la igualdad en la atención médica entre adolescentes y adultos. Este enfoque permite la estimación de un conjunto de valores que se puede utilizar en la toma de decisiones de asignación de recursos de salud que es directamente relevante para los adolescentes.


Palabras clave: Adolescentes; EQ-5D-Y-3L; DCE

# Abstract

Background: There is a growing interest in measuring youth preferences for decision making. The EQ-5D-Y-3L is generic preference-based instrument developed by the EuroQol Group that is appropriate for children and adolescents. Recent evidence demonstrates the need of specific value sets for the EQ-5D-Y-3L. Our study analyses the validity of DCE responses in adolescents and explores a potential methodology to obtain an adolescents' value set in Spain from their own perspective.

Methods: We conducted an EQ-5D-Y-3L DCE, including quality control, in an adolescent sample. As preferences heterogeneity is known to be present in DCE data, we tested random coefficient mixed logit and latent class models. Specifically, for latent class models we followed an iterative approach, increasing the number of classes until no further improvement in model predictions and goodness of fit (BIC) was obtained. Once the best model was selected, we anchored the DCE latent scale utilities onto QALYs scale using the "anchoring on the pits state" approach.

Results: 38% participants who completed the DCE failed the quality control rules. All models reported consistent results, however, the latent class model with 4 classes had the best performance.

Discussion/conclusion: The validity of DCE responses in adolescents is lower than in adults. However, the sample after the quality control procedure is valid and it is feasible to obtain a DCE latent scale utilities. Anchoring on the adults' pits state advocate for health care equality between adolescent and adults. This approach allows the estimation of a value set that can be used to inform decision-making in the health resource allocation that is directly relevant to adolescents.


Keywords: Adolescents; EQ-5D-Y-3L; DCE

# Table of contents

## Introduction

Nowadays, large amounts of health expenditure resources are spent on making illness more bearable. However, resources available to satisfy all the demands in health sector are often limited. Economic evaluations are needed to prioritize between the different alternatives, considering the costs and the clinical benefits provided by each of them. When an economic evaluation measures the clinical benefit using a generic measure of health gain, it is usually referred as cost-utility analysis (CUA). The term "utility" refers to the preferences individuals or society may have for any set of health states. Generic preference-based instruments are needed to estimate utilities. The development of this type of instruments has been fast in the last decades and nowadays there are many available, such us EQ-5D(1), Short Form 6D(2), 15D instrument(3) and Health Utility Index(4).

The EQ-5D is a standardized generic instrument of health related quality of life, developed by the EuroQol group, that is widely used in population health surveys, clinical and economic studies(5). Some health technology evaluation organizations, including NICE, specifically request that the estimates of health-related utilities used in quality-adjusted life-years (QALYs) calculation have been obtained by using the EQ-5D questionary(6). The EQ-5D-5L consists in a descriptive system with 5 domains (mobility, self-care, usual activities, pain/discomfort, and anxiety/depression) and a visual analogue scale. Each dimension presents either 3 levels of problems (EQ-5D-3L) or 5 levels of problems (EQ-5D-5L). In addition to the descriptive system, to convert each of the health states into a single numerical index, it is essential to develop a scoring algorithm or value set, i.e., a set of weights to each of the levels in each EQ-5D dimension. EQ-5D national value sets based on general population preferences are available for many countries, including Spain, where the EQ-5D-3L(7) and EQ-5D-5L(8) value sets are available.

While instruments such as the EQ-5D-3L and EQ-5D-5L are widely used to measure HRQoL in adults, they should not be automatically used in children and adolescent because the issues and concepts may differ(9). To cover the growing interest in measuring youth preferences in decision making, the EuroQol group adapted the original EQ-5D-3L to make it appropriate for children ensuring that the language and concepts embodied in the instrument were correctly understood(9). This new instrument was called EQ-5D-Y-3L.

Given the similarities between the EQ-5D-3L and the EQ-5D-Y-3L, some authors have wrongly used the adults EQ-5D-3L value sets to perform cost-utilities analysis among paediatric populations(10). Recent evidence demonstrates that different value sets are necessary, since differences in wording/perspective of the instrument results in different values for the EQ-5D-3L than those for the corresponding EQ-5D-Y-3L and greater differences were found among health states with higher severity(11).

However, issues have been raised in the process of developing value sets for children(11). One of the points under discussion is in the EQ-5D-Y-3L valuation is who should value the health states of children and adolescents. There is a normative argument that the general population should be the most relevant because CUA to inform decision making should follow a societal perspective. However, if the general population value the EQ-5D-Y-3L, it is adults assessing the severity of health problems of younger age groups, and there is evidence that adult's and adolescent's preferences about identical health states differ(12). Some institutions regard adolescents health preferences from own perspective for decision-making that is directly related to them(13)(14).

Recently, the first version of the EQ-5D-Y-3L valuation protocol was published, which suggests obtaining preferences from a sample of the general adult population to keep the taxpayer perspective(15). The experimental design described in the protocol combines two different valuation techniques: Discrete Choice experiment (DCE) and Composite Time Trade Off (C-TTO). Following the protocol, Ramos-Goñi et al. (16) recently proposed modelling strategies for the EQ-5D-Y-3L valuation and presented preliminary results of the Spanish EQ-5D-Y-3L value set from adults' perspective.

Despite the EQ-5D-Y-3L valuation protocol, some countries or health technology assessments may consider adolescents' health preferences from own perspective more appropriate. The experimental design in adolescents is limited to DCE, since ethical issues may arise including C-TTO tasks that involve the state of being dead. Therefore, two questions may arise when analysing adolescents' preferences: First, whether adolescents are capable to provide valid responses in a discrete choice experiment. And secondly, if that was the case, whether it is feasible to anchor DCE latent scale utilities onto QALYs scale, enabling the creation of an adolescents' value set from own perspective.

The objective of our study was to explore the validity of DCE responses on 11-17 years old adolescents. In addition, this research explores a potential methodology to obtain a young population value set in Spain from own perspective.

## Methods

### *The EQ-5D-Y valuation protocol*

The EQ-5D-Y-3L valuation protocol(15) recommends the use of discrete choice experiments to estimate the relative importance of the EQ-5D-Y-3L dimensions. The DCE task considered in this study considers pairs of health states described by the five attributes of the EQ-5D-3L-Y (Figure 1). The DCE design recommended in the EQ-5D-Y-3L valuation protocol was estimated using a D-efficiency maximization procedure and uses 150 DCE pairs blocked into 10 balanced blocks. The design used a 2-dimension overlap for all pairs, in other words, the two health states included in each pair have the same level in 2 dimensions and different levels in the other 3 dimensions.

To anchor the latent scale value set obtained from DCE data onto the appropriate scale (1=Full health, 0=Dead), the EQ-5D-Y valuation protocol recommends using C-TTO data. In a C-TTO task, participants must indicate the length of remaining life expectancy that they are willing to trade-off to avoid remaining in a health state. C-TTO acceptability and appropriateness in adolescent is questionable for two main reasons. First, whether adolescents can make reasonable choices in tasks involved the state of being dead; and second, whether it would cause distress for adolescents concerns ethics committees(17). Therefore, we only collected DCE data.

### *DCE Interview*

The DCE online survey included the following sections:

1) Information sheet and consent for participation. The aim of the project was explained. As the target population was 11-17 years old adolescents, this information was referred to a participants' parent or legal guardian, who must provide consent for participation. Otherwise, the survey only showed a message thanking for considering participating.
2) Demographic questions including age range and gender. This allowed us to delimit the quotas defined to ensure the sample representativeness.
3) The EQ-5D-Y-3L instrument. Participants were asked to self-complete the EQ-5D-Y-3L instrument as a warm-up task.
4) 3 questions related to experience with illness.
5) 15 + 3 DCE tasks. In addition to the 15 pairs suggested in the protocol, we added 3 dominant pairs (same for all participants) for quality control. A pair is called dominant when a health state is clearly better than the alternative, in other words, all dimensions have equal

or lower severity level. Each participant completed the DCE part of the survey in the following order: a fixed dominant pair, the 15 pairs with 1 fixed dominant pair shown in a random order, and lastly the third dominant pair. The left-right positions of each health state within a pair were randomized. To ensure participant's attendance, each dimension's severity level was underlined and differences between the health states presented within a pair were highlighted using bold fonts (Figure 1).

6) Additional demographic questions. This section included participant's main activity, siblings, and questions related to serious illness in minors.

### Sampling and quality control

We aimed to collect 700 participants for the DCE. Sample size allow the estimation of reliable models, as previous evidence shows that one rarely requires more than 20-30, optimal 50 respondents per pair(19)(20). We increased the sample size to 700 participants because the incremental cost was very low, and it enables futures subgroups analyses. Sample representativeness with Spanish general population was ensured by quotas based on participants' age range (11-14 years old and 15-17 years-old) and gender, taking as reference data from the Office for National Statistics(21). The links to the implemented survey was shared with participants by a panel company.

We implemented a quality control procedure to detect and discard those surveys with suspicious response patterns. We established two rules to determine whether a participant was sufficiently engaged with the survey: participants who completed the DCE tasks in less than 2 minutes and 30 seconds (speeders) or incorrectly answered at least 2 out of 3 dominant pairs (inconsistent participants) were excluded. This allowed us to divide our DCE total sample as follows: included and excluded (only speeders, only inconsistent, speeders and inconsistent).

### Statistical analysis

Descriptive statistics were used to summarize respondent's characteristics, including the interviews failing the quality control rules. We used proportions for categorical variables and mean ± standard errors for continuous variables. To provide further details on the quality control, we analysed the time taken to complete the DCE task and the number of dominant pairs incorrectly answered by each participant.

We tested two type of models that consider the presence of random heterogeneity on the sample to analyse the DCE responses: random coefficient mixed logit and latent class models. The mixed logit model generalizes standard logit by allowing the coefficients associated with each measured variable to vary across participants. Then, the moments of the distribution of each participant-specific parameter are estimated(22).

The latent class analysis allowed us to identify subgroups (in this paper called classes) of participants who share homogeneous health preferences (23). To do so, the model defines a categorical latent variable (a variable that is not directly measured but identified from some measured variables) in which each level represents a class. The latent class analysis is a probabilistic approach; therefore, class membership is not fixed, and each participant is assigned a probability of belonging to each class. In essence, the purpose of the DCE latent class analysis is to identify the number of classes that describe the underlying preferences in the data, estimate the prevalence in groups and estimate each participant's probability of belonging to each class.

To determine the number of classes that best describe the underlying preferences in the data, we tested having from 2 to N classes, stopping when the model with N+1 classes becomes worse than the model with N classes. For each of the tested models, we calculated the accuracy in predicting the observed choice by means of the mean square error (MSE) and mean absolute error (MAE). The goodness of fit of the tested models was also estimated using both the Akaike Information Criteria (AIC) and the Bayesian information criterion (BIC). Among both, BIC penalizes model complexity more heavily. Previous studies show that BIC correctly identifies the number of classes more consistently across all models and sample sizes(24). However, as further research is needed to facilitate a greater understanding on the performance of these information criteria, we tested the behaviour of both AIC and BIC in latent class analysis with 2 to 9 classes.

We selected the best model considering three rules: 1) consistency of the model coefficients; 2) accuracy in predicting the observed choice probability and 3) Goodness of fit by the BIC.

Once the best model was selected, and due to the lack of C-TTO information in our sample, we anchored the DCE latent scale utilities onto the 1-full health 0-dead scale by using the pits value from the national EQ-5D-5L value set. The followed approach was “anchoring on the pits state utility value”, and it is mathematically equivalent to using linear rescaling of the DCE by using the pits state value (utility assigned to the most severe health state, usually coded as 55555)(25)(26).

Our analysis used incremental dummies, hence the estimated coefficients represents the difference between successive levels. The final value set is also reported using regular dummies. All analyses were performed using Microsoft Excel and Stata.

# Results

## *Descriptive statistics*

The DCE sample consisted in 702 out of 1132 respondents who completed the interview and passed the quality control criteria. Therefore, 430 participants were excluded from the study. Table 1 shows demographic characteristics of our sample. Overall, the excluded and included participants had similar demographic characteristics. In the included DCE sample, the average age was 14.09±2.06. The proportion of males was slightly higher than that of female, with 54.4% of the sample. Most participants (93.4%) were students, being the compulsory secondary education (55.8%) and the baccalaureate (25.9%) the levels of studies with the greatest representation. Half participants (50.1%) had at least a relative with a serious illness. The self-reported health using the EQ-5D-Y-3L showed that 49.1% participants reported problems in anxiety or depression. The following domains with the highest incidence of reported problems are usual activities (15.1%) and pain or discomfort (13.4%).

As previously commented, 38% (430/1132) participants who completed the DCE part were excluded in the quality control. Table 2 shows further details of the rules applied and the reasons that justify these participants withdrawal. Most of the excluded participants (77.7%) were discarded because they completed the DCE part of the survey in less than 150 seconds, i.e., they were classified as speeders. 22.3% participants were speeders and wrongly answered at least two dominant pairs (most of them completed the first dominant pair properly and failed the two following ones). Participants who were not speeders but inconsistent represents only 5.1% of the excluded sample.

## *Modelling results*

Latent class models outperformed the random coefficient mixed logit (Figure 2). The implied ranking (reflecting the size of the coefficients) was the same in all the tested models and can be found in Table 1S. All the tested latent class models were consistent. The model with the best results was the one with 4 classes, since it had the lowest BIC and no differences in terms of MAE and MSE in comparison with more complex models.

The selection of the BIC as the most appropriate information criteria (instead of the AIC) was further justified with the examination of both information criteria's performance in latent class models with 2-9 classes. Figure 3 shows that BIC reach its minimum value in the model with 4 classes and AIC decrease while the number of classes increase. Therefore, AIC is not properly reflecting model complexity increase when moving from N to N+1 classes.

Results after anchoring the 4-classes latent scale on the pits state were also consistent (Table 3). The class with the highest prevalence was class 4, representing the 33.6% of the study sample. The coefficients obtained provide information on the relative importance of each dimension within each, therefore allowing an easier interpretation of the preference heterogeneity present in the data. Hence, class 1 participants showed little variability between the relative importance of the dimensions, in other words, for these individuals the 5 dimensions included in the EQ-5D-Y-3L are almost equally important. For participants who belong to class 2 the most import dimension was mobility. Pain/discomfort is considered the most important dimension for adolescents included in class 3 and class 4. However, while for class 3 participants “mobility” is considered as the second most important dimension; in class 4 “anxiety/depression” is the second most important dimension.

The value set, i.e., overall results for the 4 classes is expressed in incremental and regular variables. The most important domains were “pain/discomfort” and “anxiety/depression” (Table 3).

## Discussion

In this study we applied the DCE side of the recent International Valuation Protocol to obtain health preferences in terms of the EQ-5D-Y-3L from a sample of adolescents in Spain.

The feasibility and validity of obtaining adolescents health preferences using their own perspective has been investigated in previous studies. Daziel et al.(27) collected EQ-5D-Y-3L health preferences from two perspectives: adults responding from the perspective of a 10-years old and adolescents aged 11-17 from own perspective. The authors analyzed different sample validation statistics (completion time, consistency in ordering, rational responses to dominant choice sets and balanced distributing of best-worst frequencies) and concluded that adolescents were able to engage and complete the task, providing rational responses. However, they used a BWS experiment instead of a DCE as we are using in our experiment. Mott et al.(28) followed a similar methodology, but using a DCE, and stated that, in comparison with adults, adolescents did not struggle significantly with the discrete choice experiment and they were able to complete the task. These results are aligned with findings for other instruments used to measure youth health preferences, such us the CHU-9D(29). However, as we used more dominant pairs than Mott et al hence allowing us to better estimate the probabilities to have valid responses. In our case, 38% participants who completed the DCE part of the survey failed the quality control rules. Ramos-Goñi et al. (16), applying the same quality control rules in adults, found that 30% did not pass the quality control rules. Therefore, there is a significant higher proportion of adolescents who were not

sufficiently engaged with the survey. These findings suggest that the validity of the responses in adolescents may be lower than in adults. Therefore, these experiments should be taken with caution in adolescent.

Even so, it is fair to say that the sample after the quality control procedure seems valid and it was feasible to obtain a DCE latent scale utilities. However, this the DCE latent scale is not anchored in a QALY scale (1-full health, 0-death) and only provide information about the relative importance of the different levels. In this manuscript, we have provided an approach to overcome this issue. As stated in methods section, our approach consists in taking an external anchor to rescale the latent scale results from the DCE. Taking that external anchor value and using the pits state anchoring methods explained by Stolk et al.(25) and Ramos-Goñi et al.(26) it is possible to have a final value set that allows QALY calculations.

The selection of a pit state utility based on adults' health preferences as anchoring value is justified with the results shown in Ramos-Goñi et al.(16), where majority respondents indicated that adults and adolescents should have equal priority when it comes to decisions on health care resource allocation. To ensure an equal resource allocation, adults and adolescents health preferences should be on the same scale.

Furthermore, the EQ-5D-5L pits state was selected instead of the EQ-5D-3L pits state for two main reasons: first, because due to the differences in labels and levels between the two instruments the pits state of the EQ-5D-3L is considerably lower than that of the EQ-5D-5L from the adult perspective, partly because the level 3 for mobility in the EQ-5D-3L is described as "confined to bed", while in the EQ-5D-5L is "Unable to walk about" which is more similar than the corresponding level for the EQ-5DY-3L ("A lot of problems walking about). Indeed, Ramos-Goñi et al.(16), found small differences on pits states values when estimating Spanish value sets for the EQ-5D-5L and EQ-5D-Y-3L.

Secondly, the most recent evidence about the Spanish cost-effectiveness threshold(30), defined as the maximum amount considered acceptable to pay per QALY, is based on EQ-5D-5L values collected in the Spanish Health Survey conducted in 2011-2012. Therefore, by rescaling by the EQ-5D-5L pit state utility, we are ensuring that our results are consistent with the latest estimation of the cost-effectiveness threshold. This approach advocate for health care equality between adolescent and adults.

A limitation of this study was that the DCE data was collected online. Previous research has found that online administration of DCE for health preferences has limitations in regard to participants' engagement(31). However, we minimized the impact of this limitation by the quality control process, based on the time taken to complete the DCE part of the survey and three dominant pairs.

In addition to participants engagement, the complexity of the DCE tasks in adolescents can be considered a limitation of our study. However, our DCE design minimized the impact of both limitations, since it used a 2-dimension overlap for all pairs (each pair have the same level in 2 dimensions) and bold coding to easily identify the differences between the health states presented within a pair(32).

Finally, the left-right bias reported by previous DCE (alternatives positioned to the left are chosen more frequently compared to alternatives positioned to the right, irrespectively of health state presented) was also minimized since we randomized the left-right positions of each health state within a pair.

## Conclusion

Our results provide a potential value set that allows cost-utility analyses in younger populations in Spain using their own perspective. The proposed value set may be to use to inform decision-making in the health resource allocation that is directly relevant to adolescents.

## References

1. Devlin NJ, Brooks R. EQ-5D and the EuroQol Group: Past, Present and Future. Appl Health Econ Health Policy [Internet]. 2017 Apr 13;15(2):127–37. Available from: http://link.springer.com/10.1007/s40258-017-0310-5

2. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ [Internet]. 2002 Mar;21(2):271–92. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0167629601001308

3. Sintonen H. The 15D instrument of health-related quality of life: properties and applications. Ann Med [Internet]. 2001 Jan 8;33(5):328–36. Available from: http://www.tandfonline.com/doi/full/10.3109/07853890109002086

4. Horsman J, Furlong W, Feeny D, Torrance G. The Health Utilities Index (HUI): concepts, measurement properties and applications. Health Qual Life Outcomes [Internet]. 2003 Oct 16;1:54. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14613568

5. Wisløff T, Hagen G, Hamidi V, Movik E, Klemp M, Olsen JA. Estimating QALY Gains in Applied Studies: A Review of Cost-Utility Analyses Published in 2010. Pharmacoeconomics [Internet]. 2014 Apr 30;32(4):367–75. Available from: http://link.springer.com/10.1007/s40273-014-0136-z

6. National Institute of Health and ClinicalExcellence. Guide to the Methods of Technology Appraisal. Natl Inst Heal Clin Excell London. 2008;

7. Badia X, Roset M, Herdman M, Kind P. A Comparison of United Kingdom and Spanish General Population Time Trade-off Values for EQ-5D Health States. Med Decis Mak [Internet]. 2001 Feb 2;21(1):7–16. Available from: http://journals.sagepub.com/doi/10.1177/0272989X0102100102

8. Ramos-Goñi JM, Craig BM, Oppe M, Ramallo-Fariña Y, Pinto-Prades JL, Luo N, et al. Handling Data Quality Issues to Estimate the Spanish EQ-5D-5L Value Set Using a Hybrid Interval Regression Approach. Value Heal [Internet]. 2018 May;21(5):596–604. Available from: https://linkinghub.elsevier.com/retrieve/pii/S109830151733663X

9. Wille N, Badia X, Bonsel G, Burström K, Cavrini G, Devlin N, et al. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. Qual Life Res. 2010 Aug;19(6):875–86.

10. Griebsch I. Quality-Adjusted Life-Years Lack Quality in Pediatric Care: A Critical Review of Published Cost-Utility Studies in Child Health. Pediatrics [Internet]. 2005 May 1;115(5):e600–14. Available from: http://pediatrics.aappublications.org/cgi/doi/10.1542/peds.2004-2127

11. Kreimeier S, Oppe M, Ramos-Goñi JM, Cole A, Devlin N, Herdman M, et al. Valuation of EuroQol Five-Dimensional Questionnaire, Youth Version (EQ-5D-Y) and EuroQol Five-Dimensional Questionnaire, Three-Level Version (EQ-5D-3L) Health States: The Impact of Wording and Perspective. Value Heal. 2018 Nov;21(11):1291–8.

12. Ratcliffe J, Stevens K, Flynn T, Brazier J, Sawyer MG. Whose Values in Health? An Empirical Comparison of the Application of Adolescent and Adult Values for the CHU-9D and AQOL-6D in the Australian Adolescent General Population. Value Heal [Internet]. 2012 Jul;15(5):730–6. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1098301512015483

13. Deparmet of Health Chief Medical Offier's annual report. Our Children Deserve Better:Prevention Pays. 2012.

14. Ministerio de Sanidad y Consumo. Ganar salud con la juventud. 2008.

15. Ramos-Goñi JM, Oppe M, Stolk E, Shah K, Kreimeier S, Rivero-Arias O, et al. International Valuation Protocol for the EQ-5D-Y-3L. Pharmacoeconomics [Internet]. 2020 Jul 16;38(7):653–63. Available from: http://link.springer.com/10.1007/s40273-020-00909-3

16. Ramos-Goñi JM, Oppe M, Estévez-Carrillo A, Rivero-Arias O. Modelling strategies for generating country-specific EQ-5D-Y value sets. In: EuroQol Special Interest Groups conference. 2020.

17. Rowen D, Rivero-Arias O, Devlin N, Ratcliffe J. Review of Valuation Methods of Preference-Based Measures of Health for Economic Evaluation in Child and Adolescent Populations: Where are We Now and Where are We Going? Pharmacoeconomics [Internet]. 2020 Apr 6;38(4):325–40. Available from: http://link.springer.com/10.1007/s40273-019-00873-7

18. Kreimeier S, Åström M, Burström K, Egmar A-C, Gusi N, Herdman M, et al. EQ-5D-Y-5L: developing a revised EQ-5D-Y with increased response categories. Qual Life Res [Internet]. 2019 Jul 9;28(7):1951–61. Available from: http://link.springer.com/10.1007/s11136-019-02115-x

19. Lancsar E, Louviere J. Conducting Discrete Choice Experiments to Inform Healthcare Decision Making. Pharmacoeconomics [Internet]. 2008;26(8):661–77. Available from: http://link.springer.com/10.2165/00019053-200826080-00004

20. de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. Health Econ [Internet]. 2012 Feb;21(2):145–72. Available from: http://doi.wiley.com/10.1002/hec.1697

21. Insitituto Nacional de Estadística. Población por sexo, comunidades y provincias y edad (hasta 100 y más). [Internet]. 2020. Available from: https://www.ine.es/up/Cwidbvxki1

22. Revelt D, Train K. Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level. Rev Econ Stat [Internet]. 1998 Nov;80(4):647–57. Available from: http://www.mitpressjournals.org/doi/10.1162/003465398557735

23. Kongsted A, Nielsen AM. Latent Class Analysis in health research. J Physiother [Internet]. 2017 Jan;63(1):55–8. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1836955316300443

24. Nylund KL, Asparouhov T, Muthén BO. Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. Struct Equ Model A Multidiscip J [Internet]. 2007 Oct 23;14(4):535–69. Available from: https://www.tandfonline.com/doi/full/10.1080/10705510701575396

25. Stolk EA, Oppe M, Scalone L, Krabbe PFM. Discrete Choice Modeling for the Quantification of Health States: The Case of the EQ-5D. Value Heal [Internet]. 2010 Dec;13(8):1005–13. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1098301511718344

26. Ramos-Goñi JM, Rivero-Arias O, Errea M, Stolk EA, Herdman M, Cabasés JM. Dealing with the health state 'dead' when using discrete choice experiments to obtain values for EQ-5D-5L heath states. Eur J Heal Econ [Internet]. 2013 Jul 31;14(S1):33–42. Available from: http://link.springer.com/10.1007/s10198-013-0511-2

27. Dalziel K, Catchpool M, García-Lorenzo B, Gorostiza I, Norman R, Rivero-Arias O. Feasibility, Validity and Differences in Adolescent and Adult EQ-5D-Y Health State Valuation in Australia and Spain: An Application of Best–Worst Scaling. Pharmacoeconomics [Internet]. 2020 May 24;38(5):499–513. Available from: http://link.springer.com/10.1007/s40273-020-00884-9

28. Mott DJ, Shah KK, Ramos-Goñi JM, Devlin NJ, Rivero-Arias O. Valuing EQ-5D-Y Health

States Using a Discrete Choice Experiment: Do Adult and Adolescent Preferences Differ? OHE Res Pap [Internet]. 2019; Available from: https://www.ohe.org/publications/valuing-eq-5d-y-health-states-using-discrete-choice-experiment-do-adult-and-adolescent

29. Ratcliffe J, Huynh E, Stevens K, Brazier J, Sawyer M, Flynn T. Nothing About Us Without Us? A Comparison of Adolescent and Adult Health-State Values for the Child Health Utility-9D Using Profile Case Best-Worst Scaling. Health Econ [Internet]. 2016 Apr;25(4):486–96. Available from: http://doi.wiley.com/10.1002/hec.3165

30. Vallejo-Torres L, García-Lorenzo B, Serrano-Aguilar P. Estimating a cost-effectiveness threshold for the Spanish NHS. Health Econ [Internet]. 2018 Apr;27(4):746–61. Available from: http://doi.wiley.com/10.1002/hec.3633

31. Mulhern B, Longworth L, Brazier J, Rowen D, Bansback N, Devlin N, et al. Binary Choice Health State Valuation and Mode of Administration: Head-to-Head Comparison of Online and CAPI. Value Heal [Internet]. 2013 Jan;16(1):104–13. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1098301512041265

32. Jonker MF, Donkers B, de Bekker-Grob EW, Stolk EA. Effect of Level Overlap and Color Coding on Attribute Non-Attendance in Discrete Choice Experiments. Value Heal [Internet]. 2018 Jul;21(7):767–71. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1098301517336094

## Tables and figures

Figure 1. Example of a DCE task in the online survey.

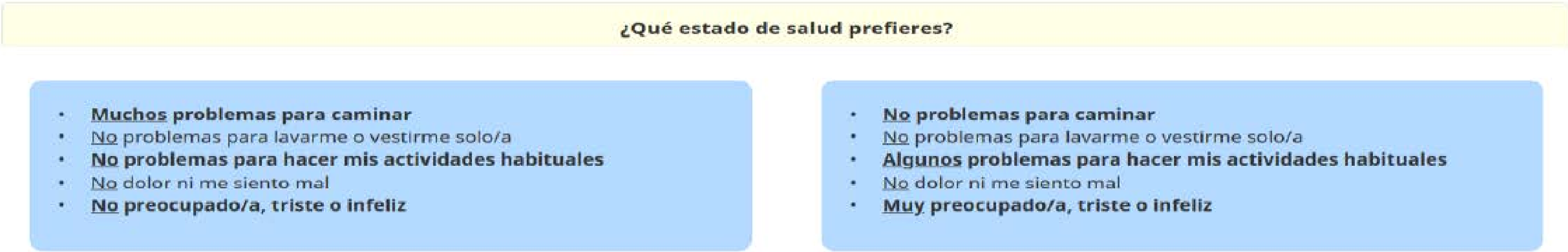

Table 1. Sample description.

| Variables | DCE Excluded (n=430) | | DCE Included (n=702) | | Total (n=1132) | |
|---|---|---|---|---|---|---|
| | N | % | N | % | N | % |
| Age [mean (SD)] | 14.15 | 2.04 | 14.09 | 2.06 | 14.11 | 2.05 |
| Age groups | | | | | | |
| 11-14 | 226 | 52.6 | 379 | 54.0 | 605 | 53.4 |
| 15-17 | 204 | 47.4 | 323 | 46.0 | 527 | 46.6 |
| Sex | | | | | | |
| Male | 250 | 58.1 | 382 | 54.4 | 632 | 55.8 |
| Female | 180 | 41.9 | 320 | 45.6 | 500 | 44.2 |
| Employment status | | | | | | |
| Student | 355 | 82.6 | 656 | 93.4 | 1011 | 89.3 |
| Apprentice | 11 | 2.6 | 5 | 0.7 | 16 | 1.4 |
| Employee | 51 | 11.9 | 33 | 4.7 | 84 | 7.4 |
| None | 10 | 2.3 | 4 | 0.6 | 14 | 1.2 |
| Missing | 3 | 0.7 | 4 | 0.6 | 7 | 0.6 |
| Education | | | | | | |
| Compulsory secondary education | 218 | 50.7 | 392 | 55.8 | 610 | 53.9 |
| Baccalaureate | 113 | 26.3 | 182 | 25.9 | 295 | 26.1 |
| Other | 24 | 5.6 | 82 | 11.7 | 106 | 9.4 |
| Missing | 75 | 17.4 | 46.00 | 6.6 | 121 | 10.7 |
| Experience with illness | | | | | | |
| Personal (%yes) | 79 | 18.4 | 82 | 11.7 | 161 | 14.2 |
| Relatives (%yes) | 209 | 48.6 | 352 | 50.1 | 561 | 49.6 |
| Others (%yes) | 60 | 14.0 | 57 | 8.1 | 117 | 10.3 |
| Self-reported EQ-5D-Y | | | | | | |
| Mobility | | | | | | |
| No problems | 373 | 86.7 | 665 | 94.7 | 1038 | 91.7 |
| Some problems | 52 | 12.1 | 30 | 4.3 | 82 | 7.2 |
| A lot of problems | 5 | 1.2 | 7 | 1.0 | 12 | 1.1 |
| Self-care | | | | | | |
| No problems | 381 | 88.6 | 685 | 97.6 | 1066 | 94.2 |
| Some problems | 40 | 9.3 | 15 | 2.1 | 55 | 4.9 |
| A lot of problems | 9 | 2.1 | 2 | 0.3 | 11 | 1.0 |
| Usual activities | | | | | | |
| No problems | 340 | 79.1 | 596 | 84.9 | 936 | 82.7 |
| Some problems | 78 | 18.1 | 71 | 10.1 | 149 | 13.2 |
| A lot of problems | 12 | 2.8 | 35 | 5.0 | 47 | 4.2 |
| Pain/discomfort | | | | | | |
| No pain or discomfort | 340 | 79.1 | 608 | 86.6 | 948 | 83.7 |
| Some pain or discomfort | 81 | 18.8 | 92 | 13.1 | 173 | 15.3 |
| A lot of pain or discomfort | 9 | 2.1 | 2 | 0.3 | 11 | 1.0 |
| Anxiety/depression | | | | | | |
| No worried, sad or unhappy | 243 | 56.5 | 357 | 50.9 | 600 | 53.0 |
| A bit worried, sad or unhappy | 167 | 38.8 | 311 | 44.3 | 478 | 42.2 |
| Very worried, sad or unhappy | 20 | 4.7 | 34 | 4.8 | 54 | 4.8 |

Table 2. Quality control and inclusion/exclusion criteria

| | DCE Excluded | | | | | | DCE Included | |
|---|---|---|---|---|---|---|---|---|
| | Non-speeders | | Speeders | | Total | | Total | |
| | n | % | n | % | n | % | n | % |
| **<2 Inconsistencies** | - | - | **334** | **77.7** | **334** | **77.7** | **702** | **100** |
| 0 | - | - | 209 | 62.6 | 209 | 62.6 | 622 | 88.6 |
| 1 | - | - | 125 | 37.4 | 125 | 37.4 | 80 | 11.4 |
| **≥2 Inconsistencies** | **22** | **5.1** | **74** | **17.2** | **96** | **22.3** | **-** | **-** |
| 2 | 18 | 80 | 67 | 90.5 | 85 | 88.5 | - | - |
| 3 | 4 | 20 | 7 | 9.5 | 11 | 11.5 | - | - |
| **Total** | **22** | **5.1** | **408** | **94.9** | **430** | **100** | **702** | **100** |

Figure 2. DCE models comparison

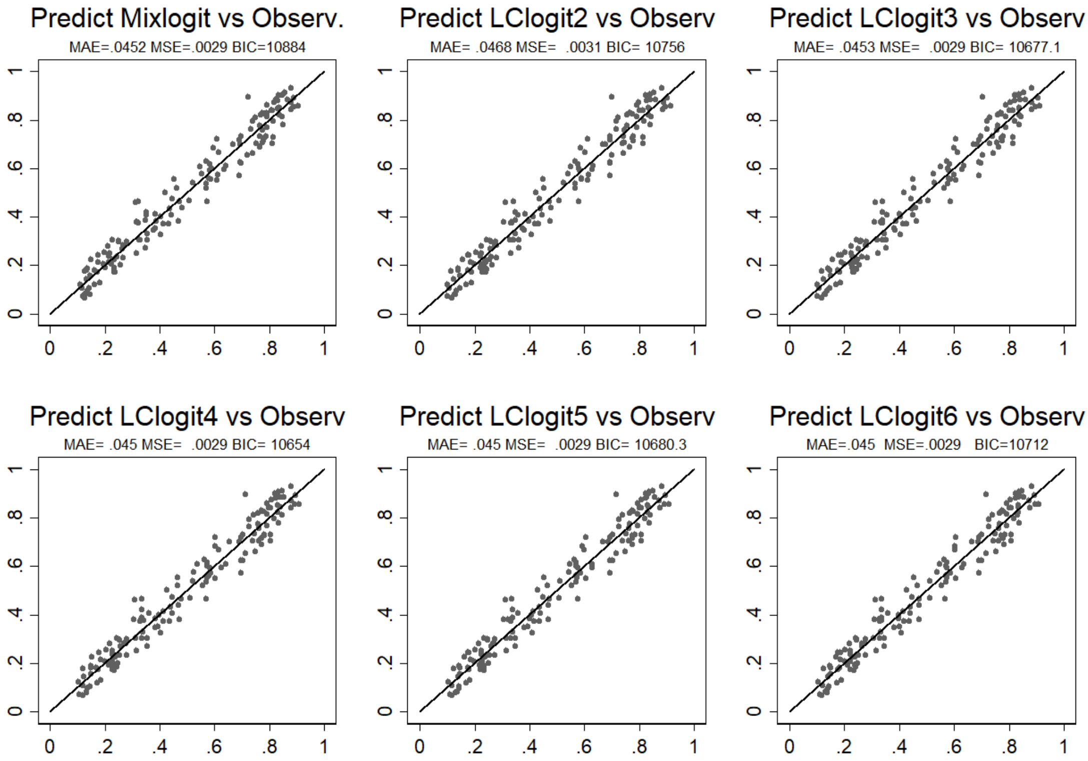

Figure 3. Information criterias' performance in latent class models with increasing number classes

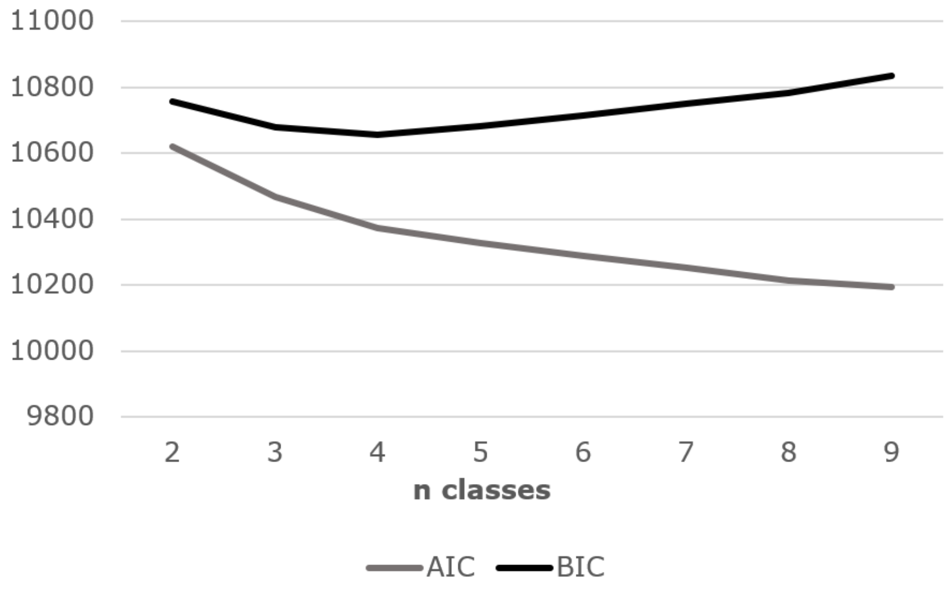

**Table 3.** Results after anchoring for each of the 4 latent classes.

| Dimension | Independent variables of the model | LC1 | LC2 | LC3 | LC4 | Value set | |
|---|---|---|---|---|---|---|---|
| | | | | | | Incremental | Regular |
| | Class Share | 0.242 | 0.168 | 0.254 | 0.336 | | |
| Mobility | No problems to some problems | -0.084 | -0.262 | -0.058 | -0.044 | -0.072 | -0.072 |
| | Some problems to a lot of problems | -0.152 | -0.576 | -0.310 | -0.125 | -0.222 | -0.294 |
| Self-Care | No problems to some problems | -0.142 | 0.014 | 0.014 | -0.057 | -0.041 | -0.041 |
| | Some problems to a lot of problems | -0.203 | -0.230 | -0.182 | -0.112 | -0.155 | -0.196 |
| Usual Activities | No problems to some problems | -0.092 | -0.022 | -0.071 | -0.109 | -0.088 | -0.088 |
| | Some problems to a lot of problems | -0.199 | -0.113 | -0.130 | -0.153 | -0.149 | -0.237 |
| Pain/ Discomfort | No to some | -0.109 | 0.032 | -0.082 | -0.180 | -0.124 | -0.124 |
| | Some to a lot of | -0.219 | -0.183 | -0.348 | -0.263 | -0.275 | -0.399 |
| Anxiety/ Depression | Not to a bit | -0.027 | -0.069 | -0.058 | -0.131 | -0.091 | -0.091 |
| | A bit to very | -0.188 | -0.007 | -0.190 | -0.240 | -0.199 | -0.290 |
| | U (33333) | -0.416 | -0.416 | -0.416 | -0.416 | -0.416 | -0.416 |

## Appendix

Table 1S. DCE models results (incremental dummies).

| Dimension | Independent variables of the model | Mix logit | Latent class | | | | |
|---|---|---|---|---|---|---|---|
| | | | 2 classes | 3 classes | 4 classes | 5 classes | 6 classes |
| Mobility | No to some problems | -0.489 | -0.388 | -0.420 | -0.455 | -0.462 | -0.473 |
| | Some to a lot of problems | -1.551 | -1.236 | -1.346 | -1.396 | -1.597 | -1.564 |
| Self-care | No to some problems | -0.332 | -0.252 | -0.297 | -0.261 | -0.370 | -0.321 |
| | Some to a lot of problems | -1.153 | -0.903 | -0.934 | -0.974 | -1.085 | -1.060 |
| Usual activities | No to some problems | -0.615 | -0.489 | -0.552 | -0.557 | -0.642 | -0.613 |
| | Some to a lot of problems | -1.050 | -0.840 | -0.924 | -0.938 | -1.056 | -1.043 |
| Pain/discomfort | No to some pain or discomfort | -0.899 | -0.688 | -0.780 | -0.781 | -0.846 | -0.836 |
| | Some to a lot of pain or discomfort | -2.011 | -1.570 | -1.687 | -1.732 | -1.868 | -1.862 |
| Anxiety/depression | No to a bit worried, sad or unhappy | -0.621 | -0.497 | -0.578 | -0.574 | -0.608 | -0.598 |
| | A bit to a very worried, sad or unhappy | -1.412 | -1.088 | -1.214 | -1.252 | -1.322 | -1.320 |