Social Science & Medicine 292 (2022) 114519

Contents lists available at ScienceDirect

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed

An EQ-5D-5L value set for Italy using videoconferencing interviews and feasibility of a new mode of administration

Aureliano Paolo Finch [a,c,*], Michela Meregaglia [b,1], Oriana Ciani [b,1], Bram Roudijk [a], Claudio Jommi [b]

[a] EuroQol Office, EuroQol Research Foundation, Rotterdam, the Netherlands
[b] Centre for Research on Health and Social Care Management (CERGAS), SDA Bocconi School of Management, Bocconi University, Milan, Italy
[c] Health Values Research and Consultancy, Amsterdam, the Netherlands

ARTICLE INFO

Keywords:
EQ-5D-5L
cTTO
Preferences
Valuation study
Mode of administration
Videoconferencing
Video-telephony
Feasibility

ABSTRACT

Objectives: To test the feasibility of using videoconferencing (VC) administered interviews and to derive an EQ-5D-5L value set for Italy.
Methods: Preferences were collected using the EuroQol standardized valuation protocol (EQ-VT) administered via VC. Two valuation methods were employed, composite time trade-off (cTTO) and discrete choice experiment (DCE). Technical, organizational and protocol feasibility were tested in a pilot of 198 interviews. Upon positive assessment, data collection continued with a target sample of 1000–1200 participants including the pilot. Quality control (QC) procedures were employed to monitor interviewers' performance during the pilot and the data collection. Data were modelled using GLS, Tobit, Logit and Hybrid models with different error specifications. Monotonicity of coefficients, statistical significance, and theoretical considerations informed the model choice.
Results: Dropouts and technical problems occurred in less than 5% of the 198 pilot interviews. Protocol compliance was demonstrated with significant improvements in QC parameters and limited interviewers' effects, for all interviewers. Overall, interviewers were satisfied with this mode of administration, highlighting it allows flexibility and efficient scheduling. Based on these results, VC was deemed as a feasible mode of administration. The study collected preferences for 1182 responders, including the pilot interviews. The demographic characteristics of the sample were representative of the Italian general population for age, gender and geographical macro-areas. The hybrid Tobit heteroscedastic model without constant estimated on the full sample (including pilot) was selected for the derivation of the value set. Values ranged from −0.571 for the worst health state (55555) to 1 for the best health state (11111). Pain/discomfort registered the largest decrement, followed by mobility, anxiety/depression, self-care, and usual activities. 523 health states were worse than dead.
Conclusions: VC is viable for the conduct of valuation interviews. The Italian value set for the EQ-5D-5L can be used for value determinations of health technologies.

1. Introduction

Health related quality of life (HRQoL) measures are frequently used to report patients' health (Brazier et al., 1999). While most HRQoL measures use the summation of their dimensions' levels as a scoring system, a subclass of them uses preferences. These measures are sometimes referred to as generic preference-based measures, health state utility value measures, multi-attribute utility measures (Finch et al., 2019) and preference accompanied measures (Devlin, 2020).

One of the main advantages of employing preferences as a scoring system is that this allows for the calculation of quality adjusted life years (QALYs), the outcome of cost-utility analysis (CUA) of health care interventions (Finch et al., 2019; Devlin, 2020). The 2020 guidelines of the Italian Drug Agency (AIFA), the Italian health technology assessment body, state that CUAs are a mandatory component of the national-level pricing and reimbursement (P&R) dossier and that, whenever possible, utility data should be referred to the Italian context (Linee Guida, 2020). CUAs may also be employed at the regional level for prioritizing

* Corresponding author. EuroQol Office, EuroQol Research Foundation, Marten Meesweeg 107, 3068, AV, Rotterdam, the Netherlands.
E-mail address: finch@euroqol.org (A.P. Finch).
[1] These authors contributed equally.

https://doi.org/10.1016/j.socscimed.2021.114519
Received 9 July 2021; Received in revised form 22 September 2021; Accepted 25 October 2021
Available online 28 October 2021

reimbursement decisions for technologies which are not appraised at the central level (Favaretti et al., 2009) or for the evaluation of other technologies such as medical devices (Tarricone et al., 2021). For the conduct of CUAs, the Italian Health Economics Association recommends using the EQ-5D as their preferred instrument (Fattore, 2009).

The EQ-5D is the most widely used generic preference accompanied measure of health (Richardson et al., 2015). Its original descriptive system i.e. EQ-5D-3L comprises of five health dimensions described in terms of three response levels (EuroQol Group (1990a, 1990b)). More recently, the EuroQol Group has developed a new measure with increased "granularity", where for each of the five dimensions the response levels were increased to five i.e. EQ-5D-5L (Herdman et al., 2011). The EQ-5D is valid and responsive in multiple disease areas, conditions and cultural contexts (Janssen et al., 2018; Finch et al., 2018; Qian et al., 2019), with the EQ-5D-5L reporting better psychometric properties than the EQ-5D-3L (Janssen et al., 2013a; Devlin et al., 2018; Gandhi et al., 2019; Jin et al., 2019). While there is an Italian EQ-5D-3L value set i.e. scoring based on preferences (Scalone et al., 2013), there is not an EQ-5D-5L one.

EQ-5D-5L value sets are developed following the EuroQol Valuation Technology (EQ-VT) protocol (Pickard et al., 2019; Augustovski et al., 2016; Xie et al., 2016), an important requirement of which is the conduct of one-on-one, face-to-face interviews (Oppe et al., 2016; Ramos-Goñi et al., 2017a). This mode of administration (MoA) has been seen to produce data of better quality compared to group administered interviews and online self-completed interviews (Shah et al., 2013; Norman et al., 2010), in part due to the greater support and guidance provided by the interviewers to the interviewee (Shah et al., 2013; Edelaar-Peeters et al., 2014). As a response to the Covid-19 pandemic, numerous governments, including the Italian one, imposed travel restrictions, establishments closures and limitations of in-person meetings (Bosa et al., 2021), which prevented the possibility of collecting valuation data in one-on-one, face-to-face interviews.

An alternative MoA that may preserve the required interaction between interviewers and interviewees is one-on-one videoconferencing (VC) administered interviews. The strength and weakness of VC interviews have been discussed in the literature in general (e.g. Deakin and Wakefield, 2014, Bertrand et al., 2010) and for valuations in particular (Lipman, 2020). However, it remains unclear whether VC is feasible for conducting a full EQ-5D-5L valuation for multiple reasons. First, the only current evidence is based on a single interviewer with prior experience in valuation interviews (Lipman, 2020), while EQ-5D-5L valuations (and to a lesser degree EQ-5D-3L and EQ-5D-Y) require the involvement of multiple and often inexperienced interviewers. Second, the study only monitored a partial set out of the key feasibility indicators (Lipman, 2020), resulting in scattered evidence. Third, this study assessment and reporting is often unstructured, i.e. evidence can be based on the interviewer's recollection of events.

Three types of feasibility are of importance when assessing the use of VC interviews in the context of an EQ-5D valuation study, namely technical, organizational and protocol feasibility. Technical feasibility refers to the possibility of using the EQ-VT software in a VC administered setting with limited technical (e.g. connectivity, audio-visual quality) issues. Organizational feasibility relates to the possibility of setting up a VC administered valuation study, scheduling of hundreds of interviews and handling of delays or no show in a manageable way. Protocol feasibility refers to the possibility of complying with the EQ-VT protocol (Oppe et al., 2016), and its QC procedures (Ramos-Goñi et al., 2017a) when using VC interviews. Feasibility is further presented in the methods section.

The current study had two aims. First, it tested the feasibility of using VC administered interviews for EQ-5D valuation studies. Second, conditional on an assessment of VC feasibility, it derived an EQ-5D-5L value set for Italy using this MoA.

2. Methods

This study was approved by the Ethics Committee of Bocconi University in date October 6, 2020 (approval number: 2020-SA000136.4). It was conducted following the EQ-VT protocol (Oppe et al., 2016; Ramos-Goñi et al., 2017a) and its reporting complies with the CREATE checklist for reporting valuation studies (Xie et al., 2015).

2.1. Descriptive system and EQ VAS

The EQ-5D-5L descriptive system comprises of five items related to five dimensions of health, namely mobility, self-care, usual activities, pain/discomfort, and anxiety/depression (Herdman et al., 2011). Each item is described in terms of 5 levels, no problems, slight problems, moderate problems, severe problems and extreme problems/unable to. This results in 3125 possible combinations of dimensions levels, commonly referred to as health states.

The EQ-5D comes with a self-rating question of overall health status, measured on a 0–100 visual analogue scale (EQ VAS) (EuroQol Group (1990a, 1990b)).

2.2. Preference elicitation methods

Preferences were collected using the EuroQol Valuation Technology (EQ-VT), a software used for the conduct of EQ-5D-5L valuation studies. Two elicitation methods were used, composite time trade-off (cTTO) and discrete choice experiment (DCE).

cTTO is an indifference procedure that applies classical TTO for the valuation of health states considered better than dead (BTD) and lead-time TTO for the valuation of health states considered worse than dead (WTD) (Oppe et al., 2016; Janssen et al., 2013b). In classical TTO the responder is asked to choose between life A in full health for x years or life B in an intermediate health state for t years. The value of x varies between 10 years and 0 years, with the smallest unit of iteration set at 6 months, while t is set at 10 years. The task ends when the participant expresses indifference between life A and life B, at which point the health state value is given by time x/10. For some health states, some respondents may trade off the whole time in life A, indicating the value for that state is less than or equal to 0, in which case lead-time TTO is presented. In lead-time TTO, the intermediate health state lasts for 10 years but is preceded by a lead-time of 10 years. The responder trades x of the 10 years of lead-time of life A to achieve an indifference between life A and life B, once again using 6 months as the possible smallest iteration. At the indifference point the health state value is given by ((x-10)/10). cTTO attainable values range between 1 and -1. For a more thorough explanation of this method please refer to Devlin and colleagues (Devlin et al., 2013).

DCE presents the participant with a single choice situation with two paired EQ-5D-5L health states, option A and option B, without a duration attribute. Participants are asked "which is better, A or B" forcing them to choose between the two options.

2.3. Health state selection

The standard EQ-VT design was employed for both cTTO and DCE (Oppe and Hout, 2017; Oppe et al., 2014). The cTTO design comprised of 86 health states covering a range of possible severities, divided into 10 blocks. Each block included 1 mild state i.e. health state with only 1 deviation from full health, the most severe health state i.e. 55555 and 8 unique states. The DCE design comprised of 196 choice tasks, 186 of which selected from an efficient design maximized for the D-error, and 10 of which selected as mild pairs comparisons (Oppe and Hout, 2017). Choice tasks were assigned to one of 28 blocks, with 7 choice tasks per block.

2.4. Sampling, recruitment, and data collection procedures

The target sample was 1000–1200 participants (Oppe and Hout, 2017) including the pilot, representative of the Italian adult non-institutionalized population. Quotas were used for age, gender and geographical distribution for macro-areas (north-east, north-west, centre, south and islands) (ISTAT).

A survey agency experienced in quantitative and qualitative healthcare research, Pepe Research s.r.l., organized the recruitment and interviews scheduling. Candidate responders were identified using a panel and a network of local recruiters. Participants were provided information on the study objectives and the technical requirements for part taking[2] and upon agreement an interview was scheduled. A scheduling assistant software (TIMIFY) was used to facilitate scheduling and communication between the company, the interviewers, and the interviewees (an example is available in Appendix Fig. 1). The day prior to the interview, each participant received a phone reminder.

Data were collected between October 2020 and February 2021 using computer assisted personal interviews (CAPI) administered through Zoom as first choice VC software, and Lifesize as a backup option. Interviewers first launched the EQ-VT software, then connected to Zoom and allowed participants into the VC call, and finally shared their screen. In case a participant did not access the link at the scheduled time he/she was contacted by phone and received technical support if needed (based on a technical support document available to all interviewers).

In the interviews, responders first completed the EQ-5D-5L, the EQ VAS and a familiarization session. The familiarization session comprised a classical TTO and a lead-time-TTO (asking responders to imagine living in a wheelchair for 10 years and then dying, or in a state much worse than being in a wheelchair for 10 years and then dying), as well as three practice states (one mild, one severe and one difficult to imagine). Subsequently, the 10 real cTTO questions, the feedback module and 7 real DCE questions were presented. The feedback module presents responders with a ranking of health states according to the assigned values in the valuation exercise, and offers the possibility of highlighting whether the ordering is inconsistent based on the responders self-assessment (Ramos-Goñi et al., 2017a). Upon completion of the interview, participants received a 20€ voucher.

2.5. Interviewers' training and QC procedures

Eleven interviewers participated in a VC administered training. All interviewers were members of the staff, MSc, or PhD students at Bocconi University. The training familiarized interviewers with the EQ-5D, the EQ-VT protocol, the elicitation techniques, and the content of the QC reports.

QC reports comprise two important aspects, protocol compliance and interviewers' effects. Protocol compliance is assessed based on 4 quantitative indicators related to completion times, inconsistency size and use of sections of the valuation interview.[3] Interviewers' effects are assessed by comparing the distribution of cTTO values per interviewer to the overall distribution of values for all interviewers. Other meta-data e.g. number of moves, severity gradient of values per interviewer etc. are employed to aid the interpretation of the report.

After the training, each interviewer conducted 10 practice interviews with family and friends, using both face-to-face and VC administered interviews. The real interviews (including pilot interviews) were all conducted using VC. The remainder of the fieldwork was implemented in batches of 10 interviews per interviewer initially, with this requirement being relaxed at later stages of the data collection. Interviewers' performance was assessed using the QC report in a joint call with the EuroQol VT support team after each round of data collection and the feedback received was discussed in a group call with all interviewers. Similar procedures have been already employed for the collection of EQ-5D-5L data and have shown to improve data quality (Purba et al., 2017a).

2.6. Feasibility of videoconferencing interviews

VC administered interviews were piloted in 2 consecutive batches of 10 interviews per interviewer between October and November 2020 by monitoring and reporting indicators of technical, organizational and protocol feasibility. After feasibility was demonstrated in the pilot, monitoring continued for the duration of the whole study. This formal assessment of the compliance with the feasibility aspects of VC administered interviews occurred in a call with the EQ-VT support team based on discussion of feasibility indicators rather than on a predefined formal threshold set for study continuation. Compliance with all three feasibility aspects was required to continue the study. Data collected up to that point were considered part of the full sample for the EQ-5D-5L value set determination if feasibility was demonstrated.

More specifically, for each dimension of feasibility investigated we tracked: a) technical feasibility indicators, i.e. the frequency of problems such as unstable internet connection, difficulty in accessing the VC link, activating the microphone, visualizing the screen, etc., b) organizational feasibility indicators, i.e. frequency and percentages of dropouts, rescheduled interviews, etc. c) protocol feasibility indicators, i.e. frequency of problems related to the QC compliance and interviewers' effects (Stolk et al., 2019). A predefined spreadsheet was developed to collect information from the interviewers about these indicators. A customized feedback form featuring 20 Likert scale questions (five levels from 'strongly agree' to 'strongly disagree') adapted from the Post-Study system usability questionnaire (PSSUQ) was used to collect interviewers evaluation on the use of VC (13 questions) and how this interacted with the EQ-VT software (7 questions). Additionally, 4 open-ended questions were included to probe interviewers on acceptability of VC interviews, specific aspects of using this MoA compared to face-to-face interviews, and potential actions to improve the use of VC interviews for valuation studies (Appendix Table 1). The survey was administered via Qualtrics® XM platform. Quantitative data were tabulated and analysed descriptively using Microsoft® Excel v16.50. Qualitative data from the VC interviews evaluation form were analysed thematically and summarised in a narrative fashion.

2.7. cTTO and DCE data analysis

Data were analysed using STATA version MP V14.1. To inform the most appropriate modelling method, multiple alternative regression models were compared, for cTTO and DCE data alone, and using both preference data in combination.

For cTTO the dependent variable was expressed as the disutility of the observed cTTO value i.e. 1 –observed value while for DCE it was the respondent stated choice i.e. choice A or choice B. Both cTTO and DCE regressed the dependent variable over twenty dummy variables, which represented the EQ-5D-5L dimensions levels different from 1. In cTTO, dummy coefficients should be interpreted as decrements associated to departures from full health. In DCE, coefficients are not anchored on the QALY scale, meaning that direct comparisons are not possible. Rescaling of the latent coefficients was based on two anchoring approaches. In the first, a rescaling parameter was obtained by dividing the observed value for the worse cTTO state i.e. 55555 by the DCE predicted 55555 value. DCE coefficients were then multiplied by this rescaling parameter (Shah et al., 2020). In the second, the exponential of the hybrid Tobit

[2] Laptop with minimum screen size of 11 inches to ensure a correct interaction with the interviewer and a stable internet connection, among the others.

[3] The 4 protocol compliance indicators are: i) at least 3 min spent on the two wheelchair examples; ii) the interviewer entered the WTD component in at least 1 of the 2 wheelchair examples; iii) there is no serious inconsistency for the state 55555; iv) the interviewer spent at least 5 min on the 10 real TTO questions.

heteroscedastic model theta parameter was multiplied by the coefficients of the DCE data (Ramos-Goñi et al., 2017b).

Modelling of cTTO data first investigated a generalized least square (GLS) random intercept model which accounted for the panel structure of the data. Homoscedasticity of the data was tested using the Breusch–Pagan/Cook-Weisberg test. In presence of heteroscedasticity, a linear heteroscedastic model was fitted to the data. Alternative specifications were investigated based on the size and significance of the regression' coefficients and intercept. The frequency of censored data i.e. feature of the cTTO method that does not allow to trade beyond the value of −1 was assessed to determine the appropriateness of Tobit modelling. Tobit random effects models that accounted for the panel structure of the data, and possible heteroscedasticity, were also considered.

The analysis further explored whether modelling could be improved using approaches that combine cTTO and DCE data, commonly referred as hybrid models (Stolk et al., 2019). The same assumptions of the cTTO data were tested in this suit of models, including the hybrid, hybrid Tobit and hybrid Tobit heteroscedastic models.

Models' performance was evaluated in terms of logical consistency of the parameter estimates, i.e. monotonicity (non-decreasing coefficients at increasing severity levels) and parameters significance. Theoretical considerations (e.g. handling of censored values) were also used to inform the final model choice, as well as mean absolute error (sum of absolute errors between observed and predicted states) and mean absolute errors for states at the top and bottom of the scale.

To test the robustness of the selected model two sensitivity analyses were performed. In the first, the selected model was re-estimated using a reduced sample that removed the pilot data (n = 986). In the second, the selected model was re-estimated excluding the health states identified in the feedback module.

3. Results

In total, 1182 VC interviews were completed between October 2020 and February 2021, including 198 feasibility pilot interviews completed at the moment of the formal assessment with the VT support team. The mean interview time was 42.56 ± 17.48 min. The sample was representative of the Italian adult population for gender, geographical distribution, and mean age, although the age group 65+ was underrepresented. The sample included responders with different educational levels, professional status, household gross salary levels, composition of the household and marital status. Most responders had a child and were not affected by a chronic condition. The background characteristics of the sample compared with general population data (as reported by the Italian National Institute of Statistics) (Devlin et al., 2013; Stolk et al., 2019) are reported in Table 1.

3.1. Feasibility of videoconferencing

During the pilot phase, there were 25 (12.6%) technical feasibility issues reported out of 198 interviews performed. The most common problem was activation of the microphone (9, 4.5%), followed by wrong link sent to the interviewee or difficulties in accessing the link (5, 2.5%). The issues were solved with combination of VC software with phone audio (7, 3.5%), and use of Lifesize as alternative software (1, 0.5%). The frequency of the technical issues reported was stable throughout the data collection period, with each single problem impacting less than 5% of the interviews completed (Table 2).

The interviewers maintained each a rate of 9–10 interviews per week, which corresponded to a saturation of about 70% of their allotted slots, albeit this indicator increased over time. During the pilot phase, 12 dropouts were reported (6%), whilst there were 58 (6%) in total.

Table 1
Background characteristics of the sample and national adult population (2020).

	Sample n = 1182	General population (18+) n = 50,208,329
Age, mean (SD)	48.29 (16.06)	52.05
Age groups, n (%)		
18-24	109 (9.22%)	4,121,339 (8.21%)
25-34	166 (14.04%)	6,410,935 (12.77%)
35-44	200 (16.92%)	7,759,655 (15.45%)
45-54	251 (21.24%)	9,626,469 (19.18%)
55-64	211 (17.85%)	8,430,841 (16.79%)
65+	245 (20.72%)	13,859,090 (27.60%)
Gender, n (%)		
Male	575 (48.75%)	24,195,125 (48.19%)
Female	606 (51.27%)	26,013,204 (51.81%)
Other	1 (0.08%)	NA
Geographical distribution [a]		
North-West	317 (27.16%)	13,498,616 (26.88%)
North-East	225 (19.28%)	9,790,372 (19.50%)
Centre	230 (19.71%)	10,012,074 (19.95%)
South and Islands	395 (33.85%)	16,907,267 (33.67%)
Education [b]		
Elementary	1 (0.08%)	8263 (15.90%)
Middle inferior	76 (6.43%)	16,733 (32.19%)
High school	637 (53.89%)	19,038 (36.63%)
Academic degree	468 (39.59%)	7944 (15.28%)
Profession [c]		
Employed	487 (41.20%)	18,183,000 (36.21%)
Self-employed	150 (12.69%)	5,302,000 (10.56%)
Student	112 (9.48%)	2,202,487 (4.39%)
Pensioner	234 (19.8%)	16,000,000 (31.87%)
Unemployed	92 (7.78%)	NA
Housewife	96 (8.12%)	7,338,000 (14.61%)
Other	11 (0.93%)	1,182,842 (2.36%)
Annual gross salary of the household		
< € 14,000	93 (7.87%)	/
€ 14,000 - € 20,999	135 (11.42%)	/
€ 21,000 - € 27,999	168 (14.21%)	/
€ 28,000 - € 34,999	160 (13.54%)	/
€ 35,000 - € 41,999	159 (13.45%)	/
€ 42,000 - € 48,999	64 (5.41%)	/
€ 49,000 - € 55,999	90 (7.61%)	/
€ 56,000 - € 62,999	50 (4.23%)	/
€ 63,000 - € 69,999	40 (3.38%)	/
€ 70,000 - € 90,999	43 (3.64%)	/
> € 91,000	13 (1.10%)	/
Prefer not to answer	167 (14.13%)	/
Marital status [d]		
Single	350 (29.61%)	15,966,146 (31.80%)
Married or living with partner	727 (61.51%)	28,012,121 (55.80%)
Separated or divorced	78 (6.60%)	1,850,178 (3.68%)
Widow	27 (2.28%)	4,379,884 (8.72%)
Children [e]		
Yes	691 (58.46%)	8766 (62.13%)
No	491 (41.54%)	5343 (37.87%)
Household size [f]		
One	138 (11.67%)	8410 (32.85%)
Two	369 (31.22%)	7086 (27.69%)
Three	285 (24.11%)	4860 (18.99%)
Four	275 (23.27%)	3907 (15.27%)
Five or more	115 (9.73%)	1330 (5.20%)
Chronic conditions [g]		
No	721 (61.00%)	31,989 (26.08%)
Yes	461 (39.00%)	90,643 (73.92%)
Self-reported EQ VAS, mean (SD)	81.82 (13.52)	/
Self-reported health status (EQ-5D-5L)		
Mobility		
No Problems	1039 (87.9%)	/
Slight problems	103 (8.71%)	/
Moderate problems	30 (2.54%)	/
Severe problems	10 (0.85%)	/
Unable to	0 (0%)	/
Selfcare		
No Problems	1132 (95.77%)	/
Slight problems	37 (3.13%)	/

(continued on next page)

Table 1 (*continued*)

	Sample n = 1182	General population (18+) n = 50,208,329
Moderate problems	9 (0.76%)	/
Severe problems	4 (0.34%)	/
Unable to	0 (0.00%)	/
Usual Activities		
No Problems	1045 (88.41%)	/
Slight problems	108 (9.14%)	/
Moderate problems	24 (2.03%)	/
Severe problems	4 (0.34%)	/
Unable to	1 (0.08%)	/
Pain/Discomfort		
No	671 (56.77%)	/
Slight	383 (32.4%)	/
Moderate	113 (9.56%)	/
Severe	13 (1.10%)	/
Extreme	2 (0.17%)	/
Anxiety/Depression		
No	694 (58.71%)	/
Slight	372 (31.47%)	/
Moderate	106 (8.97%)	/
Severe	6 (0.51%)	/
Extreme	4 (0.34%)	/

Note.

[a] Data of geographical distribution were not recorded for 15 interviews, as these were collected by a previous panel company with which the study team terminated the contract.

[b] Education of the general public was calculated on a sample of 51,978 residents aged above 15.

[c] Occupational data are approximations of ISTAT data; the number of students was calculated as the sum of university students and those enrolled in the last year of high school (aged 18).

[d] ISTAT classification of separated is within the married category.

[e] Number of children in the general public is calculated on a sample of 14,109 couples where the woman is aged above 15.

[f] Number of people living in the same household is calculated on a sample of 25,593 families.

[g] Number of chronic conditions in the general public is calculated on a sample of 122,632 people aged 18+.

Table 2
Selected technical and organizational feasibility issues after pilot phase and overall.

Technical feasibility issues	Pilot phase (n = 198), n (%)	Full sample (n = 1182), n (%)
Activate microphone	9 (4.5%)	54 (4.6%)
Poor wi-fi connection	2 (1%)	20 (1.7%)
Issues with EQ-VT software	1 (0.5%)	18 (1.5%)
Problems with access to link	5 (2.5%)	15 (1.3%)
Other	1 (0.5%) [a]	15 (1.3%) [b]
Forgot appointment	1 (0.5%)	19 (1.6%)
Wrong link	5 (2.5%)	18 (1.5%)
Interview delayed	1 (0.5%)	11 (0.9%)

[a] Did not know the interview required a PC.

[b] Did not know the interview required a PC/Webcam not functioning/PC not functioning/Could not speak up in the room/Changed her mind/Emergency health issues/Somebody in the background suggesting the answers.

Overall, 146 (12%) interviews required rescheduling. The overall study response rate was 13.92%.[4] Of those that did not take part in the study, 33.94% refused to be interviewed and 52.14% were excluded as their quota was full.

After 198 interviews, 4 interviews (2%) were flagged for data quality reasons: 3 for a clear inconsistency in the TTO ratings for the state 55555, 1 due to less than 3 min spent on the wheelchair example. Overall, the median rate of flagged interviews per each interviewer was 0%. Mean interview time for each interviewer ranged from 41 to 62 min, with the time taken to complete a single TTO task ranging from 69 to 98 seconds. Visual inspection of TTO value distribution for each interviewer vis-à-vis the overall distribution did not reveal alarming clustering on critical values, with discrepancies always <5.4% (Appendix Table 2).

All interviewers completed the anonymous VC evaluation form. Overall, 9 out of 11 (82%) interviewers were strongly or somewhat satisfied with this MoA ("Overall, I am satisfied with videoconferencing interviews") and 10 out of 11 (91%) were strongly or somewhat satisfied with how the EQ-VT software interacted with the videoconferencing software ("Overall, I am satisfied with how the EQ-VT software and the videoconferencing software work together"). However, the vast majority (82%) did not find it easy to retrieve information for problem resolution directly in Zoom (Fig. 1). Qualitative analysis of open-ended questions revealed three emerging themes according to the interviewers' perspective, broadly synthesised as upsides, downsides, and recommended improvements. In the current high-income context of application, online interviews were considered positively due to time savings, flexibility, efficient scheduling, and a comfortable, familiar and safe setting for participants. Interviewers reported that they felt most interviewees were at ease with this MoA. Among possible limitations of this MoA, interviewers reported that engaging participants may be at times difficult, the illustration of the tasks may be more complex than in a face-to-face setting due to screensharing and that technical issues (internet connection/audio/video/outdated or poorly performing equipment from the responders) may complicate the interaction with the interviewees. Some interviewers reported that with this MoA interviewees may be more subject to distractions ("*sometimes you notice that they are checking their phone while you are talking*"), they may rush to answers, or, when bugs occur, information from that interview may be lost. Suggested improvements regarded focusing on the coordination routines among participants, panel company and interviewers, enhanced hardware and software related aspects, and exploring the technological capabilities and attitudes of interviewees prior to the conduct of the interview.

3.2. cTTO and DCE data

The 1182 interviews provided 11,820 cTTO responses and 8274 DCE responses. In cTTO, the mean value was 0.308, and the mean value for state 55555 was −0.485. The sample included 8 non-traders and 1 responder who gave the same value to all health states. Eighty-four health states were flagged with the feedback module.

Sixty-eight responders had an inconsistency for the state 55555 (5.75%), 65 of which (5.49%) remained after the feedback module. The aggregate data were well dispersed, with the highest clustering being reported for 1 (11.14%) and −1 (7.92%). The interviewer reporting the highest clustering at 1 registered 16.30% of observations on this value and the one with the highest clustering at −1 12.20%. Clustering at 0 were low (2.04%), with the interviewer registering the highest clustering reporting 4.80% observations on this value. Appendix Fig. 2 reports the cTTO response distribution for the aggregate data.

The mean time taken by responders to complete the 7 DCE choice tasks was 4.13 min. Unusual responses were rare, with only 2 responders always choosing option A, 14 responders alternating between option A and then B, and 9 responders alternating between option B and then A.

3.3. Modelling results

Table 3 reports the parameters estimates for the cTTO data, number of inconsistencies, non-monotonicities, mean absolute errors and ranking of dimensions by level 5 coefficients. The linear random intercept model (Model 1) showed monotonic coefficients, but mobility level

[4] Based on the second panel company data.

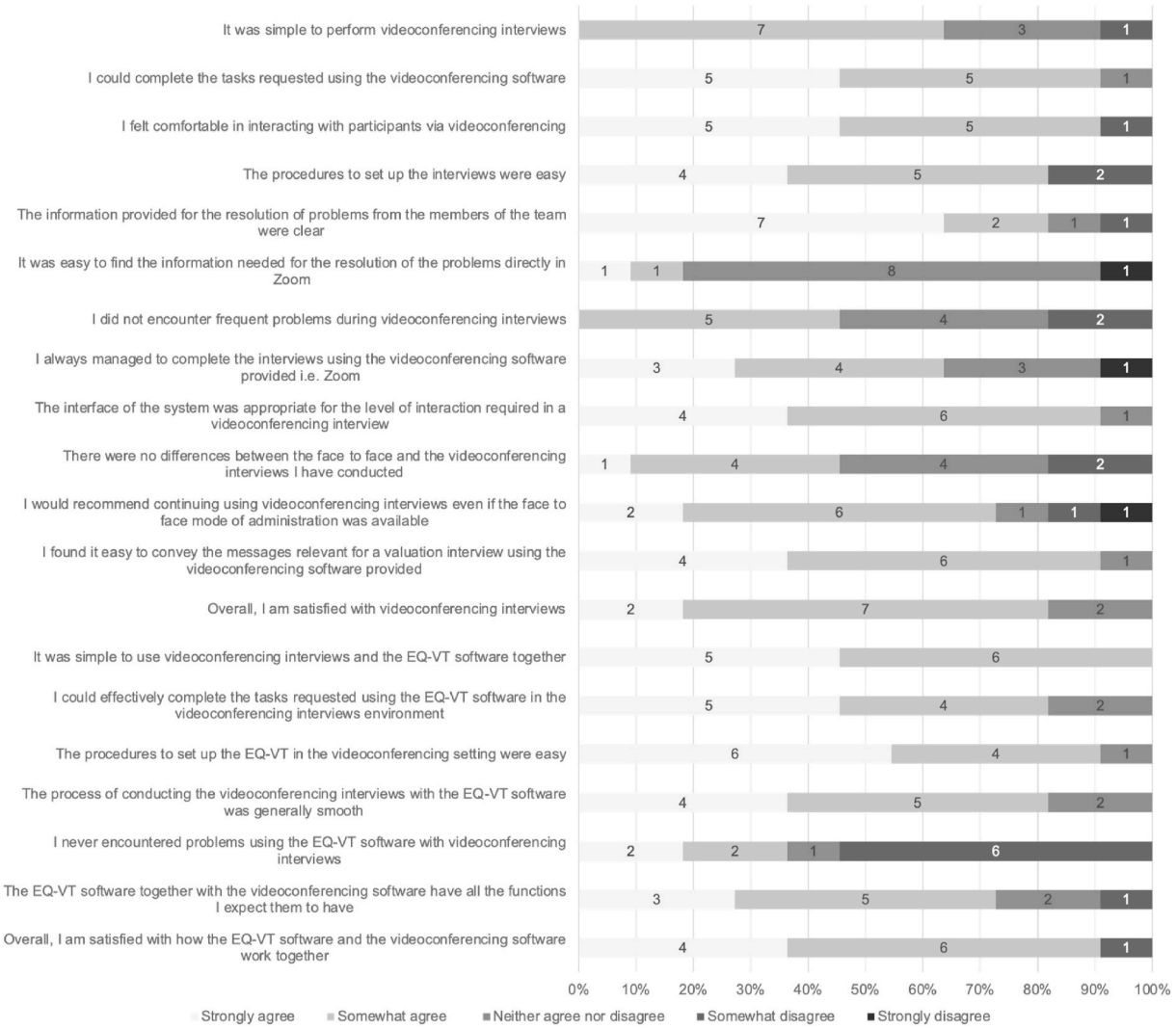

Fig. 1. Videoconferencing evaluation form results (n = 11).

2 was non-significant. The constant was non-significant and close to 1 and the Breusch–Pagan/Cook-Weisberg test suggested the presence of heteroscedasticity in the data. It was therefore decided to suppress the constant and account for the data heteroscedasticity (Model 2). Model 2 reported monotonically decreasing coefficients which were significant but presented an inconsistency for usual activity level 4. As left censoring i.e. censoring at −1 was observed in the data, a Tobit random effect model (Model 3) and a Tobit random effect heteroscedastic model (Model 4) were also specified. Model 3 was consistent, but mobility level 2 was not significant, while Model 4 had a significant constant and a non-significant parameter for mobility level 3.

Table 3 also reports the DCE coefficients rescaled using state 5555and the theta hybrid parameters. As it can be seen, all parameters were significant, but the generated tariff was inconsistent for anxiety/depression level 4 and usual activities level 4.

Table 4 presents the parameter estimates for the hybrid models, non-significant parameters, non-monotonicities, mean absolute errors and ranking of dimensions by level 5 coefficients. Model 6, a linear hybrid heteroscedastic model, reported a non-monotonicity for anxiety/depression level 4. Model 7, a hybrid Tobit heteroscedastic model was consistent, and the constant did not differ significantly from 1. It was therefore decided to constraint the constant at 1. The resulting hybrid Tobit heteroscedastic without constant model (Model 8) was monotonically consistent, and all parameters were significant. This model accounted for left censoring of the data i.e. censoring at −1, a feature considered valuable (Jensen et al., 2021) due to the construction of the task in the EQ-VT software (Devlin et al., 2013; Jensen et al., 2021), and for the heteroscedasticity observed in the data. As previously mentioned, this study reported 7.92% of observed values at −1, therefore justifying the choice of the Tobit model empirically. The model reported a mean absolute error of 0.051, mean absolute errors ranging between ±0.02, and predicted appropriately mild states i.e. level sum score below 7, with a mean absolute error for the mild states together of 0.011. This was selected as a preferred model for the derivation of an EQ-5D-5L value set for Italy. The two sensitivity analyses showed the model results were robust when testing alternative data exclusions. More specifically, the mean absolute error of the model increased when excluding the pilot, and only marginally improved when excluding the feedback module states but at the cost of higher mean absolute errors for the mild states.

Coefficients from Model 8 can be used for the calculation of the utility associated with different EQ-5D-5L health states. This is done by subtracting the coefficients associated with each of the dimension's levels from 1 i.e. full health. For example, EQ-5D-5L state 34212 would be calculated as: 1–0.064-0.216–0.050-0.000-0.044 = 0.626. The utility associated to state 55555 i.e. worst attainable health is −0.571, with 523 health states being WTD. Appendix Table 3 reports the STATA syntax for the computation of the EQ-5D-5L value set for Italy.

Table 3
Parameter estimates for cTTO and DCE models, insignificant parameters, non-monotonicities, mean absolute error and ranking of dimensions.

Dimensions and Levels	Model 1: cTTO (Linear, RE, BTW REG EST)		Model 2: cTTO (Linear, HET, NOCONS)		Model 3: cTTO (Tobit censored at −1, RE)		Model 4: cTTO (Tobit censored at −1, HET)		Model 5: DCE (conditional logit)			
	Coeff, SIG, SE		Coeff, SIG, SE		Coeff, SIG, SE		Coeff, SIG, SE		Coeff, SIG, SE		Rescaled state 55555	Rescaled Hybrid Theta
Mobility 2	0.007	0.012	0.045**	0.009	0.000	0.013	0.027**	0.010	0.315**	0.053	0.068	0.075
Mobility 3	0.038**	0.013	0.048**	0.016	0.030*	0.014	0.026	0.017	0.437**	0.065	0.094	0.103
Mobility 4	0.193**	0.014	0.206**	0.017	0.190**	0.015	0.196**	0.018	1.214**	0.066	0.262	0.287
Mobility 5	0.265**	0.013	0.260**	0.014	0.276**	0.014	0.277**	0.016	1.577**	0.075	0.340	0.373
Selfcare 2	0.040**	0.012	0.057**	0.008	0.038**	0.013	0.040**	0.010	0.148**	0.058	0.032	0.035
Selfcare 3	0.065**	0.014	0.070**	0.014	0.062**	0.015	0.070**	0.014	0.243**	0.062	0.052	0.057
Selfcare 4	0.211**	0.014	0.226**	0.016	0.214**	0.015	0.230**	0.017	0.918**	0.064	0.198	0.217
Selfcare 5	0.260**	0.012	0.249**	0.013	0.279**	0.013	0.273**	0.014	1.084**	0.062	0.234	0.256
Usual activities 2	0.041**	0.013	0.054**	0.009	0.039**	0.014	0.037**	0.010	0.267**	0.054	0.058	0.063
Usual activities 3	0.096**	0.014	0.088**	0.014	0.096**	0.015	0.075**	0.014	0.248**	0.061	0.053	0.059
Usual activities 4	0.230**	0.013	0.250**	0.015	0.235**	0.014	0.246**	0.016	0.932**	0.063	0.201	0.220
Usual activities 5	0.241**	0.012	0.233**	0.015	0.259**	0.013	0.254**	0.016	1.125**	0.065	0.243	0.266
Pain/Discomfort 2	0.041**	0.012	0.040**	0.007	0.038**	0.012	0.023**	0.008	0.352**	0.058	0.076	0.083
Pain/Discomfort 3	0.089**	0.014	0.080**	0.016	0.088**	0.015	0.073**	0.017	0.445**	0.062	0.096	0.105
Pain/Discomfort 4	0.366**	0.012	0.361**	0.015	0.382**	0.013	0.382**	0.016	1.518**	0.068	0.328	0.359
Pain/Discomfort 5	0.382**	0.013	0.414**	0.016	0.402**	0.014	0.430**	0.018	1.741**	0.071	0.376	0.412
Anxiety/Depression 2	0.037**	0.013	0.045**	0.008	0.035**	0.014	0.029**	0.009	0.240**	0.059	0.052	0.057
Anxiety/Depression 3	0.093**	0.015	0.098**	0.015	0.087**	0.016	0.075**	0.016	0.513**	0.062	0.111	0.121
Anxiety/Depression 4	0.286**	0.013	0.297**	0.015	0.290**	0.014	0.294**	0.016	1.427**	0.071	0.308	0.337
Anxiety/Depression 5	0.297**	0.013	0.314**	0.014	0.311**	0.013	0.338**	0.015	1.352**	0.068	0.292	0.320
Constant	0.974	0.016	NA		0.974		0.973**	0.018	NA		NA	NA
Non-significant parameters[a]	1		0		1		1		0			
Non-monotonicities	0		1		0		1		2			
Mean Absolute Error	0.040		0.042		0.044		0.048		N/R			
Mean Absolute Error (mild states)[b]	0.018		0.010		0.018		0.009		N/R			
Ranking of dimensions[c]	PD > AD > MO > SC > UA		PD > AD > MO > SC > UA		PD > AD > SC > MO > UA		PD > AD > MO > SC > UA		PD > MO > AD > UA > SC			

Legend: RE: random effects; HET: heteroscedasticity; BTW REG EST: between regression estimator; NOCONS: No constant; *p < 0.05; **p < 0.01; NA: Not applicable; Coeff: Coefficients; SIG: Significance; SE: Standard error.

[a] Excluding possible intercept.

[b] Mild states are defined as those with a level sum score below 7.

[c] Based on level 5. MO: mobility; SC: self-care; UA: usual activities; PD: pain/discomfort; AD: anxiety/depression.

4. Discussion

This study investigated the technical, organizational and protocol feasibility of VC interviews as a new MoA for the conduct of valuation studies. Having deemed the MoA feasible, we collected preferences from Italian members of the general public for a set of EQ-5D-5L health states using both cTTO and DCE and modelled them to derive a value set. The hybrid Tobit heteroscedastic model without constant, which made use of both cTTO and DCE data, was selected to generate the EQ-5D-5L Italian value set.

The frequency of technical issues recorded in this study was limited, with each single problem impacting less than 5% of the interviews completed. Most common issues related to using the computer microphone or to wi-fi connectivity. Interviewers were provided with a VC resolution guide and were instructed to phone participants in case of issues. Phoning participants was useful in limiting dropouts due to technical problems, although it imposed additional financial and ethical requirements which should be carefully considered by teams intending to use this tool. Of note, to account for problems with the VC audio some interviewers used phones as external microphones despite having been instructed not to do so. This use of phones may complicate the interaction between interviewers and interviewees which is at the base of the EQ-VT data quality and should be discouraged.

From an organizational point of view, a relatively high rate of weekly interviews per interviewers was maintained for a period of about 12 weeks of data collection using VC. This accounted for a rate of 6% dropouts and 12% rescheduled interviews. Organizational issues were sporadic and did not prevent completing the data collection. Overall, the majority of interviewers were somehow satisfied with how the EQ-VT software and the videoconferencing software work together. This suggests that the current study set up is feasible for the conduct of VC administered valuation studies of the EQ family of instruments in a high-income context. The possibility of employing a similar set up in a middle or low-income context still needs verification and might require different choices from the study teams and/or supply of adequate technological equipment. As the current set up involved multiple choices, e.g. market research company for the recruitment, online appointment scheduling assistant, phone reminders the day prior to the interview etc. and isolating the contribution of each of them to the success of a VC administered valuation study is difficult, teams intending to use this MoA should make explicit judgments and considerations of the likely impact of changes, e.g. postal recruitment, to the feasibility of the study.

Protocol feasibility was generally good. All 11 interviewers adhered to the QC protocol compliance indicators at the first round of data collection, as well as registering limited interviewers' effects from early rounds of the fieldwork. This resulted in low clustering of values which were generally comparable to those of other high-quality recent EQ-VT EQ-5D-5L valuation studies (Pickard et al., 2019; Purba et al., 2017b; Ludwig et al., 2018; Rencz et al., 2020), with no interviewer exhibiting substantially different patterns from the aggregate distribution. Therefore, it is possible to achieve comparable data quality to face-to-face administration with VC administered interviews. Yet, two of the current study team investigators are part of the EQ-VT support team. This may have ensured closer resemblance of the interviewers' training with the principal investigators' training provided remotely by the EQ Office and more frequent feedback to the interviewers, which in turn could have increased interviewers' motivation. Greater variability in

Table 4
Parameter estimates for hybrid models, insignificant parameters, non-monotonicities, mean absolute error, ranking of dimensions and selected model for value set derivation (gray).

Dimensions and Levels	Model 6: Hybrid (HET, NOCONS) Coeff, SIG, SE		Model 7: Hybrid (Tobit censored at -1, HET) Coeff, SIG, SE		Model 8 (selected model): Hybrid (Tobit censored at -1, HET, NOCONS) Coeff, SIG, SE		Sensitivity - Hybrid (Tobit censored at -1, HET, NOCONS) Excluding pilot n=984 Coeff, SIG, SE		Sensitivity - Hybrid (Tobit censored at -1, HET, NOCONS) Excluding feedback flagged states Coeff, SIG, SE	
Mobility 2	0.053**	0.007	0.054**	0.007	0.051**	0.007	0.046**	0.008	0.051**	0.007
Mobility 3	0.067**	0.010	0.065**	0.011	0.064**	0.011	0.059**	0.011	0.065**	0.011
Mobility 4	0.238**	0.010	0.246**	0.010	0.244**	0.010	0.242**	0.011	0.245**	0.010
Mobility 5	0.309**	0.009	0.329**	0.010	0.329**	0.010	0.320**	0.011	0.331**	0.010
Selfcare 2	0.046**	0.007	0.050**	0.007	0.046**	0.007	0.044**	0.007	0.045**	0.007
Selfcare 3	0.058**	0.009	0.058**	0.009	0.056**	0.009	0.057**	0.010	0.056**	0.009
Selfcare 4	0.209**	0.010	0.219**	0.010	0.216**	0.010	0.221**	0.011	0.217**	0.010
Selfcare 5	0.241**	0.009	0.258**	0.009	0.257**	0.009	0.261**	0.010	0.258**	0.009
Usual activities 2	0.051**	0.007	0.054**	0.007	0.050**	0.007	0.047**	0.007	0.050**	0.007
Usual activities 3	0.068**	0.009	0.066**	0.009	0.064**	0.009	0.071**	0.010	0.063**	0.009
Usual activities 4	0.217**	0.009	0.227**	0.010	0.225**	0.010	0.231**	0.011	0.226**	0.010
Usual activities 5	0.241**	0.009	0.257**	0.009	0.255**	0.009	0.256**	0.010	0.255**	0.009
Pain/ Discomfort 2	0.050**	0.006	0.052**	0.006	0.047**	0.006	0.047**	0.007	0.047**	0.006
Pain/ Discomfort 3	0.091**	0.010	0.090**	0.010	0.088**	0.010	0.083**	0.011	0.089**	0.010
Pain/ Discomfort 4	0.338**	0.010	0.353**	0.011	0.353**	0.011	0.344**	0.011	0.355**	0.011
Pain/ Discomfort 5	0.383**	0.010	0.410**	0.011	0.408**	0.011	0.411**	0.012	0.409**	0.011
Anxiety / Depression 2	0.046**	0.006	0.048**	0.006	0.044**	0.006	0.042**	0.007	0.044**	0.006
Anxiety / Depression 3	0.110**	0.009	0.111**	0.010	0.109**	0.010	0.108**	0.011	0.108**	0.010
Anxiety / Depression 4	0.308**	0.009	0.320**	0.010	0.318**	0.010	0.303**	0.010	0.318**	0.010
Anxiety / Depression 5	0.301**	0.009	0.324**	0.010	0.322**	0.010	0.306**	0.010	0.323**	0.010
Constant	NA		0.991	0.005	NA		NA		NA	
Non-significant parameters table[a]	0		0		0		0		0	
Non-monotonicities	1		0		0		0		0	
Mean Absolute Error	0.044		0.051		0.051		0.051		0.050	
Mean Absolute Error (mild states) b	0.012		0.014		0.011		0.013		0.012	
Ranking of dimensions [c]	PD>AD>MO>UA,SC		PD>AD>MO>SC>UA		PD>MO>AD>SC>UA		PD>MO>AD>SC>UA		PD>MO>AD>SC>UA	

Legend: RE: random effects; HET: heteroscedasticity; BTW REG EST: between regression estimator; NOCONS: No constant; * p<0.05; ** p<0.01; NA: Not applicable; a: Excluding possible intercept; b: Mild states are defined as those with a level sum score below 7; [c]: Based on level 5. MO: mobility; SC: self-care; UA: usual activities; PD: pain/discomfort; AD: anxiety/depression;

interviewers' performance might be expected for VC studies where team members have less experience with the EQ-VT procedures.

The current study recruited a sample representative of the Italian adult population for age, gender and geographical distribution, which was similar to the one of a previous valuation study conducted in Italy (Scalone et al., 2013). In the context of valuation studies using VC administration, the possibility of achieving sample representativeness is linked to the technological penetration within the country. Prior to testing this MoA the current study team assessed the technological penetration in Italy, finding that 85.3% of households have access to an internet connection (Household internet access, 2020). In countries with lower technological penetration, it may be more difficult to achieve sample representativeness. Even for countries with high technological penetration, other limits in achieving representativeness may exist. These may be context specific and could be related to the level to which market research companies are established as well as the methods they use for recruiting participants. Moreover, the online administration of VC interviews may be associated with barriers preventing older people, less educated people, diseased populations and less technically skilled from being sampled, with these barriers varying between countries. Selection bias and under sampling of some groups may be risks associated to online administration. Overall, our study provides additional empirical evidence on the advantages (e.g. increased flexibility of interview scheduling) and disadvantages (e.g. partial loss of visual cues, complications with building rapport) of VC MoA, in line with previous reports (Lipman, 2020). Whilst the approach is considered feasible and effective, robust experimental studies are awaited to investigate the role of possible selection effects and respondents' perspective on VC interviews (Peasgood, 2021; Rowen et al., 2021).

This study generated a monotonically consistent value set. In the selected model, for level 5 pain/discomfort registered the largest decrement (0.408), followed by mobility (0.329), anxiety/depression (0.322), self-care (0.257) and usual activities (0.255). This reflects the Italian public perception on the importance of symptoms and disabilities as described by the EQ-5D-5L, showing high comparability with the perception of the US population, which exhibited the same ranking for level 5 dimensions (Pickard et al., 2019). The predicted value for the worst attainable health on the EQ-5D-5L i.e. state 55555 was −0.571, which is similar to the one of other recent high income countries value sets such as the United States (−0.573) (Pickard et al., 2019) and the France ones (−0.525) (Andrade et al., 2020).

This study has some limitations. First, it collected preferences during the Covid-19 outbreak. There is currently no evidence available on the impact of the pandemic on preferences for EQ-5D-5L health states. It is therefore not possible to know whether individuals' value system remained unchanged during the covid pandemic, and, if not, whether preferences would return to as they were prior to the pandemic. Second, it generated a value set using VC interviews. On the one hand, VC interviews may be associated to MoA effects. Preliminary reports suggest this MoA outcomes are similar to those of face-to-face interviews

(Lipman, 2020; Estevez-Carillo et al., 2021; Dewilde, 2021), but in absence of controlled comparison studies the possibility of MoA effects cannot be entirely ruled out. The EuroQol Foundation has recently funded two studies (Peasgood, 2021; Rowen et al., 2021) which will further investigate the equivalence of VC and face-to-face MoA using an experimental design. Either way, the quality of data achieved in this study is comparable to the one of those administering interviews face-to-face (e.g. 18). On the other hand, the use of VC administered interviews requires participants to have internet access and to be computer literate. This may have caused a selection bias in our sample, particularly for older adults. Third, different model specifications from the standard ones were not explored for the derivation of this study value set, as the EQ-VT protocol optimizes for main effect models. Fourth, albeit this is frequent also in other valuation studies (e.g. 18), also in the current study the recruited sample was younger than the general population estimates. While some measures were taken to increase the proportion of older adults participating e.g. recruitment through local recruiters, perfect representativeness could not be achieved. Yet, use of local recruiters increased the proportion of older adults sampled. Future studies are invited to consider this as an option to improve representativeness. Fifth, the current study did not collect evidence on feasibility from the interviewee and in that it does not cover their experiences with this MoA. Finally, the interviewer evaluation form may have been prone to acquiescence bias, due to the presentation of the best response category on the left.

Despite this, the current study has also important strengths. It is the first study reporting evidence on the feasibility of conducting a full valuation study using VC administered interviews, demonstrating this could be a useful new data collection approach that ensures equivalent data quality to face-to-face interviews. It represents the first valuation for one of the EQ family of instruments that uses VC interviews for data collection. It estimated for the first time the EQ-5D-5L value set for Italy, using an international protocol which allows comparability with other international EQ-5D-5L studies.

5. Conclusion

VC appears feasible for the conduct of the EQ family of instruments valuation studies. The data collected were of equivalent quality to the ones of recent face-to-face studies. An EQ-5D-5L value set for Italy was generated. This can be used for the conduct of Italian CUAs. Furthermore, the availability of an EQ-5D-5L value set for Italy allows for international comparisons of HTA assessments across jurisdictions. The EuroQol Foundation recently funded two equivalence studies which will use an experimental design with random allocation to interviewers and VC or face to face administration. These studies will ensure homogeneity in the observed and unobserved characteristics between the two MoA, informing on the possible presence of MoA effects with VC administration.

Credit authors statement

Aureliano Paolo Finch: Funding acquisition, Conceptualization, Methodology, Formal analysis, Project administration and writing. Michela Meregaglia: data acquisition, Methodology, Project administration.Oriana Ciani: data acquisition, Funding acquisition, Formal analysis, writing. Bram Roudijk: Project administration, Formal analysis. Claudio Jommi: Funding acquisition, Project administration.

Grant support

20190460 / EuroQol Research Foundation.

Abbvie Italy, Fondazione SmithKline, Merck Sharp & Dohme Italy, Roche Italy, Sanofi Italy.

Declaration of competing interest

AF and BR work for the EuroQol Office and are members of the EuroQol Group. The other authors declare no conflict of interest in relation to this study.

Acknowledgements

The authors are grateful to the EuroQol Research Foundation, Abbvie Italy, Fondazione SmithKline, Merck Sharp & Dohme Italy, Roche Italy and Sanofi Italy for their unconditional grants. A special thanks goes to the strong team of interviewers, which included (in alphabetical order) Andrea Moro, Arianna Bertolani, Camilla Falivena, Carla Rognoni, Carlotta Variale, Francesco Malandrini, Giovanni Andrulli, Ludovica Borsoi, Rachele Freddi and Riccardo Consadori. We would like to thank the EuroQol scientific team leader Elly Stolk for the useful comments given during the planning, training, and execution of the study. We would like to thank Pepe Research, and in particular Elena Salvi, Paola Merulla and Marta Muchetti for the constant support in the collection of data and monitoring of feasibility indicators.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.socscimed.2021.114519.

References

Andrade, L.F., Ludwig, K., Goni, J.M.R., Oppe, M., de Pouvourville, G., 2020 Apr. A French Value Set for the EQ-5D-5L. PharmacoEconomics [cited 2021 Jul 2];38(4): 413–25. Available from: http://link.springer.com/10.1007/s40273-019-00876-4.

Augustovski, F., Rey-Ares, L., Irazola, V., Garay, O.U., Gianneo, O., Fernández, G., et al., 2016 Feb 1. An EQ-5D-5L Value Set Based on Uruguayan Population Preferences. Qual Life Res. https://doi.org/10.1007/s11136-015-1086-4 [cited 2021 Feb 17];25 (2):323–33. Available from:

Bertrand, C., Bertrand, C., Bourdeau, L., 2010. Research interviews by Skype: a new data collection method. In: Esteves, J. (Ed.), Proceedings from the 9th European Conference on Research Methods. IE Business School, Spain, pp. 70–79, 2010 Jul 1.

Bosa, I., Castelli, A., Castelli, M., Ciani, O., Compagni, A., Galizzi, M.M., et al., 2021 Mar 1. Response to COVID-19: Was Italy (Un)prepared [cited 2021 Jul 1]. Health Econ Policy Law, pp. 1–13. Available from: https://europepmc.org/articles/PMC7985656.

Brazier, J., Deverill, M., Green, C., 1999 Jul 1. A review of the use of health status measures in economic evaluation [cited 2021 Feb 16] J. Health Serv. Res. Pol. 4 (3), 174–184. https://doi.org/10.1177/135581969900400310. Available from:

Deakin, H., Wakefield, K., 2014 Oct 1. Skype Interviewing: Reflections of Two PhD Researchers. Qual Res. https://doi.org/10.1177/1468794113488126 [cited 2021 May 11];14(5):603–16. Available from:

Devlin, N., 2020. 'Preference-based Measure' Is Misleading – Can We Agree on Something Better? The Academic Health Economists' Blog [cited 2021 Feb 16]. Available from: https://aheblog.com/2020/08/12/preference-based-measure-is-misleading-can-we-agree-on-something-better/.

Devlin, N., Buckingham, K., Shah, K., Tsuchiya, A., Tilling, C., Wilkinson, G., et al., 2013. A Comparison of Alternative Variants of the Lead and Lag Time Tto. Health Econ [cited 2021 Jul 2];22(5):517–32. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/hec.2819.

Devlin, N., Brazier, J., Pickard, A.S., Stolk, E., 3L, 5L, 2018 Jun 1. What the L? A NICE conundrum [cited 2021 Feb 16] PharmacoEconomics 36 (6). https://doi.org/10.1007/s40273-018-0622-9, 637–40. Available from:

Dewilde, S., 2021. Conducting a Valuation Study during the Coronavirus Pandemic: the Effect of Online Video Interviews on QC Parameters and on TTO Value. Poster presented at the 2021. EuroQol Academy Meeting.

Edelaar-Peeters, Y., Stiggelbout, A.M., Van Den Hout, W.B., 2014 Jul 1. Qualitative and Quantitative Analysis of Interviewer Help Answering the Time Tradeoff. Med Decis Making. https://doi.org/10.1177/0272989X14524989 [cited 2021 May 11];34(5): 655–65. Available from:

Estevez-Carillo, A., Ramos-Goñi, J.M., Rand-Hendriksen, K., 2021. A Comparison of Face-To-Face and Video Conference-Based C-TTO Interviews. EuroQol Virtual Academy Meeting.

EuroQol Group, 1990a. EuroQol-A new facility for the measurement of health-related quality of life. Health Pol. 16 (3), 199–208. - Cerca con Google. [cited 2021 Feb 16]. Available from:https://www.google.com/search?q=EuroQol+Group+(1990).+EuroQol-A+new+facility+for+the+measurement+of+health-related+quality+of+life.+Health+Policy%2C+16(3)%2C+199%E2%80%93208&rlz=1C1CHBF_enNL878NL878&oq=EuroQol+Group+(1990).+EuroQol-A+new+facility+for+the+measurement+of+health-related+quality+of+life.+Health+Policy%2C+16(3)%

2C+199%E2%80%
93208&aqs=chrome..69i57.351j0j7&sourceid=chrome&ie=UTF-8.
EuroQol Group. EuroQol - a new facility for the measurement of health-related quality of life. Health Pol.. 1990b Dec 1 [cited 2021 Sep 6];16(3):199–208. Available from: https://www.sciencedirect.com/science/article/pii/0168851090904219.
Fattore on behalf of Associazione Italiana di Economia Sanitaria, 2009. Proposta di linee guida per la valutazione economica degli interventi sanitari. Aprile-Giugno Politiche Sanitarie 10 (2). Cerca con Google. [cited 2021 Feb 16]. Available from: https://www.google.com/search?q=Fattore+on+behalf+of+Associazione+Italiana+di+Economia+Sanitaria.+Proposta+di+linee+guida+per+la+valutazione+economica+degli+interventi+sanitari.+Politiche+Sanitarie%2C+10+(2)%2C+Aprile-Giugno+2009.&rlz=1C1CHBF_enNL878NL878&oq=Fattore+on+behalf+of+Associazione+Italiana+di+Economia+Sanitaria.+Proposta+di+linee+guida+per+la+valutazione+economica+degli+interventi+sanitari.+Politiche+Sanitarie%2C+10+(2)%2C+Aprile-Giugno+2009.&
aqs=chrome..69i57.257j0j9&sourceid=chrome&ie=UTF-8.
Favaretti, C., et al., 2009. Health Technology Assessment in Italy. International Journal of Technology Assessment in Healthcare. Cerca con Google [cited 2021 Feb 16]. Available from: https://www.google.com/search?q=Favaretti+C+et+al.%2C+Health+technology+assessment+in+Italy.+International+Journal+of+Technology+assessment+in+healthcare.+2009.&rlz=1C1CHBF_enNL878NL878&oq=Favaretti+C+et+al.%2C+Health+technology+assessment+in+Italy.+International+Journal+of+Technology+assessment+in+healthcare.+2009.&
aqs=chrome..69i57.229j0j4&sourceid=chrome&ie=UTF-8.
Finch, A.P., Brazier, J.E., Mukuria, C., 2018 May 1. What is the evidence for the performance of generic preference-based measures? A systematic overview of reviews [cited 2020 Jan 13] Eur. J. Health Econ. 19 (4), 557–570. https://doi.org/10.1007/s10198-017-0902-x. Available from:
Finch, A.P., Brazier, J.E., Mukuria, C., 2019 Jan 1. Selecting bolt-on dimensions for the EQ-5D: examining their contribution to health-related quality of life [cited 2021 Feb 16] Value Health 22 (1), 50–61. Available from: https://www.sciencedirect.com/science/article/pii/S1098301518323179.
Gandhi, M., Ang, M., Teo, K., Wong, C.W., Wei, Y.C.-H., Tan, R.L.-Y., et al., 2019 Aug 1. EQ-5D-5L Is More Responsive than EQ-5D-3L to Treatment Benefit of Cataract Surgery. Patient - Patient-Centered Outcomes Res. https://doi.org/10.1007/s40271-018-00354-7 [cited 2021 Feb 16];12(4):383–92. Available from:
Herdman, M., Gudex, C., Lloyd, A., Janssen, M.F., Kind, P., Parkin, D., et al., 2011 Dec 1. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L) [cited 2020 Jan 22] Qual Life Res 20 (10). https://doi.org/10.1007/s11136-011-9903-x, 1727–36. Available from:
Household internet access in Italy 2020. Statista. [cited 2021 May 25]. Available from: https://www.statista.com/statistics/377722/household-internet-access-in-italy/.
Janssen, M.F., Pickard, A.S., Golicki, D., Gudex, C., Niewada, M., Scalone, L., et al., 2013 Sep 1. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study [cited 2021 Feb 16] Qual. Life Res. 22 (7). https://doi.org/10.1007/s11136-012-0322-4, 1717–27. Available from:
Janssen, B.M.F., Oppe, M., Versteegh, M.M., Stolk, E.A., 2013 Jul. Introducing the composite time trade-off: a test of feasibility and face validity [cited 2021 Sep 6] Eur. J. Health Econ. 14 (S1), 5–13. Available from: http://link.springer.com/10.1007/s10198-013-0503-2.
Janssen, M.F., Bonsel, G.J., Luo, N., 2018 Jun 1. Is EQ-5D-5L better than EQ-5D-3L? A head-to-head comparison of descriptive systems and value sets from seven countries [cited 2021 Feb 16] PharmacoEconomics 36 (6). https://doi.org/10.1007/s40273-018-0623-8, 675–97. Available from:
Jensen, C.E., Sørensen, S.S., Gudex, C., Jensen, M.B., Pedersen, K.M., Ehlers, L.H., 2021 Feb 2. The Danish EQ-5D-5L Value Set: A Hybrid Model Using cTTO and DCE Data. Appl Health Econ Health Policy [cited 2021 May 21]; Available from: http://link.springer.com/10.1007/s40258-021-00639-3.
Jin, X., Al Sayah, F., Ohinmaa, A., Marshall, D.A., Johnson, J.A., 2019 Sep 1. Responsiveness of the EQ-5D-3L and EQ-5D-5L in Patients Following Total Hip or Knee Replacement. Qual Life Res. https://doi.org/10.1007/s11136-019-02200-1 [cited 2021 Jul 2];28(9):2409–17. Available from:
AIFA, 2020. Linee Guida per la compilazione del dossier a supporto della domanda di rimborsabilita e prezzo di un medicinale, p. 64.
Lipman, S., 2020 Jan 1. Time for Tele-TTO? Lessons Learned from Digital Interviewer-Assisted Time Trade-Off Data Collection. Patient Patient-Centered Outcomes Res [cited 2021 Feb 17]; Available from: https://repub.eur.nl/pub/132689/.
Ludwig, K., Graf von der Schulenburg, J.-M., Greiner, W., 2018 Jun. German Value Set for the EQ-5D-5L. PharmacoEconomics [cited 2021 Sep 6];36(6):663–74. Available from: http://link.springer.com/10.1007/s40273-018-0615-8.
Norman, R., King, M.T., Clarke, D., Viney, R., Cronin, P., Street, D., 2010 May 1. Does mode of administration matter? Comparison of online and face-to-face administration of a time trade-off task [cited 2021 May 11] Qual. Life Res. 19 (4), 499–508. https://doi.org/10.1007/s11136-010-9609-5. Available from:
ISTAT. Home English [internet]. Time series [cited 2021 May 4]. Available from: http://seriestoriche.istat.it/index.php?id=1&no_cache=1&L=1.
Oppe, M., Hout, B., 2017 Oct 17. The 'power' of Eliciting EQ-5D-5L Values: the Experimental Design of the EQ-VT.
Oppe, M., Devlin, N.J., Hout, B van, Krabbe, P.F.M., Charro, F de, 2014 Jun 1. A Program of Methodological Research to Arrive at the New International EQ-5D-5L Valuation Protocol. Value Health [cited 2021 May 4];17(4):445–53. Available from: https://www.valueinhealthjournal.com/article/S1098-3015(14)01829-4/abstract.
Oppe, M., Rand-Hendriksen, K., Shah, K., Ramos-Goñi, J.M., Luo, N., 2016 Oct. EuroQol protocols for time trade-off valuation of health outcomes. Pharmacoeconomics 34 (10), 993–1004.
Peasgood, T., 2021. Randomised Equivalence Study to Compare Online Interviews versus Face-To-Face Interviews to Value the EQ-5D-5L Using cTTO: Australian Arm. EuroQol Research Foundation.
Pickard, A.S., Law, E.H., Jiang, R., Pullenayegum, E., Shaw, J.W., Xie, F., et al., 2019 Aug 1. United States Valuation of EQ-5D-5L Health States Using an International Protocol. Value Health [cited 2021 Feb 17];22(8):931–41. Available from: https://www.sciencedirect.com/science/article/pii/S1098301519301433.
Purba, F.D., Hunfeld, J.A.M., Iskandarsyah, A., Fitriana, T.S., Sadarjoen, S.S., Passchier, J., et al., 2017 May. Employing Quality Control and Feedback to the EQ-5D-5L Valuation Protocol to Improve the Quality of Data Collection. Qual Life Res [cited 2021 May 11];26(5):1197–208. Available from: http://link.springer.com/10.1007/s11136-016-1445-9.
Purba, F.D., Hunfeld, J.A.M., Iskandarsyah, A., Fitriana, T.S., Sadarjoen, S.S., Ramos-Goñi, J.M., et al., 2017 Nov. The Indonesian EQ-5D-5L Value Set. PharmacoEconomics [cited 2021 Jul 2];35(11):1153–65. Available from: http://link.springer.com/10.1007/s40273-017-0538-9.
PSSUQ (Post-Study System Usability Questionnaire) - UIUX Trend [cited 2021 May 20]. Available from: https://uiuxtrend.com/pssuq-post-study-system-usability-questionnaire/.
Qian, X., Tan, R.L.-Y., Chuang, L.-H., Luo, N., 2019 Nov 25. Measurement Properties of Commonly Used Generic Preference-Based Measures in East and South-East Asia: A Systematic Review [cited 2020 Jan 14]. PharmacoEconomics. https://doi.org/10.1007/s40273-019-00854-w. Available from:
Ramos-Goñi, J.M., Oppe, M., Slaap, B., Busschbach, J.J.V., Stolk, E., 2017 Mar 1. Quality Control Process for EQ-5D-5L Valuation Studies. Value Health [cited 2021 May 5];20(3):466–73. Available from: https://www.valueinhealthjournal.com/article/S1098-3015(16)34066-9/abstract.
Ramos-Goñi, J.M., Pinto-Prades, J.L., Oppe, M., Cabasés, J.M., Serrano-Aguilar, P., Rivero-Arias, O., 2017 Jul. Valuation and Modeling of EQ-5D-5L Health States Using a Hybrid Approach. Med Care [cited 2021 Jul 2];55(7):e51–8. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7659432/.
Rencz, F., Brodszky, V., Gulácsi, L., Golicki, D., Ruzsa, G., Pickard, A.S., et al., 2020 Sep 1. Parallel Valuation of the EQ-5D-3L and EQ-5D-5L by Time Trade-Off in Hungary. Value Health [cited 2021 Sep 6];23(9):1235–45. Available from: https://www.valueinhealthjournal.com/article/S1098-3015(20)32117-3/abstract.
Richardson, J., Iezzi, A., Khan, M.A., 2015 Aug. Why do multi-attribute utility instruments produce different utilities: the relative importance of the descriptive systems, scale and 'micro-utility' effects. Qual Life Res Int J Qual Life Asp Treat Care Rehabil 24 (8), 2045–2053.
Rowen, D., Yang, Y., Mukuria, C., 2021. Pilot of UK EQ-5D-5L TTO Valuation to Assess Equivalence across Modes. EuroQol Research Foundation.
Scalone, L., Cortesi, P.A., Ciampichini, R., Belisari, A., D'Angiolella, L.S., Cesana, G., et al., 2013 Jul 1. Italian Population-Based Values of EQ-5D Health States. Value Health [cited 2021 Feb 17];16(5):814–22. Available from: https://www.sciencedirect.com/science/article/pii/S109830151301807X.
Shah, K.K., Lloyd, A., Oppe, M., Devlin, N.J., 2013 Jul. One-to-one versus group setting for conducting computer-assisted TTO studies: findings from pilot studies in England and The Netherlands [cited 2021 May 11] Eur. J. Health Econ. 14 (S1), 65–73. Available from: http://link.springer.com/10.1007/s10198-013-0509-9.
Shah, K.K., Ramos-Goñi, J.M., Kreimeier, S., Devlin, N.J., 2020 Sep. An exploration of methods for obtaining 0 = dead anchors for latent scale EQ-5D-Y values [cited 2021 Sep 8] Eur. J. Health Econ. 21 (7), 1091–1103. Available from: https://link.springer.com/10.1007/s10198-020-01205-9.
Stolk, E., Ludwig, K., Rand, K., van Hout, B., Ramos-Goñi, J.M., 2019 Jan. Overview, Update, and Lessons Learned from the International EQ-5D-5L Valuation Work: Version 2 of the EQ-5D-5L Valuation Protocol. Value Health [cited 2021 Jul 2];22(1):23–30. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1098301518361643.
Tarricone, R., Amatucci, F., Armeni, P., Banks, H., Borsoi, L., Callea, G., et al., 2021 May 1. Establishing a national HTA program for medical devices in Italy: overhauling a fragmented system to ensure value and equal access to new medical technologies [cited 2021 Jul 1] Health Policy 125 (5), 602–608. Available from: https://www.sciencedirect.com/science/article/pii/S0168851021000750.
Xie, F., Pickard, A.S., Krabbe, P.F.M., Revicki, D., Viney, R., Devlin, N., et al., 2015 Aug. A Checklist for Reporting Valuation Studies of Multi-Attribute Utility-Based Instruments (CREATE). PharmacoEconomics [cited 2021 Jul 2];33(8):867–77. Available from: http://link.springer.com/10.1007/s40273-015-0292-9.
Xie, F., Pullenayegum, E., Gaebel, K., Bansback, N., Bryan, S., Ohinmaa, A., et al., 2016 Jan. A time trade-off-derived value set of the EQ-5D-5L for Canada [cited 2021 Feb 17] Med Care 54 (1), 98–105. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4674140/.