Article:
Rowen, D. orcid.org/0000-0003-3018-5109, Mukuria, C., Bray, N. et al. (6 more authors) (2022) Assessing the comparative feasibility, acceptability and equivalence of videoconference interviews and face-to-face interviews using the time trade-off technique. Social Science & Medicine, 309. 115227. ISSN 0277-9536

https://doi.org/10.1016/j.socscimed.2022.115227

Social Science & Medicine 309 (2022) 115227

Contents lists available at ScienceDirect

Social Science & Medicine

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

Assessing the comparative feasibility, acceptability and equivalence of videoconference interviews and face-to-face interviews using the time trade-off technique

Donna Rowen [a,*], Clara Mukuria [a], Nathan Bray [b], Jill Carlton [a], Louise Longworth [c], David Meads [d], Ciaran O'Neill [e,f], Koonal Shah [g], Yaling Yang [h]

[a] School of Health and Related Research, University of Sheffield, UK
[b] School of Health Sciences, Bangor University, UK
[c] PHMR Consulting, UK
[d] Academic Unit of Health Economics, University of Leeds, UK
[e] Centre for Public Health, Queen's University, Belfast, Northern Ireland, UK
[f] National University of Ireland, Galway, Ireland
[g] National Institute for Health and Care Excellence, UK
[h] Nuffield Department of Primary Care Health Sciences, University of Oxford, UK

ARTICLE INFO

Keywords:
EQ-5D-5L
Preference elicitation
Time trade-off
Face to face interview
Videoconference interview

ABSTRACT

This study examines the comparative equivalence, feasibility and acceptability of video and in-person interviews in generating time trade-off (TTO) values. Sample participants in England were recruited using a blended approach of different methods and sampled based on age, gender, ethnicity, and index of multiple deprivation. Participants were randomly allocated to be interviewed either via video or in-person. Participants completed TTO tasks for the same block of 10 EQ-5D-5L health states using the EQ-VTv2 software. Feasibility, acceptability and equivalence was assessed across mode using: sample representativeness; participant understanding, engagement and feedback; participant preferred mode of interview; data quality; mean utility and distribution of values for each health state; and regression analyses assessing the impact of mode whilst controlling for participant characteristics. The video and in-person samples had statistically significant differences in ethnicity and income but were otherwise broadly similar. Video interviews generated marginally lower quality data across some criteria. Participant understanding and feedback was positive and similar across modes. TTO values were similar across modes; whilst mean in-person TTO values were lower for the more severe states, mode was insignificant in most regression analyses. There was no clear preference of mode across all participants, though the characteristics of participants preferring to be interviewed in-person or by video differs. Video and in-person TTO interviews were feasible, acceptable and generated good-quality data, though video interviews had lower quality data across some criteria. Whilst TTO values differed across modes for the more severe states, mode does not appear to be the cause. The study found that the characteristics of people preferring each mode differed, and this should be taken into account in future valuation studies since sample representativeness for some characteristics, and therefore potentially TTO values, could be affected by the choice of mode.

1. Introduction

Health preference research is required to adapt to the challenges of conducting research during the COVID-19 pandemic, and by exploring solutions to emerging challenges this may force innovation via sustainable changes to our research that make the most of modern technologies and consider the equitable inclusion of members of the general population (Kaur et al., 2021). The EQ-5D-5L (Herdman et al., 2011)

Abbreviations: DCE, Discrete choice experiment; EQ-VT, EuroQol Valuation Technology; IMD, Index of multiple deprivation; TTO, Time trade-off; VAS, Visual analogue scale.

* Corresponding author. School of Health and Related Research, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK.
E-mail address: d.rowen@sheffield.ac.uk (D. Rowen).

https://doi.org/10.1016/j.socscimed.2022.115227
Received 11 April 2022; Received in revised form 14 June 2022; Accepted 16 July 2022
Available online 5 August 2022

international valuation protocol involves the use of computer-assisted personal interviews conducted face-to-face in-person (in-person) using the EuroQol Valuation Technology (EQ-VT) system that involves the time trade-off (TTO) and (usually) discrete choice experiment (DCE) techniques (Oppe et al., 2014; Stolk et al., 2019). However, undertaking in-person interviews during the COVID-19 pandemic has presented considerable challenges worldwide due to national and local lockdowns, social distancing, work from home policies, and shielding of vulnerable participants. There may be reticence of people to participate in interviews undertaken in-person, and this may be more pronounced for some such as elderly or vulnerable people. Over the course of the pandemic, organisations and individuals have adapted to using technology such as videoconferencing to enable day-to-day activities including research to carry on. Continued widespread use of this technology is expected, and the way that research - and many other activities - are undertaken is unlikely to return to pre-pandemic practice. Valuation studies conducted by interview prior to the COVID-19 pandemic were conducted in-person but pandemic restrictions have necessitated the move to videoconferencing. Whilst in-person interviews for the TTO technique in particular are not necessarily viewed as the gold standard, they were most commonly used and thus movement away from this should be carefully considered.

In online video-conference (video) interviews, the interviewer and interviewee meet using videoconference software such as Zoom which uses both audio to enable them to converse and 'screen share' of the survey so both can see the tasks simultaneously. Video interviews have the advantage that they can be safely conducted with computer literate people in the context of pandemic restrictions. However, video administration of the interview may impact on preferences, understanding and engagement. In addition, it is common in studies using this technology that participants must be computer literate with access to a computer/tablet and internet connection. This means they may exclude some groups within the general population, unless provisions are made to provide participants with a computer/tablet and location with an internet connection, and even then some participants may not feel comfortable or may not be able to proficiently use the technology.

Studies conducted prior to the pandemic found that mode impacted on TTO values elicited via in-person interviews and a remote online survey (with no interviewer present) (Norman et al., 2010; Jiang et al., 2021) and data quality (Jiang et al., 2021). Recent studies have examined the feasibility of video interviews (Lipman, 2021; Finch et al., 2021; Estévez-Carrillo et al., 2022), but not their equivalence in generating TTO values that are comparable to those elicited in-person. One study assessed the quality and feasibility of undertaking video interviews using the EQ-VTv2 protocol, finding that this was both feasible and appropriate (Finch et al., 2021). One study switched data collection from in-person to video interviews due to the pandemic using a single interviewer with previous experience in the conduct of these studies, and found that video interviews were feasible (Lipman, 2021). Two EQ-5D-5L valuation studies that had started using in-person interviews and continued data collection using video interviews found comparable data quality for both modes (Estévez-Carrillo et al., 2022). These studies provide promising evidence around the feasibility of video interviews for eliciting TTO values, but cannot assess equivalence given the limited sample size who valued the same health states for video and in-person interviews.

This study's objectives are to examine the feasibility, acceptability and equivalence (in TTO values, distribution and data quality) of in-person and video interviews in generating TTO values. The survey findings will be informative for ongoing and future health state valuation studies.

2. Methods

Participants were interviewed either via video or in-person, using the EQ-VTv2 software involving only the TTO technique, with the same set of 10 health states for all participants. Ethical approval for the project was granted by the University of Sheffield School of Health and Related Research (ScHARR) Research Ethics Committee.

2.1. Recruitment, sampling, and randomisation

A blended recruitment approach was used to ensure a mix of people in Sheffield and Oxford were contacted to be invited to participate in the study: postal mailouts (targeted using postcode); adverts on social media (including Facebook and Twitter); adverts on websites for participants interested in undertaking interviews or surveys; flyers in cafes and shops; newspaper adverts; and snowballing via word of mouth. Interested participants completed a short online survey recording their age, gender, ethnicity, postcode, whether their health limits their day-to-day activities, and contact details. A multi-stage stratified quota approach was used for sampling, with quota groups for age and gender, and across the sample quotas for ethnicity and socio-economic group using the index of multiple deprivation (IMD) using postcode (but not within each quota group for age and gender). Quotas for age, gender and ethnicity were determined from the latest available census for England (2011), and IMD using quintiles (lowest quintile, middle three quintiles, highest quintile). The inclusion of participants with and without health problems was also ensured. Interested participants were allocated to be interviewed either via video or in-person, with the allocation at random initially and purposively to ensure sample representativeness towards the end of data collection. Interviews were conducted from September to December 2021, with the study ending recruitment due to increased transmission of COVID-19 in England.

2.2. Sample size

Varied and multiple analyses were used to compare a range of data by mode, meaning sample size cannot be determined using a single calculation. To inform sample size selection, a sample size was estimated for the comparison of the TTO data by mode. Assuming power 0.8, significance level 0.05, standard deviation 0.3, and 0.1 expected difference in TTO values requires 73 valuations per mode and in total 146 completed interviews, as used previously for equivalent research questions involving TTO (Rowen et al., 2012, 2015). However, a 0.3 standard deviation is lower than typically observed for severe states, and expected difference of 0.1 could be too large. To take both this and the varied analyses proposed into account, overall sample size was increased to 400 completed interviews, with the expectation that up to 60 pilot interviews may be excluded.

2.3. Selection of health states

Ten EQ-5D-5L health states were selected from the standard 86 health states used in the EQ-VT protocol using a single TTO block that consisted of plausible health states (Yang et al., 2019) covering the severity range. EQ-5D-5L health states are reported as a 5 digit number generated using the response level (1 = no problems through to 5 = extreme problems) to the dimensions (mobility, self-care, usual activities, pain/discomfort, anxiety/depression) (Herdman et al., 2011).

2.4. Interviewer training and monitoring

Six interviewers were trained via videoconference and in-person. Each interviewer undertook approximately 10 pilot interviews which were retained for interviewers with protocol compliance. All interviewers received further individualised training following the pilot. Data quality was monitored throughout using the EQ-VTv2 quality control process (Ramos-Goñi et al., 2017).

2.5. The interviews

Informed consent was taken prior to the start of the interview. In the interview the participant answered socio-demographic questions and completed the EQ-5D-5L for their own health today. The composite TTO technique was explained using warm-up tasks that comprised consideration of being in a wheelchair, a state either better than or worse than being in a wheelchair (to ensure explanation of TTO tasks for states better than and worse than dead), and three EQ-5D-5L states (mild, moderate and difficult to imagine). The participant completed TTO tasks for 10 EQ-5D-5L health states, then the "feedback module" where the implied ranking of the 10 health states generated using their TTO responses were presented, and participants were asked to highlight any they would now reconsider (though they do not undertake further TTO tasks for any highlighted states). Finally, participants were asked about their understanding of the tasks, what they thought of the interview and mode, and completed additional health and socio-demographic questions. Participants were thanked for their participation with £50, offered as a choice of two different vouchers.

Interviews used the digital EQ-VTv2 software controlled only by the interviewer. In-person interviews were conducted in accessible meeting rooms with social distancing (2 metre distance between participant and interviewer) and 2 screens, one each for participant and interviewer. Video interviews were conducted on Zoom, and participants were instructed to keep their cameras on to enable greater interaction and interviewer monitoring of understanding and engagement.

Prior to launching the interviews, a pre-pilot was undertaken with a convenience sample of 12 participants (6 video and 6 in-person interviews) recruited from Sheffield and Oxford. These interviews assessed the acceptability and sufficiency of the adaptations made (due to COVID-19 and mode) and assessed whether the questions added to the end of the interview were appropriate and correctly interpreted and understood. Minor changes were made iteratively throughout these interviews.

2.6. Public involvement

Public involvement via the Patient Involvement Programme (PIP) at National Institute for Health and Care Excellence (NICE) was undertaken prior to data collection to inform: information provided to participants prior to the interview including postal mailout and questions to determine socioeconomic group; questions asked to assess the appropriateness of mode; thank you payments; and a priori criteria specified to determine equivalence. Public involvement involved two video meetings each with two researchers and two public involvement participants, involving four public involvement participants in total.

2.7. Analysis

Feasibility, acceptability and equivalence was assessed across mode. Feasibility was assessed using participant understanding, engagement and feedback and data quality. Acceptability was assessed using participant preferences about how they would prefer to be interviewed and participant feedback. Equivalence was assessed using: sample representativeness; data quality; mean TTO value and distribution of TTO values for each health state; and regression analyses assessing the impact of mode whilst controlling for the sociodemographic characteristics of participants. The analyses focus on statistical significance at the 1% or 5% level though are also reported at the 10% level.

2.7.1. Sample by mode and sample representativeness

Sample representativeness was assessed by comparison to the 2011 UK census. The samples were compared by mode, using a two-sample test of proportions where appropriate (for greater information this is reported for characteristics of interest e.g. employed, retired, rather than by question e.g. employment status).

2.7.2. Participant understanding, engagement and feedback and preference of mode

Responses to questions assessing understanding, engagement, feedback on the interview and mode preference were reported by mode of interview conducted and compared using the Chi-squared test. The questions on understanding and engagement are included within the standard EQ-VT system, and the additional questions seeking feedback on the interview and mode preference were developed bespoke for this study (and piloted as discussed in section 2.5). Reasons for the preference to be interviewed by video or in-person are tabulated for participants with each preference by mode and overall. Sociodemographic characteristics of participants by their preferred choice of mode were compared using the Chi-squared test, though these were not statistically compared to those with no preference of mode since these participants would be willing to participate in studies using either mode.

2.7.3. Data quality

Data quality was assessed by mode using a range of criteria, including criteria reported in the recent quality assurance of the EQ-5D-5L value set for England (Hernandez Alava et al., 2020) and the EQ-VT quality control process (Ramos-Goñi et al., 2017). Criteria were based around:

- Clustering of TTO values at values that may indicate that the participant is exiting the task quickly or is unwilling to consider exact preferences (−1, −0.5, 0, 0.5 or 1);
- Avoidance of negative values since this can indicate an unwillingness to state any health state is below dead and constrains all values to zero and above;
- Few distinct TTO values and only integer TTO values which can indicate a lack of distinction between health states and unwillingness to report exact preferences;
- Logically inconsistent responses, where health states that are better or the same across each dimension, in comparison to another health state, are given a lower (worse) TTO value. This can indicate a lack of understanding and engagement which is stronger if the logical inconsistency occurs with health states with a larger difference in severity (for example states 11212 and 55555);
- Unclear preferences for the worst state, where TTO value for the worst state is not at the lowest or uniquely lowest value for an individual, which may indicate a lack of understanding or lack of distinction between health states;
- Participant understanding and effort as perceived and reported by the interviewer, where low understanding or engagement may indicate poorer data quality.

Interviewer comparability by mode and between interviewers was also examined by looking at EQ-VT quality control reports and protocol compliance criteria, clustering of TTO values, TTO value distribution, mean TTO task duration, mean interview duration and mean feedback module duration. Data quality was also assessed by mode by comparing the number of moves in the TTO task to reach the TTO value, clustering effects per state, and the percentage of times each state was flagged in the feedback module.

2.7.4. Mean TTO value and distribution of TTO values for each health state

The distribution of TTO values per state is compared by mode. TTO values per state are summarised by mode using mean, standard deviation (SD), median, lower quartile and upper quartile, and the differences in mean and median by mode are reported. Two tests are reported to compare TTO values by mode: 1) Wilcoxon's rank-sum test of difference in means; and 2) equality of standard deviations using Levene's robust test statistic (that is robust when the distribution is non-normal). The analysis was repeated when removing TTO values flagged by the feedback module.

Table 1
The sample, by mode.

		Video interview N = 224 (%)	In-person interview N = 136 (%)	P-value (two-sample test of proportions)	Target for representativeness based on UK Census 2011 (England percentages)
Gender	Male	44.6	44.1	0.926	
	Female	55.4	55.2	0.971	
	Prefer not to say	0	0.7	0.541	
Mean age (SD)		44.5 (16.25)	47.67 (15.43)		
Female	Age 18 to 40	24.1	21.3	0.541	18.8
	Age 41 to 64	21.4	25.7	0.347	19.6
	Age 65 and over	9.8	8.1	0.588	12.7
Male	Age 18 to 40	18.8	15.4	0.411	19.2
	Age 41 to 64	17.4	20.6	0.449	19.1
	Age 65 and over	8.5	8.1	0.894	10.7
Prefer not to say	Prefer not to say	0	0.7	0.541	
Ethnicity	White	73.2	86.0	0.005	80%
	White British	63.8	76.5	0.012	
	White non-British	9.4	9.6	0.950	
	Asian/Asian British	6.7	8.1	0.619	
	Black/African/Caribbean/Black British	11.2	1.5	0.001	
	Mixed/Multiple ethnic groups	4.5	1.5	0.126	
	Other ethnic group	3.1	0.7	0.132	
	Prefer not to say	1.3	2.2	0.515	
IMD Index of multiple deprivation	1 or 2 (most deprived quintile)	9.4	8.8	0.848	20
	3,4,5,6,7,8	56.7	55.1	0.767	60
	9,10 (least deprived quintile)	30.4	33.8	0.501	20
	Prefer not to say	3.6	2.2	0.455	
Marital status	Single	26.8	32.4	0.256	
	Married/Partner	61.2	60.3	0.865	
	Separated	3.1	1.5	0.345	
	Divorced	5.8	4.4	0.564	
	Widowed	3.1	0.7	0.132	
	Prefer not to say	0	0.7	0.210	
Employment status	In employment or self-employment	54.9	60.3	0.316	
	Retired	18.8	19.1	0.944	
	Housework	1.8	0.7	0.387	
	Student	10.7	11.0	0.929	
	Seeking work	1.8	0	0.116	
	Unemployed	2.7	1.5	0.456	
	Long-term sick	5.4	1.5	0.065	
	Carer or volunteer	2.2	2.2	1.0	
	Prefer not to say	1.8	3.7	0.265	
Educated after age of 16	Yes	95.5	94.9	0.795	
	No	4.0	5.2	0.593	
	Prefer not to say	0.5	0		
Degree or equivalent professional qualification	Yes	84.6	87.6	0.430	
	No	15.0	11.6	0.363	
	Prefer not to say	0.5	0.8	0.724	
Home ownership/rental	Own your home outright, or with a mortgage	57.1	65.4	0.119	
	Rent from a local authority	14.0	7.4	0.057	
	Rent from the private sector	23.2	22.1	0.809	
	Other	4.5	4.4	0.965	
	Prefer not to say	1.4	0.7	0.543	
Annual household income before tax (including benefits)	Up to £5199	2.2	2.2	1.0	
	£5200 and up to £10,399	7.6	2.2	0.030	
	£10,400 and up to £15,599	6.3	4.4	0.447	
	£15,600 and up to £20,799	9.4	3.7	0.043	
	£20,800 and up to £25,999	9.8	10.3	0.878	
	£26,000 and up to £31,199	9.8	10.3	0.878	
	£31,200 and up to £36,399	6.7	8.1	0.619	
	£36,400 and up to £51,999	13.4	14.7	0.730	
	£52,000 and above	21.4	27.9	0.161	
	Prefer not to say	13.4	16.2	0.464	
Day-to-day activities limited because of a health problem or disability	Yes, limited a lot	10.3	6.6	0.232	
	Yes, limited a little	22.3	27.9	0.231	
	No	67.4	65.4	0.696	
Parent or guardian for a child or children aged under 18 years	Yes	28.6	22.1	0.174	
	No	71.0	77.9	0.150	
	Prefer not to say	0.5	0	0.409	
Have experienced ill health	In you, yourself	35.3	30.2	0.320	
	In your family	77.7	84.6	0.111	
	In caring for others	46.9	46.3	0.912	
	Video interview N = 222, %		In-person interview N = 138, %		

(*continued on next page*)

Table 1 (*continued*)

				Video interview N = 224 (%)		In-person interview N = 136 (%)		P-value (two-sample test of proportions)	Target for representativeness based on UK Census 2011 (England percentages)	
EQ-5D-5L level	**Mobility**	**Self-care**	**Usual activities**	**Pain/ discomfort**	**Anxiety depression**	**Mobility**	**Self-care**	**Usual activities**	**Pain/ discomfort**	**Anxiety depression**
1	80.4	89.7	70.5	50.9	48.7	81.6	91.2	77.9	51.5	55.9
2	9.4	4.9	17.0	32.6	33.9	12.5	4.4	12.5	36.0	32.4
3	7.1	4.0	7.6	9.8	11.6	2.9	3.7	6.6	9.6	8.8
4	3.1	1.3	4.5	4.5	3.6	2.2	0	2.9	1.5	2.2
5	0	0	0.5	2.2	2.2	0.7	0.7	0	1.5	0.7
EQ-5D-5L: Mean (SD)	1.33 (0.74)	1.17 (0.55)	1.47 (0.85)	1.75 (0.96)	1.77 (0.95)	1.28 (0.70)	1.15 (0.54)	1.35 (0.73)	1.65 (0.83)	1.60 (0.80)
EQ-5D-5L: *t*-test P-value (two-sample test of means)						0.519	0.704	0.145	0.360	0.077
VAS: Mean (SD)	77.87 (17.51)					79.96 (14.91)				
VAS: Median	80					82.5				
VAS: Interquartile range	70, 90					75, 90				

Notes: P-values have been generated using a two-sample test of proportions by mode.

2.7.5. Regression analyses

Regression analyses assess whether mode impacts on TTO values after controlling for other factors that may impact on the values. The TTO data has censoring at −1, since participants cannot express a lower TTO value than −1 for any health state, though they may wish to do so, and repeated observations per participant, as each participant values all 10 TTO health states. Taking this into account, regression analysis was undertaken with TTO value as the dependent variable using a random effects Tobit model with censoring at −1 (and explored using a random effects generalised least squares model and heteroscedastic Tobit model, reported in Supplementary Materials). Four model specifications were explored: Model 1 assesses the impact of mode and controls for health state through health state dummies; Model 2 also controls for socio-demographic characteristics of participants; Model 3 also controls for interviewer effects; and Model 4 also includes interactions for the health state and in-person mode. OLS and Tobit regressions were also estimated for each state separately for these model specifications, to determine whether the impact of mode differed across the different states, to determine whether the impact of mode differed by interviewer, and also assessing the impact of education by mode (reported in Supplementary Materials). Analyses were conducted in Stata version 15.

3. Results

3.1. Sample by mode and sample representativeness

Forty interviews were excluded following the pilot due to protocol non-compliance. The video sample (n = 224) is considerably larger than the in-person sample (n = 136). In comparison to England population norms, both the video and in-person samples have larger proportions of females aged 18 to 64 and smaller proportions of males and females aged 65 and over (see Table 1). The video and in-person samples are very similar for the sampled characteristics of age, gender and IMD, though for ethnicity there are statistically significant differences, with the video interview sample having a smaller proportion of White British individuals and a larger proportion of Black/African/Caribbean/Black British individuals. The samples are statistically significantly different for income, with individuals with lower income levels more highly represented in the video sample. Both samples are highly educated. The video sample has worse health than the in-person sample (though this is not statistically significant).

3.2. Participant understanding, engagement and feedback and preference of mode

There are no statistically significant differences in participant-reported understanding by mode, where the majority of participants reported that they agreed that the questions they were asked were easy to understand (see Table 2). Overall feedback on the interviews was positive. Statistically significant differences by mode were observed for "I got bored during the interview" and "I could hear the interviewer clearly" where in-person interviews performed more favourably. Nearly 17% of video participants agreed that they had technical issues (e.g. internet connection, sound). Nearly 20% of in-person participants agreed that they would have preferred not to travel to the interview, though less than 1% disagreed that they felt safe travelling to the interview, and no participants disagreed that they felt comfortable and safe during the interview.

There was no clear preference of mode across individuals (see Table 2), though the most common response was to prefer the mode by which they were interviewed. The sociodemographic characteristics of participants preferring to be interviewed in-person or by video interview had statistically significant differences by age, gender, employment status, home ownership (regarding rental from private sector), parent/guardian status and ethnicity (see Supplementary Materials Table A1). The most common selected reasons for preferring a video interview were convenience and that there was no time of travel involved, and in contrast the most common reason selected for preferring an in-person interview was that they would feel most at ease being interviewed that way.

3.3. Data quality

Overall, video interviews have lower quality across some metrics, but this impact is small and statistically significant only in two instances, where the participant gives utility of zero in at least 2 health states with no utility below zero, and interviewer reported it was doubtful whether the participant understood the exercises (see Table 3). Key points to note, though any differences are not statistically significant:

- The data does not have large clustering of TTO values (−1, −0.5, 0, 0.5 or 1) for either mode;
- The proportion of negative values is similar across modes but a larger proportion of participants in the video interviews do not report any negative TTO values (31% vs 22%);
- Neither mode has logically inconsistent TTO values between mild states and the worst state (<1%), and the proportion of logical

Table 2
Feedback questions relating to ease of understanding, ease of task and mode.

	Video interviews, n = 224 (%)					In-person interviews, n = 136 (%)					P-value (Chi squared test
	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree	
It was easy to understand the questions I was asked	51.8	42.0	3.1	2.7	0.5	50.0	35.3	7.4	6.6	0.7	0.112
I found it difficult to decide on the exact points where Life A and Life B were about the same	35.3	50.0	8.9	4.9	0.9	37.5	48.5	5.2	8.1	0.7	0.526
I found it easy to tell the difference between the lives I was asked to think about	21.4	48.7	12.5	14.7	2.7	19.9	46.3	12.5	16.9	4.4	0.875
I found the interview straightforward	49.1	42.0	4.0	4.9	0	50.0	35.3	10.3	3.7	0.7	0.087
I think the interview worked well	51.3	46.0	2.2	0.5	0	53.7	39.0	5.2	2.2	0	0.135
I got bored during the interview	0.5	0.9	4.0	54.0	40.6	0	1.5	4.4	37.5	56.6	0.035
I think the interviewer was clear and approachable	64.3	33.9	1.3	0.5	0	74.3	25.7	0	0	0	0.138
I could hear the interviewer clearly	61.2	33.5	3.6	1.8	0	77.9	20.6	1.5	0	0	0.007
I think the visual display during the interview was appropriate	57.6	36.2	4.9	0.9	0.5	59.6	34.6	2.2	2.9	0.7	0.411
I had technical issues e.g. internet connection, sound[1]	1.8	15.1	5.4	30.7	47.0						
I think the instructions during the interview were clear	57.1	40.6	2.2	0	0	63.2	33.8	1.5	1.5	0	0.164
I felt comfortable and safe during the interview	71.4	28.6	0	0	0	80.2	19.1	0.7	0	0	0.063
I felt comfortable and safe in the interview location	75.5	23.7	0.9	0	0	79.4	19.9	0.7	0	0	0.687
I would have preferred not to travel to the interview (on site only)[2]						3.0	16.4	23.1	32.1	25.4	
I felt safe travelling to the interview (on site only)[2]						61.9	34.3	3.0	0.8	0	
I think having my own laptop screen worked well (on site only)[2]						35.8	18.7	44.0	1.5	0	
"**If you had a choice would you have chosen to be interviewed online or in-person?**"											<0.001
Online interview	47.8					20.6					
On site in-person interview	15.2					50.0					
Don't mind	37.1					29.4					
Prefer not to say	0					0					
"**Why would you prefer to be interviewed online?" (if above answer was online, participants selected all that apply)**	**N (N = 107)**		**%**			**N (N = 28)**		**%**			**Overall %, N = 135**
More convenient	95		88.8			21		75.0			85.9
Would feel most at ease being interviewed that way	17		15.9			3		10.7			14.8
No cost of travel involved	22		20.6			5		17.9			20.0
No time of travel involved	42		39.3			14		50.0			41.5
Concerns about COVID-19	26		24.3			3		10.7			21.5
Other	7[3]		6.5			3[a]		10.7			7.4
"**Why would you prefer to be interviewed onsite?" (if above answer was onsite, participants selected all that apply)**	**N (N = 34)**		**%**			**N (N = 68)**		**%**			**Overall %, N = 102**
More convenient	7		20.6			7		10.3			13.7
Would feel most at ease being interviewed that way	22		64.7			39		57.4			59.8
Concerns about the technology involved	5		14.7			6		8.8			10.8
Don't like to spend too much time online	1		2.9			5		7.4			5.9

(*continued on next page*)

Table 2 (*continued*)

	Video interviews, n = 224 (%)					In-person interviews, n = 136 (%)					P-value (Chi squared test
	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree	
Concerns over privacy	0		0.0			1	1.5				1.0
Other	9[b]		26.5			24[c]	35.3				32.4

Notes: [1]n = 166.[2]n = 134.[3] Participants specified the following reasons: childcare (n = 2); caring responsibilities; environmental reasons; less anxiety and overwhelming and more honest responses; parking issues; travel problems.

[a] Participants specified the following reasons: anxiety leaving the house; easier for cognition; parking issues.

[b] Participants specified the following reasons: face-to-face contact; more interesting experience; prefer speaking in-person; read body language; miss out on body language and interactions and prefer face-to-face; eye contact difficult on a screen; quiet; more communication and body language; prefer face-to-face.

[c] Participants specified the following reasons: easier to/can ask questions (n = 3); better to talk face-to-face; easier understanding; like meeting and interacting with people; nice to get out of the house; more interactive and enjoyable; more interesting; more personal; prefer face-to-face due to constant online interviews; prefer to discuss in-person; works from home, nice to have in-person meetings; would not have understood online – more help understanding through face-to-face; easy to communicate face-to-face; human interaction; get out of the house and more pleasant, human contact; like meeting people in-person; might not get good explanation online; more congenial; more neutral environment; social contact; technology involved; ability to talk through the questions with the interviewer.

inconsistencies against all potential logical inconsistencies is small across both modes;

- The proportion of participants where the value for the worst state is not at the uniquely lowest value is larger for video interviews (55% vs 45%) and this remains unaffected by the feedback module. The proportion of participants where TTO value for the worst state is not at the lowest value is much lower and more similar by mode (17% vs 11%, and 8% vs 7% after the feedback module);
- Participant understanding and effort as perceived and reported by the interviewer was higher for in-person interviews.

Mean TTO task duration, mean interview duration and mean feedback module duration and compliance with EQ-VT quality criteria was similar across modes (Supplementary Materials Figures A3 to A5; Table A3). The number of TTO moves taken to reach indifference in the TTO tasks did not significantly differ by mode with the exception of one state (23152) where fewer moves were used in the video interviews (Supplementary Materials Table A6). There was evidence of some interviewer effects, where data differed across interviewers, though these analyses do not control or adjust for the sociodemographic characteristics of participants interviewed by each interviewer and mode (Supplementary Materials, Figure A6).

3.4. Mean TTO value and distribution of TTO values for each health state

TTO values by health state (mean, SD, median, lower and upper quartiles) are largely similar across the two modes (Table 4 and Fig. 1). However, TTO values are generally higher for the in-person interviews for the milder states and lower for the moderate and severe states. The difference in mean values is larger than 0.05 for four states, which are all moderate and severe states (21345, 23152, 43514, 55555), with the largest difference for the worst state (55555). Mean TTO values are statistically significantly different across the two modes for states 21111, 55555, and standard deviation is statistically significantly different for state 55555.

3.5. Regression analyses

The dummy variable for mode (in_person dummy variable) is not statistically significant (Table 5, Supplementary Materials Table A10 and A11). The only statistically significant sociodemographic characteristic impacting on TTO values across all models is gender, where males have higher TTO values. Being married is weakly statistically significant across most models. Interviewer effects are statistically significant for two interviewers. Dummy variables for the moderate and severe states are statistically significant, whereas dummy variables for states 11212 and 12112 are not (in comparison to the baseline of state 21111). Interactions between the state and mode are negative and statistically significant for states 21345 and state 55555.

Regressions estimated separately for each health state (Supplementary Materials Tables A12-A18) find that the dummy variable for mode is only (weakly) statistically significant for state 55555. No variables are statistically significant for all health states, and the only variables that are statistically significant for a larger proportion of the health states are those reflecting interviewer effects, and interviewer effects for a single mode are rarely significant. The impact of not being highly educated (no degree dummy variable) and its interaction with mode (in-person no degree) is statistically significant (at the 5% or 10% level) for some health states (across all model specifications it is significant for 4 of 10 health states).

4. Discussion

The results demonstrate that video and in-person TTO interviews are feasible and acceptable. Both generated good-quality data, though video interviews performed lower across some criteria. Whilst TTO values differed across the modes for the more severe states, thus questioning equivalence for more severe states by mode, mode does not appear to be the cause when controlling for sociodemographic characteristics and interviewer effects. The sample is highly educated across both modes, and it is possible that data quality and TTO values by mode could differ in a less educated sample. The study results suggest that TTO data collection using either mode is feasible, acceptable and will not in itself affect TTO values, though the choice of mode is likely to impact on the sample of participants willing to be interviewed via that mode, which in turn may potentially impact on the TTO values. The study included only TTO and not DCE, since this study was undertaken to inform the UK valuation of the EQ-5D-5L where only TTO will be used.

The imbalance in sample size for video and in-person interviews was due to both a higher acceptance rate of those invited to video interview (54.1%) than in-person interview (38.0%) (see supplementary materials Table A20), and a higher percentage allocated to video interview (53.6%) as in the final stage of data collection in-person interviews were halted due to the COVID-19 situation and data collection was then stopped. In the interests of learning from this study for a large future valuation study, and concerns about reopening data collection for predominantly in-person interviews at a different point in the pandemic which could in itself potentially affect responses, we did not re-open data collection once the COVID-19 situation eased. Whilst the sample imbalance was not intended, and an equal number in each mode would

Table 3
Data quality, by mode.

Problematic responder type	Video interviews N = 224, %	In-person interviews N = 136, %	P-value (two-sample test of proportions)
Individual values all 10 health states with the same value	0.4	0.7	0.211
Individual reports utility of −1, −0.5, 0, 0.5 or 1 for all 10 health states	1.8	2.2	0.790
Proportion of values at 1	12.0	14.3	0.528
Proportion of values at 0.5	7.1	4.8	0.381
Proportion of values at 0	4.0	3.3	0.734
Proportion of values at −0.5	3.1	2.9	0.914
Proportion of values at −1	6.6	10.1	0.233
Proportion of negative values	27.1	29.6	0.609
Individual reports no negative value	31.3	22.1	0.059
Individual gives utility of zero in at least 2 health states with no utility below zero	5.8	1.5	0.048
Individual reports fewer than 5 distinct values	14.7	14.0	0.855
Individual gives only integer values (No use of half-year increments in TTO)	24.6	25.7	0.815
Individual reports mild health states (11212, 12111, 21111) with same or lower value of 55555	0.9	0.7	0.839
Individual with any inconsistencies between the logical ordering of health states and the TTO valuation (where logically better state is valued lower) [see Table A5 in Supplementary Materials] excluding inconsistencies with state 55555	4.0	2.9	0.586
Number of logical inconsistencies across all interviews	89/5600 potential inconsistencies, 1.6%	27/3400 potential inconsistencies, 0.8%	0.516
Value for 55555 is not at the uniquely lowest value given by the individual	55.4	44.9	0.053
Value for 55555 is not at the lowest value given by the individual	16.5	11.0	0.150
State 11212 is valued strictly lower than 55555	0	0	
State 12112 is valued strictly lower than 55555	0	0	
State 21111 is valued strictly lower than 55555	0	0	
State 21345 is valued strictly lower than 55555	4.0	3.7	0.887
State 23152 is valued strictly lower than 55555	3.1	1.5	0.345
State 34244 is valued strictly lower than 55555	4.5	1.5	0.126
State 43514 is valued strictly lower than 55,555	2.7	0.7	0.183
State 44553 is valued strictly lower than 55555	7.6	5.1	0.356
State 55424 is valued strictly lower than 55555	6.7	2.2	0.058
Value for 55555 is 0.5 higher than the value for one or more other states	3.6	1.5	0.242
Value for 55555 is not at the uniquely lowest value given by the individual after the feedback module	54.9	44.9	0.066
Value for 55555 is not at the lowest value given by the individual after the feedback module	8.4	6.6	0.535
Value for 55555 is 0.5 higher than the value for one or more other states after the feedback module	2.2	0	0.082
Interviewer reporting of understanding and effort			
Understanding			
Understood and performed exercises easily	68.8%	75.7%	0.160
Some problems but seemed to understand the exercises in the end	25.5%	22.8%	0.564
Doubtful whether the respondent understood the exercises	5.8%	1.5%	0.048
Effort and concentration			
Concentrated very hard and put a great deal of effort into it	70.5%	79.4%	0.062
Concentrated fairly hard and put some effort into it	22.8%	16.2%	0.131
Didn't concentrate very hard and put little effort into it	4.0%	1.5%	0.182
Concentrated at the beginning but lost interest/ concentration before reaching the end	2.7%	2.9%	0.911

Table 4
Average (Mean & Median) TTO values and distribution of values for each health state, by mode (all responses, none excluded using feedback module).

	Video interviews					In-person interviews					Difference		Tests of significance	
	Mean	SD	Median	Lower quartile	Upper quartile	Mean	SD	Median	Lower quartile	Upper quartile	Difference in mean TTO values	Difference in median TTO values	Difference in mean TTO values[1]	Equality of standard deviations[2]
11212	0.903	0.130	0.95	0.9	1.0	0.911	0.150	0.95	0.9	1.0	0.008	0	0.052	0.454
12112	0.892	0.141	0.95	0.9	1.0	0.901	0.157	0.95	0.9	1.0	0.009	0	0.051	0.496
21111	0.934	0.115	0.95	0.9	1.0	0.950	0.103	1	0.95	1.0	0.016	−0.05	0.030	0.252
21345	0.084	0.614	0.3	−0.6	0.5	−0.024	0.604	0.2	−0.6	0.5	−0.107	0.1	0.064	0.997
23152	0.247	0.596	0.4	0.0	0.7	0.177	0.605	0.35	−0.25	0.65	−0.070	0.05	0.187	0.708
34244	−0.007	0.597	0.175	−0.6	0.5	0.004	0.581	0.2	−0.525	0.5	0.011	−0.025	0.960	0.454
43514	0.196	0.575	0.375	−0.025	0.6	0.114	0.578	0.3	−0.2	0.5	−0.081	0.075	0.125	0.868
44553	−0.204	0.596	−0.1	−0.8	0.3	−0.231	0.550	−0.075	−0.775	−0.075	−0.027	−0.025	0.641	0.100
55424	−0.061	0.606	0.075	−0.6	0.425	−0.083	0.594	0.05	−0.7	0.475	−0.021	0.025	0.634	0.571
55555	−0.439	0.532	−0.6	−0.95	0.025	−0.558	0.472	−0.7	−1	−0.05	−0.119	0.1	0.031	0.018

Notes: [1]P-value testing difference in means (Wilcoxon's rank-sum test). [2]P-value testing equality of standard deviations using Levene's robust test statistic.

be preferable, meaningful results can still be obtained across the range of different analyses conducted. Each interviewer conducted interviews in both modes throughout, with the exception of the final stage of data collection where in-person interviews were halted, meaning that the learning curve for interviewers would not be expected to differ across modes.

Taking into account all analyses, state 55555 is the only state where TTO value may differ by mode. Whilst this could be due to differential preferences around whether a health state is valued as worse than dead by mode, this only has an impact for the TTO value of the worst state across all analyses. It is difficult to understand or reason why a difference in values for only the worst state may occur by mode, since other severe health states were also valued. Further research assessing whether the value for state 55555 differs by mode is encouraged, and further exploration of whether this issue is apparent for other more severe states would also be beneficial.

The study findings are consistent with the recent study conducting (only) video interviews for TTO, that also found that video interviews were feasible and acceptable (Finch et al., 2021). The two studies (Lipman, 2021; Estévez-Carrillo et al., 2022) comparing video and in-person interviews, where the different modes were collected at different time points due to the COVID-19 pandemic, did not find significant impacts of mode on data quality, though note that the criteria that was used differed (one study focussed on EQ-VT quality control (Estévez-Carrillo et al., 2022) and the other on quality assurance (Lipman, 2021)) and sample size was small for at least one of the modes (n = 60/61 (Estévez-Carrillo et al., 2022) and n = 36 (Lipman, 2021)).

Minor logical inconsistencies in TTO values are to be expected in any TTO valuation study where the order of health states is randomized. These are expected because in early tasks participants may not fully understand the severity of states and there are learning effects, and fatigue effects in later tasks. Arguably more important indicators for data quality are high proportions of responses at values where the TTO tasks can be quickly concluded (1, 0, 0.5, −0.5) and logical inconsistencies in TTO values for states that clearly differ in severity (for example 21111 vs 55555). Using these criteria, the study has good data quality and better data quality than the current EQ-5D-5L value set for England (Devlin et al., 2018). For example 8.4% of the original sample gave a state with a 1 digit difference to the best state the same or lower value than the worst state (Hernandez Alava et al., 2020) in the value set for England (Devlin et al., 2018), in comparison to 0.4% and 0.7% by video and in-person interviews in this study. Whilst the value for the worst state is not at the uniquely lowest value for approximately half of the participants (55% video interviews vs 45% for in-person interviews), this was higher for the EQ-5D-5L UK value set at 66.8% in the original sample. It is also not logically inconsistent if respondents are not willing to sacrifice a different number of life years to avoid the worst state in comparison to the other state, and this can reflect a genuine preference.

Whilst video interviews have lower quality across some criteria than in-person interviews, these differences are generally small, are only significant across two criteria of the large number assessed, and are not at a level that indicates concerns in data quality. It should also be noted that these analyses do not control for sociodemographic differences of the samples across modes.

Interviewer effects are apparent in the data despite the quality control process used. Whilst this does not indicate poor quality data, it does indicate variability in data across different interviewers. This emphasises the need for good quality interviewer training, careful monitoring of data collection and informative feedback to interviewers during the study.

The study was conducted during the COVID-19 pandemic, and it is unknown whether the preferred mode of people may differ post-pandemic. Whilst TTO preferences may be different during the pandemic to pre-pandemic (Webb et al., 2021), any impact is likely to have affected responses in both modes. Due to the pandemic the in-person interviews used separate screens for the interviewer and participant in the in-person interviews, which differs to standard administration of the EQ-VT system using a single device. It is unknown whether this impacted on understanding or engagement of participants or interaction with their interviewer.

This study has several limitations. The key limitation is that the study recruited participants for a study where they were allocated to either a video or in-person interview. Therefore, the participants are unlikely to be fully representative of people willing to be interviewed by a single mode, or fully representative of the wider UK population. The study was also conducted during the COVID-19 pandemic where there may have been greater reluctance to participate in in-person interviews. It is unknown how or whether this would impact on the results or the acceptance rate of those invited to interview that was much higher for those invited to be interviewed via video interview (54.1% versus 38.0% for in-person interview).

The study has underrepresentation of the lowest socioeconomic group and less highly educated individuals. Whilst the regression results suggest that this is unlikely to impact on the TTO values, this is based on a sample with low representation of these groups and hence further research is encouraged. The requirement that participants expressing an interest in this study had to be able to complete the interview via video, with internet access and a computer/tablet, may have led to the underrepresentation. Indeed, a recent EQ-5D-5L valuation in Italy conducted via video interviews had twice the proportion of participants with a degree in comparison to the general population (39.6% vs 15.3%) (Finch et al., 2021).

Remaining limitations are the difference in the size of sample across the two modes and that the study was conducted in a single country, England, in two cities. Ongoing research in Australia is repeating the study with an equal sample size in each group (Peasgood et al., 2021)

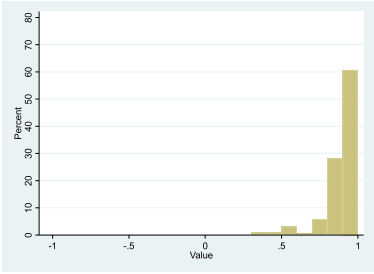
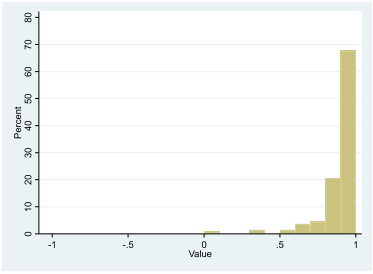
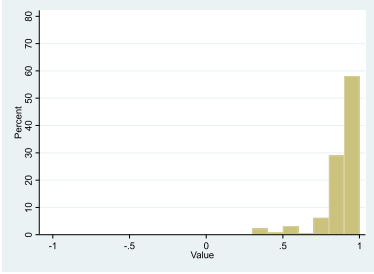
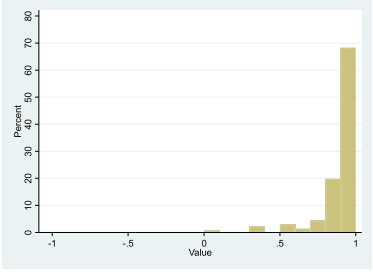
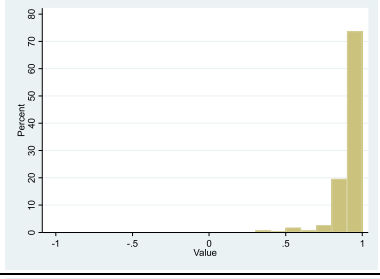
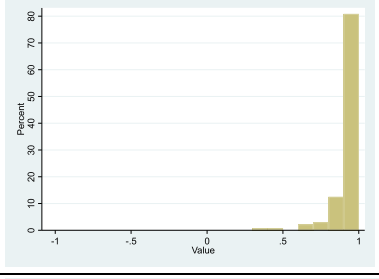
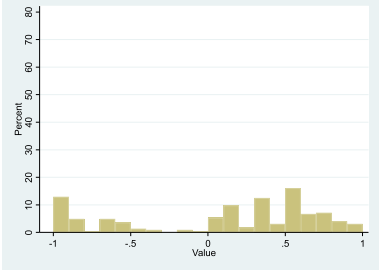
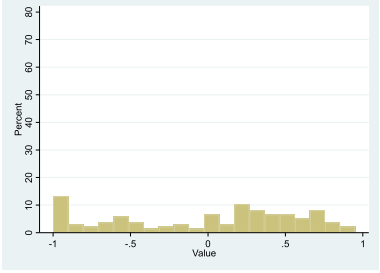
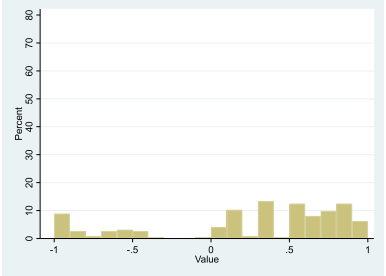
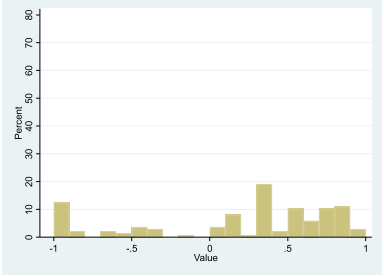
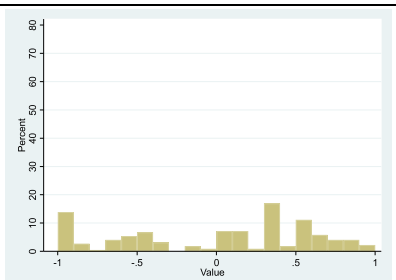
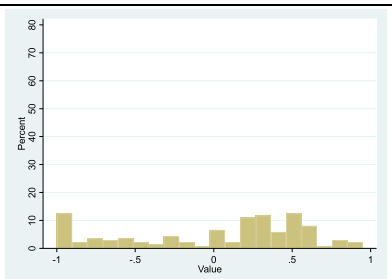
State	Video interviews	In-person interviews
11212		
12112		
21111		
21345		
23152		
34244		

Fig. 1. Histograms of TTO values by state and mode (all responses, none excluded using feedback module).

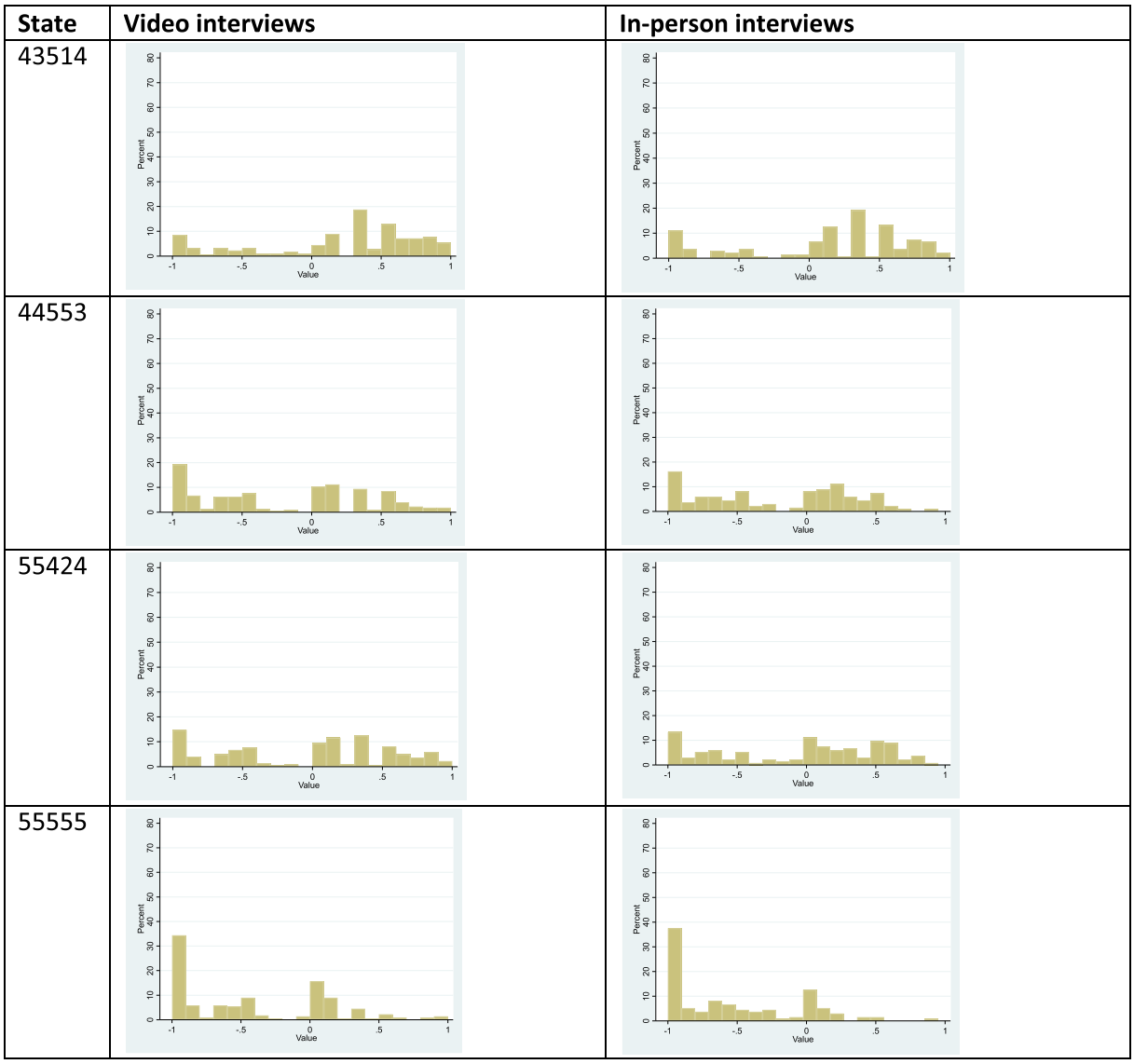


Fig. 1. (*continued*).

and will provide results for a different country, albeit another high income country, but one with different lived experience of the COVID-19 pandemic.

5. Conclusion

The study found that TTO data collection using video interviews or in-person interviews is feasible, acceptable and will not in itself affect TTO values. However, the choice of a single mode is likely to impact on the sample of participants willing to be interviewed, and this may potentially impact on the TTO values. Our results therefore suggest that offering a choice of mode in future TTO valuation studies, where feasible, will enable greater accessibility and greater inclusivity of participants into the study. There is no reason to expect this will have a substantial impact on data quality, and a small potential impact on data quality is arguably warranted due to the increased accessibility and inclusivity. There is also an efficiency argument, since it may be quicker and easier to recruit participants when offering both modes, and video interviews do not require room hire, or travel time and costs. We recommend that future TTO valuation studies consider offering both in-person and online video interviews, since these studies can be influential in their use to inform public policy, and therefore require representation of the diverse sociodemographic characteristics of the general population.

Credit author statement

Donna Rowen: Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Funding acquisition. Clara Mukuria: Conceptualization, Methodology, Investigation, Writing – review & editing, Funding acquisition. Nathan Bray: Conceptualization, Methodology, Writing – review & editing, Funding acquisition. Jill Carlton: Conceptualization, Methodology, Writing – review & editing, Funding acquisition. Louise Longworth: Conceptualization, Methodology, Writing – review & editing, Funding acquisition. David Meads: Conceptualization, Methodology, Writing – review & editing, Funding acquisition. Ciaran O'Neill: Conceptualization, Methodology, Writing – review & editing, Funding acquisition. Koonal Shah: Conceptualization, Methodology, Writing – review & editing, Funding acquisition. Yaling Yang: Conceptualization, Methodology, Investigation, Writing – original draft, Writing – review & editing, Funding acquisition.

Table 5
Random effects Tobit regression of all TTO observations.

Variables	Model 1	Model 2	Model 3	Model 4
State 11212	−0.034	−0.034	−0.034	−0.031
	(0.258)	(0.258)	(0.258)	(0.414)
State 12112	−0.045	−0.045	−0.045	−0.043
	(0.133)	(0.133)	(0.133)	(0.262)
State 21345	−0.917***	−0.917***	−0.917***	−0.866***
	(0.000)	(0.000)	(0.000)	(0.000)
State 23152	−0.732***	−0.732***	−0.732***	−0.695***
	(0.000)	(0.000)	(0.000)	(0.000)
State 34244	−0.963***	−0.963***	−0.963***	−0.959***
	(0.000)	(0.000)	(0.000)	(0.000)
State 43514	−0.789***	−0.789***	−0.789***	−0.749***
	(0.000)	(0.000)	(0.000)	(0.000)
State 44553	−1.185***	−1.185***	−1.185***	−1.162***
	(0.000)	(0.000)	(0.000)	(0.000)
State 55424	−1.033***	−1.033***	−1.033***	−1.015***
	(0.000)	(0.000)	(0.000)	(0.000)
State 55555	−1.496***	−1.496***	−1.496***	−1.430***
	(0.000)	(0.000)	(0.000)	(0.000)
Interview conducted in-person	−0.048	−0.039	−0.048	
	(0.224)	(0.321)	(0.225)	
State 11212 * interview conducted in person				0.009
				(0.881)
State 12112 * interview conducted in person				0.010
				(0.863)
State 21111 * interview conducted in person				0.016
				(0.773)
State 21345 * interview conducted in person				−0.121**
				(0.036)
State 23152 * interview conducted in person				−0.081
				(0.157)
State 34244 * interview conducted in person				0.004
				(0.939)
State 43514 * interview conducted in person				−0.089
				(0.122)
State 44553 * interview conducted in person				−0.045
				(0.433)
State 55424 * interview conducted in person				−0.031
				(0.586)
State 55555 * interview conducted in person				−0.160***
				(0.007)
Male		0.087**	0.079**	0.079**
		(0.024)	(0.039)	(0.039)
Aged 41 to 64		−0.083*	−0.065	−0.065
		(0.062)	(0.141)	(0.141)
Aged 65 and over		−0.011	−0.017	−0.017
		(0.867)	(0.788)	(0.787)
Ethnicity of White British or White Other		0.019	0.024	0.024
		(0.706)	(0.631)	(0.630)
Day-to-day activities are limited a lot because of a health problem or disability		0.081	0.068	0.068
		(0.265)	(0.343)	(0.344)
Have experienced illness in you, yourself		−0.049	−0.052	−0.052
		(0.266)	(0.228)	(0.227)
Parent/guardian for a child or children aged under 18 years		0.059	0.072	0.072
		(0.204)	(0.123)	(0.123)
In employment or self-employment		0.025	0.022	0.022
		(0.572)	(0.623)	(0.628)
Married		0.076*	0.070*	0.070*
		(0.072)	(0.090)	(0.090)
IMD most deprived quintile		−0.077	−0.041	−0.041
		(0.273)	(0.566)	(0.566)
IMD least deprived quintile		−0.071	−0.057	−0.057
		(0.103)	(0.185)	(0.185)
Rent (home) from a local authority		−0.042	−0.034	−0.034
		(0.523)	(0.598)	(0.598)
No degree or equivalent professional qualification		0.069	0.060	0.061
		(0.235)	(0.292)	(0.291)
Interviewer 1			−0.138**	−0.137**
			(0.036)	(0.036)
Interviewer 2			−0.029	−0.029
			(0.621)	(0.623)
Interviewer 3			−0.183***	−0.183***
			(0.002)	(0.002)
Interviewer 4			−0.110*	−0.110*
			(0.077)	(0.077)
Interviewer 6			−0.030	−0.030
			(0.682)	(0.685)
Constant	0.959***	0.897***	0.971***	0.947***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	3600	3600	3600	3600
Number of participants	360	360	360	360

Notes: P-values in parentheses: ***p < 0.01, **p < 0.05, *p < 0.1. Baseline for model 4: Interviewed by video interview, health state 21111, female, aged 18–40 years, ethnicity is not white, no or some limitations in daily activities as a result of health, no experience of illness in yourself, not a parent or guardian of child aged under 18 years, not employed, not married, IMD middle three quintiles, do not rent a house from a local authority, have degree or equivalent professional qualification, interviewed by interviewer 5.

Financial support

This project received funding from the EuroQol Research Foundation. The views expressed by the authors in the publication do not necessarily reflect the views of the EuroQol Foundation. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

Acknowledgements

We would like to thank our interviewers and all our research participants for taking part in our study. We would also like to thank Donna Davis and Claire Meadows for providing project management, and Nancy Devlin, Tessa Peasgood, Kim Rand, Bernhard Slaap and Elly Stolk for comments on previous versions of the analysis.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.socscimed.2022.115227.

References

Devlin, N.J., et al., 2018. Valuing health-related quality of life: an EQ-5D-5L value set for England. Health Econ. 27 (1), 7–22.

Estévez-Carrillo, A., et al., 2022. Exploring the Comparability of Face-To-Face versus Video Conference-Based Composite Time Trade-Off Interviews: Insights from EQ-5d-Y-3l Valuation Studies in Belgium and Spain. The Patient - Patient-Centered Outcomes Research.

Finch, A.P., et al., 2021. An EQ-5D-5L value set for Italy using videoconferencing interviews and feasibility of a new mode of administration. Soc. Sci. Med., 114519

Herdman, M., et al., 2011. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual. Life Res. 20, 1727–1736.

Hernandez Alava, M., Pudney, S., Wailoo, A., 2020. The EQ-5D-5L value set for England: findings of a quality assurance program. Value Health 23 (5), 642–648.

Jiang, R., et al., 2021. Comparison of online and face-to-face valuation of the EQ-5D-5L using composite time trade-off. Qual. Life Res. 30 (5), 1433–1444.

Kaur, M.N., et al., 2021. Transforming Challenges into Opportunities: Conducting Health Preference Research during the COVID-19 Pandemic and beyond. Qual Life Res, pp. 1–8.

Lipman, S.A., 2021. Time for tele-TTO? Lessons learned from digital interviewer-assisted time trade-off data collection. Patient 14 (5), 459–469.

Norman, R., et al., 2010. Does mode of administration matter? Comparison of online and face-to-face administration of a time trade-off task. Qual. Life Res. 19 (4), 499–508.

Oppe, M., et al., 2014. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. Value Health 17 (4), 445–453.

Peasgood, T., et al., *Randomised Equivalence Study to Compare Online Interviews versus Face-To-Face Interviews to Value the EQ-5D-5L Using cTTO: Australian Arm.* 2021, Funded by The EuroQol Research Foundation.

Ramos-Goñi, J.M., et al., 2017. Quality control process for EQ-5D-5L valuation studies. Value Health 20, 466–473.

Rowen, D., et al., 2012. It's all in the name, or is it? The impact of labeling on health state values. Med. Decis. Making 32 (1), 31–40.

Rowen, D., et al., 2015. Comparison of general population, patient, and carer utility values for dementia health states. Med. Decis. Making 35 (1), 68–80.
Stolk, E., et al., 2019. Overview, update, and lessons learned from the international EQ-5D-5L valuation work: version 2 of the EQ-5D-5L valuation protocol. Value Health 22 (1), 23–30.
Webb, E.J.D., et al., 2021. Does a health crisis change how we value health? Health Econ. 30 (10), 2547–2560.
Yang, Z., et al., 2019. How prevalent are implausible EQ-5D-5L health states and how do they affect valuation? A study combining quantitative and qualitative evidence. Value Health 22 (7), 829–836.