Article:
Peasgood, T., Bourke, M., Devlin, N. et al. (3 more authors) (2023) Randomised comparison of online interviews versus face-to-face interviews to value health states. Social Science and Medicine, 323. 115818. ISSN 0277-9536

https://doi.org/10.1016/j.socscimed.2023.115818

Social Science & Medicine 323 (2023) 115818

Contents lists available at ScienceDirect

Social Science & Medicine

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

Randomised comparison of online interviews versus face-to-face interviews to value health states

Tessa Peasgood [a,*], Mackenzie Bourke [a], Nancy Devlin [a], Donna Rowen [b], Yaling Yang [c], Kim Dalziel [a]

[a] Health Economics Unit, Centre for Health Policy | Melbourne School of Population and Global Health, The University of Melbourne, Australia
[b] School of Health and Related Research, University of Sheffield, UK
[c] Nuffield Department of Primary Care Health Sciences, University of Oxford, UK

ARTICLE INFO

Handling Editor: R Smith

Keywords:
EQ-5D-5L
EuroQol valuation protocol (EQ-VT)
Composite time trade off (TTO)
Videoconference

ABSTRACT

Introduction: Health state valuation studies using composite time trade-off (cTTO) interviews have historically been conducted face-to-face. The COVID-19 pandemic forced disruptive innovation meaning a number of valuation studies conducted interviews via videoconference. These studies found online interviews feasible and acceptable; however, studies were not constructed to test the impact of online versus face-to-face interviews. This study builds on its sister study from the UK and aims to assess the acceptability and equivalence of in person face-to-face interviews with online interviews on cTTO valuation outcomes and on data quality.

Methods: Participants were recruited into a randomised equivalence study via an external research company. Consenting participants were randomly allocated to complete a cTTO interview face-to-face or online, using the same 10 EQ-5D-5L health states. Mean and distribution of the cTTO values, participant understanding, data quality, demographic characteristics, participant preference, participant engagement and participant feedback were all compared across interview mode. Statistical equivalence for cTTO values for each state was tested using two one-sided t-tests by mode. Finally, regression analysis was completed to assess the impacts of interview mode on cTTO value while controlling for demographic characteristics of the participants.

Results: Mean cTTO values were shown to be equivalent for mild health states and showed no significant difference for serious health states. The proportion of individuals who expressed an interest in the study but declined to arrange an interview after finding out their randomisation was significantly higher for the face-to-face (21.6%) than the online group (1.8%). No significant difference was found between groups for participant engagement, understanding or feedback nor for any indicators of data quality.

Conclusion: Administrating interviews face to face or online did not appear to have a statistically significant impact on mean cTTO values. Offering both online and face-to-face interviews routinely allows all participants to select the most convenient option.

1. Introduction

The ability to quantify health-related quality of life (HRQoL) is valuable in health economic evaluation, as it offers a single outcome that can be used to compare impacts across different interventions (Brazier et al., 2017). Preference accompanied measures, such as the EQ-5D, are widely used to quantify HRQoL (Kennedy-Martin et al., 2020), with societal preferences standardly used to establish utility weights (Xie et al., 2015). Time trade-off (TTO) is a common method to elicit societal preferences (Lugnér and Krabbe, 2020). TTO involves trading off quantity for quality in different imaginary lives (Lugnér and Krabbe, 2020). Participants are offered the choice of living a given amount of time (e.g. 10 years) in a health state below full health or a shorter amount of time in full health. The time in full health is varied until the respondent is indifferent between the two lives; the point of indifference is used to calculate the utility value of the health state being valued. For very poor states considered worse than dead the approach can be adjusted to include lead-time in full health resulting in additional time

* Corresponding author. Health Economics Unit, Centre for Health Policy | Melbourne School of population and Global Health, The University of Melbourne, Level 4, 207 Bouverie, Carlton, Vic, 3010, Australia.
E-mail address: tessa.peasgood@unimelb.edu.au (T. Peasgood).

https://doi.org/10.1016/j.socscimed.2023.115818
Received 14 November 2022; Received in revised form 23 January 2023; Accepted 2 March 2023
Available online 8 March 2023

that participants can sacrifice. A composite time trade-off (cTTO) approach adopts a standard TTO for states better than dead and the lead-time approach for states considered worse than dead (Janssen et al., 2013).

The EuroQol Valuation Technology (EQ-VT) protocol was developed to standardise the methods used for valuing the EQ-5D instruments (Oppe et al., 2016; Stolk et al., 2019) and consists of cTTO and (usually also) Discrete Choice Experiment (DCE) questions. The quality control (QC) procedures and monitoring of interviewer protocol compliance is an integral part of the protocol (Oppe et al., 2014, 2016; Stolk et al., 2019). Data quality checks pay close attention to two further quality indicators. Firstly, the extent of clustering of values by interviewer, particularly at values of 0, 0.5, 1, −1 and −0.5 as clustering here arises when respondents undertake few trading moves and may indicate low engagement with the task and secondly, the number of half-year units used. As the cTTO routing defaults firstly to full years, the use of half-year units is indicative of higher precision in preferences arising from greater participant engagement in the task.

A preference for face-to-face interviewer administration of cTTO within the EQ-VT protocol (version 2.1) was supported by previous research that compared face-to-face interviewing in respondents' homes in London with individually completed cTTO tasks in the Netherlands using a computer program in a central venue with group level support (Oppe et al., 2016; Stolk et al., 2019; Shah et al., 2013). It was shown that face-to-face participants completed more trades per health state and had a lower proportion of clustering of values around −1, 0 and 1 than those undertaking the cTTO tasks without an interviewer (Shah et al., 2013). Concerns over online data quality was supported by two further studies that compared face-to-face interviews with unsupervised online survey formats (Norman et al., 2010; Jiang et al., 2021). However, more recent research has shown that conducting the interview online (fully supervised just as a face-to-face interview) using videoconferencing is now considered feasible (Finch et al., 2022; Lipman, 2021). Comparisons between mode of administration across a range of health state valuation techniques has also shown that online administration can be advantageous for reasons of convenience and sampling diversity (Lipman, 2021; Davies et al., 2020).

The COVID-19 pandemic, and subsequent global lockdowns of 2020 and 2021, produced an opportunity for disruptive innovation, leading researchers conducting valuation studies to explore the use of videoconferencing. In one Italian study, feasibility was investigated relating to technological, organisational and quality-control aspects of videoconferencing interviews. They found minimal issues with EQ-VT protocol non-compliance (Finch et al., 2022). As a result of the pandemic two studies using the EQ-VT protocol to value EQ-5D-3L-Y (Estévez-Carrillo et al., 2022) and the EQ-5D-5L (Lipman, 2021) were moved online during data collection. These studies reported no difference between the face-to-face and videoconferencing groups in terms of participants' engagement, as measured by task duration, number of inconsistent answers and number of half-year units utilised (Lipman, 2021; Estévez-Carrillo et al., 2022). However, as with prior research there appeared to be some demographic differences in education level and gender between online and face-to-face groups which may have impacted findings (Estévez-Carrillo et al., 2022). Furthermore, the proportion of participants failing to attend or cancelling at short notice was higher in the videoconferencing group (Lipman, 2021). One major limitation of these studies when examining the mode of administration effects was that they were opportunistic and not designed to answer the question of cTTO response equivalence between modes. Participants were not randomised into face-to-face or videoconferencing nor were the studies powered for comparative equivalence of cTTO responses. Furthermore, interviewers had more experience when conducting the online interviews, which could have differentially impacted data quality. A final important factor is that face-to-face and videoconferencing interviews were conducted before and after lockdowns respectively, when health consciousness in the general population was arguably different and this could have impacted on cTTO responses.

While recent evidence seems to suggest that videoconferencing is a viable option for conducting cTTO interviews (Finch et al., 2022) there is a clear need for a randomised comparison with face-to-face techniques in order to establish the equivalence of these techniques. A recent UK study that randomised participants to online and in-person interviews found that online and in-person cTTO interviews were acceptable and both generated good-quality data though across some criteria video interviews had lower quality data (Rowen et al., 2022). The study also found that cTTO values differed across modes for more severe EQ-5D-5L states, but when controlling for the participant characteristics it did not appear that mode was the cause. The study also identified that the characteristics of people preferring each mode differed. UK study limitations were that there were a larger proportion of participants who were interviewed online than face-to-face (62%, due to data collection being ended due to COVID-19), the sample was highly educated across both modes, and the study was conducted in 2021 when the COVID-19 pandemic was affecting life in general in the UK.

This Australian study was a sister study to the UK study (Rowen et al., 2022). In both cases, the objectives were to test the acceptability and equivalence of in person face-to-face interviews with online videoconferencing interviews on cTTO valuation outcomes and explore the impact of interview mode on attendance of different groups and on data quality. DCE comparisons were not included because health state valuation using DCE, in contrast to TTO, is more commonly conducted online and has been shown to provide equivalent values to CAPI face to face interviews (Mulhern et al., 2013). This study aims to establish whether the UK results are generalizable to a different English-speaking country with different lived experience of the COVID-19 pandemic. Given the potential impact of mode of administration on future valuation studies with large policy influence from their results, it is important to not rely upon a single study in two UK cities with one particular lived experience of COVID-19 at the time of the data collection. The study also aims to strengthen the evidence obtaining a sample that is more representative with respect to education level and in which equal proportions of participants interviewed in-person and online.

2. Methods

2.1. Recruitment, sampling and randomisation

An external market research company (Stable Research) conducted the recruitment. They invited members of their 100,000 plus panel of potential participants living in the target locations. Interested members completed a short online screening survey on their age, gender, location, education level and income. Recruitment used a multi-stage stratified sampling approach with quotas based on Australian Bureau of Statistics estimates for 2020 for all demographic components included in the screening survey. Once participants were recruited into the study they were randomised into the face-to-face or online group, after which they were contacted via phone to organise a convenient time for the interview. Participants who decided at this point that they no longer wished to take part in the research were asked why and their reasons were recorded. Those who agreed to take part in the project were sent a Plain Language Statement which described the purpose and content of the interview and links to an electronic consent form which they were required to complete prior to their interview. Participants who did not attend their prebooked appointments were contacted by the research company to find out their reasons for not attending (see Supplementary Appendix 2: Table S14). Given the difficulty of following up with participants who did not book after randomisation or did not attend their appointment, these data were incomplete and thus not included in the final analysis. Participants who completed the interview were compensated with an $40 AUD voucher for online participation and an $80 AUD voucher for face-to-face. At the initial point of invitation participants were informed that compensation would be at least $40;

after randomisation those allocated to face-to-face were informed of the higher compensation for their travel time and expenses. All interviews were conducted between the 6th of April and the June 1, 2022. Ethics approval was granted by the University of Melbourne Human Research Ethics Committee (2021-22030-20424-3).

2.2. Study design and sample size

The study was designed as a randomised controlled study to test the equivalence of mean cTTO values for online versus face-to-face interviews and also to explore any differences between the groups in terms of acceptability and data quality. Study outcomes include several different analyses to compare acceptability and data quality between modes and to compare cTTO values between modes for ten different EQ-5D-5L health states which would be expected to have substantially different standard deviations. Therefore, there is no single calculation on which to determine sample size. The minimum sample size for each arm was calculated to be powered to test for equivalence of mean value for the three mildest states. This assumed a power of 0.8, a significance level of 0.05, a standard deviation of 0.15 and an equivalence limit of 0.05. This requires 190 participants in each arm based on a two-sided equivalence test (estimated in STATA using 'ssi'). The estimate for the standard deviation is in line with that found for the mildest states in recent cTTO valuations applying EQ-VT 2.0 (Pickard et al., 2019; Jensen et al., 2021). The equivalence limit of 0.05 is reasonably aligned with most estimates of Minimum Clinically Important Difference estimates for the EQ-5D (Coretti et al., 2014; McClure et al., 2017). Furthermore, this value has been applied in other equivalence work on the EQ-5D (Gao et al., 2009; Luo et al., 2003). However, there is some uncertainty in the appropriate choice of equivalence limit for the Australian context. The standard deviations for the valuations of more serious health states tend to be much larger; up to 0.7 (Pickard et al., 2019; Jensen et al., 2021) which would require a sample of over 8000 for a similar equivalence test; this was not considered practical or necessary. The sample size of at least 400 (200 randomised to each mode) was a pragmatic choice given that findings between modes would draw upon a range of comparisons. Equivalence studies usually require more statistical power than those designed to test differences based on hypotheses. The randomised controlled design is also suitable for testing the difference between study groups for the following outcomes.

- Acceptability as measured by proportion completing interview
- Data quality
- Participant preference for interview mode of administration
- Participant understanding by mode of administration.

2.3. Selection of health states

Participants were asked to evaluate 10 health states which comprised one block in the standard EQ-VT design for valuation the EQ-5L–5L (these states are shown in Table 2). This was the same block as used in the UK study (Rowen et al., 2022) to enable a direct comparison. The EQ-5D-5L consists of five dimensions: mobility, self-care, usual activities, pain or discomfort and anxiety or depression. Within each dimension there are five levels of severity - no problem, slight, moderate, severe and unable/extreme. Health states are labelled based on the response level for each dimension for example, health state '12,111' describes a health state where someone has slight problems with self-care but no other problems. The 10 health states included ranged from mild (e.g. 21,111) to the most severe state (55,555).

2.4. The interviewers

Four interviewers were recruited to this study, all of whom had postgraduate or honours degrees in health-related subjects. Interviewers were each given a two-day training course adapted from the training content provided by the EuroQol EQ-VT support team. After completing the initial training, each interviewer conducted 10 pilot interviews using a mix of administration modes (which were not included in the dataset for analysis) and were then given feedback in accordance with the EQ-VTv2 QC protocol. Any flagged interviews that did not meet protocol compliance were discussed and more training was given if necessary. The EQ-VTv2 QC process (Ramos-Goñi et al., 2017) continued throughout data collection. Interviewers undertook a broadly similar number of interviews online as face-to-face, in the same time periods to ensure interviewer learning effects were similar in both modes. Interviewers also attended a mix of face-to-face locations.

2.5. The interviews

All interviews were conducted in line with EQ-VTv2 protocol, applying the cTTO technique (as described earlier). Participants were asked background questions about their age and gender and their own experience with serious illness, personally, in their family and in caring for others. Next, they completed the EQ-5D-5L and a 0–100 visual analogue scale (VAS) rating of their own health.

Participants were then given five example warm-up cTTO tasks which included an explanation of all the elements of the trade-off task. In the first warm-up task participants were asked to imagine having a mobility problem which required being in a wheelchair. If participants gave a positive cTTO value for this first task they were asked to imagine a state much worse than being in a wheelchair for the second warm-up example, where they were exposed to the cTTO procedure used to value states worse than dead. If participants offered a negative value for the wheelchair example, they were asked to imagine a state much better than being in a wheelchair for the second warm-up task and were exposed to the task used to value states better than dead. Interviewers ensured that participants had been exposed to both procedures used to value states considered better than and worse than dead in the warm-up tasks before moving on. Participants were given three further practice tasks, designed to familiarise respondents with the range of EQ-5D-5L health states. Participants were then asked to complete the cTTO tasks for the 10 health states included in the experiment, presented in random order.

Participants were then asked about their difficulty with the task and their understanding of it. Participants were then shown the 10 health states as they had ranked them based on utility values derived from their cTTO choices (referred to as the feedback module). At this point participants were able to flag any health states they believed were out of order. Following this feedback module participants were asked some further socio-demographic questions and then asked their (*ex post*) preferred interview mode and what they thought of the task.

Face-to-face interviews were conducted in several different locations in inner-city Melbourne, western-suburbs of Melbourne, and the inner regional areas of Ballarat and Geelong to ensure some socio-geographic diversity. Interviews in the inner-city were conducted on the University of Melbourne campus, all other interviews were conducted in hired meeting rooms. Online interviews were conducted via Zoom and all participants required either a computer or tablet with a camera that was switched on. In both online and face-to-face interviews, the screen was controlled by the interviewer and all responses were entered by the interviewer.

2.6. Analysis

Equivalence was assessed statistically by comparing the mean cTTO values by mode.

Other potential differences between mode were explored by investigating the distribution of cTTO values, sample representativeness, participant preference for interview mode and participant understanding and engagement by mode. This included regression analyses to assess the impact of mode of administration on a) initial acceptance of

Table 1
Sociodemographic characteristics of final cohort, overall and by mode.

	Total n = 403 (%)	Overall	F2F n = 190	F2F	Online n = 213	Online	p-value (two sample test of proportions)	Australian Population 2021
		%	n	%	n	%		
Age group								
18–34	106	26.30	45	23.68	61	28.64	0.2595	
35–49	119	29.53	61	32.11	58	27.23	0.2842	
50–64	102	25.31	51	26.84	51	23.94	0.5041	
65+	76	18.86	33	17.37	43	20.19	0.4701	
Gender								
Male	173	42.93	75	39.47	98	46.01	0.1858	49.3
Female	229	56.82	115	60.53	114	53.52	0.1564	50.7
Other/prefer not to say	1	0.25	0	0	1	0.47	0.3443	NA
Experience of serious illness								
In oneself	196	48.95	97	51.05	99	46.64	0.359	
In family	317	78.96	150	78.4	167	78.66	0.894	
Care for others	183	45.41	88	46.6	95	46.32	0.73	
Country born								
Australia	303	75.19	140	73.68	163	76.53	0.51	
Other	100	24.81	50	26.47	50	23.47		
Language at home								
Only English	323	80.15	149	78.42	174	81.69	0.411	
Other	80	19.85	41	21.58	39	18.31		
Aboriginal or Torres Strait Islander origin								
Yes	5	1.24	2	1.05	3	1.41	0.756	
No	398	98.95	188	98.95	210	98.59		
Highest educational level								
Year 10 or below	7	1.74	4	2.11	3	1.41	0.593	
Year 12 equivalent	61	15.14	25	13.16	36	16.9	0.2952	
Certificate/Diploma	150	37.22	63	33.16	87	40.85	0.111	
Bachelor	91	22.58	51	26.84	40	18.78	0.0533	
Above bachelor	94	23.33	47	24.74	47	22.07	0.5268	
Bachelor or above								35% (from aged 25–74 years)
Household Income level (annual before tax)								
<$19,999	9	2.23	4	2.11	5	2.35	0.8696	
$20,000-$39,999	35	8.68	21	11.05	14	6.57	0.1109	
$40,000-$59,999	46	11.41	22	11.58	24	11.27	0.9218	
$60,000-$99,999	77	19.11	34	17.89	43	20.19	0.5589	
$100,000-$149,999	86	21.34	43	22.63	43	20.19	0.55	
>$150,000	94	23.33	45	23.68	49	23	0.8721	
Prefer not to say	56	13.9	21	11.05	35	16.43	0.1191	
Victorian Median Household Income (annual)								$91468
Employment								
Employed part time	73	18.11	36	18.95	37	17.37	0.6817	
Employed full time	168	41.69	81	42.63	87	40.85	0.7165	
Self-employed part time	16	3.97	9	4.74	7	3.29	0.4566	
Self-employed full time	12	2.98	4	2.11	8	3.76	0.3305	
Retired	63	15.63	25	13.16	38	17.84	0.1963	
Housework/carer	12	2.98	7	3.68	5	2.35	0.4306	
Student	14	3.47	6	3.16	8	3.76	0.7435	
Seeking work	6	1.49	2	1.05	4	1.88	0.4947	
Unemployed	5	1.24	1	0.53	4	1.88	0.2375	
Long term sick/disability	20	4.96	12	6.32	8	3.76	0.4564	
Other	14	3.47	7	3.68	7	3.29	0.8277	
Limitation on daily activities								
Yes, limited a lot	29	7.2	14	7.37	15	7.04	0.8994	
Yes, limited a little	110	27.3	57	30	53	24.88	0.2497	
No	264	65.51	119	62.63	145	67.08	0.2511	
Marital Status								
Single	92	22.83	36	18.95	56	26.29	0.0795	
Married/Partner	262	65.01	123	64.74	139	65.26	0.9128	
Separated	10	2.48	7	3.68	3	1.41	0.1426	
Divorced	26	6.45	17	8.95	9	4.23	0.0541	
Widowed	13	3.23	7	3.68	6	2.82	0.6228	
Home Ownership								
Own/Mortgage	288	71.46	137	72.11	151	70.89	0.7878	
Rent from government	10	2.48	4	2.11	6	2.82	0.6466	
Rent from private	95	23.57	44	23.16	51	23.94	0.8528	
Other	8	1.99	4	2.11	4	1.88	0.8703	
Prefer not to say	2	0.5	1	0.53	1	0.47	0.9354	
Parent to child U18								
Yes	123	30.52	62	32.63	61	28.64	0.3849	
No	280	69.48	128	67.37	152	71.36	0.3849	
EQ-5D-5L level 1 or 2								

(*continued on next page*)

Table 1 (*continued*)

	Total n = 403 (%)	Overall	F2F n = 190	F2F	Online n = 213	Online	p-value (two sample test of proportions)	Australian Population 2021
		%	n	%	n	%		
Mobility	373	92.56	173	91.53	200	93.46	0.463	
Self-care	397	98.51	186	98.41	211	98.60	0.878	
Usual activities	354	87.84	164	86.77	190	88.79	0.537	
Pain/Discomfort	333	82.63	154	81.48	179	83.64	0.567	
Anxiety/Depression	334	82.88	151	79.89	183	85.51	0.135	
VAS (mean)	76.59		76.51		76.67		0.911	

Notes: p test for VAS is two sample *t*-test.

the invite to the interview; b) attendance of participants to the scheduled interview and c) on mean cTTO value adjusting for socio-demographic factors.

2.6.1. Sample representativeness and sample by mode

Two samples were analysed, the first was only those who completed the interview and the second included every participant who was randomised and invited to interview. For those who attended (sample 1), participants in each arm of the study were compared using two-sample tests of proportions for all sociodemographic characteristics collected in the study. This sample was also compared to Australian 2021 census data for gender, age, education and income.

The second sample was analysed to assess whether socio-demographic characteristics and mode of administration impacted the likelihood that a participant would accept and subsequently attend their interview.

2.6.2. Participant understanding, engagement and preference for mode

Following the cTTO task participants were asked to rate their perceived understanding of the interview on a scale from 1 (strongly agreeing) to 5 (strongly disagreeing) with the statement "it was easy to understand the questions I was asked." After the interview, interviewers were prompted to rate the participant's understanding of the task. Both tests of understanding were compared across the online and face-to-face groups using Chi-squared test of significance.

Participant engagement was measured subjectively with a 1–5 scale ranging from strongly agree to strongly disagree in response to the question "I got bored during the interview'. Objective measures of engagement and understanding were also compared between groups as part of the data quality assessment.

Participant preference for mode of administration was established overall and compared between groups (noting that participants were randomised to the mode by which they were interviewed). Participants were asked which mode they would have preferred to have received: online, face-to-face or no preference. They were then asked to offer one or more reasons for their choice. Participants who had a preference were compared using logistic regression to establish whether any socio-demographic characteristics were predictive of mode preference. For this analysis those who had no preference were not considered.

2.6.3. Data quality

Data quality was assessed and compared using a number of different metrics that are routinely collected as part of the EQ-VT protocol (Oppe et al., 2016). These were (a) clustering, (b) response inconsistencies, (c) the number of trades used to achieve indifference per health state (d) using only integer values, (e) using only positive values and (f) using few values. Clustering refers to the distribution of cTTO values having higher frequency around the values 1, 0.5, 0, −0.5 and −1. This can be a sign that participants are using too few values to rank health states rather than utilising the full cTTO range and thus can be a sign of participants not understanding or not engaging fully with the task.

Inconsistent responses occur when participants' cTTO values suggest that they rank health states that are logically worse across one or more dimensions above health states that are logically better. One specific form of logical inconsistency flagged within the EQ-VT QC protocol (Stolk et al., 2019) occurs when health state 55,555 is valued at least 0.5 higher than a logically better state.

The design of the cTTO task allows participants to trade off half-years, and to use negative values to a minimum of −1. Utilising these aspects of the task can be a sign that participants are engaging with the task. If participants chose to use only full year values, only positive values or less than five possible values, it may be a sign of decreased engagement with the task, thus these metrics were compared between groups. Finally, interviewers reported participant engagement and understanding after completion of the interview and these scores were compared between groups.

2.6.4. Interviewer effects

Interviewer performance was assessed as part of the EQ-VT QC protocol. Interviewers were monitored for flagged interviews and potential quality concerns throughout the data collection. Flagged interviews occurred when any of the EQ-VT QC standards were violated; i) participants were not shown the worse than dead task in the wheelchair example; ii) less than 3 min is spent on the practice wheelchair example; iii) inconsistencies in the cTTO data as shown by the pit state '55,555' being at least half a year above the lowest health state; or iv) all 10 cTTO tasks being completed in less than 5 min. Data were also collected on interviewer results for clustering, positive value only traders, non-traders, number of moves to complete cTTO tasks and amount of time to complete cTTO tasks. Interviewer effects were analysed as part of the regression analysis and were also tested for interaction effects with mode.

2.6.5. Mean cTTO values, standard deviation and median for each health state

Statistical equivalence for mean cTTO values for each state was tested. The first analysis used two one-sided t-tests (TOST) by mode (Schuirmann, 1987). The null hypothesis of non-equivalence was rejected if both one-sided t-tests reported a p-value of less than 5%. Secondly, standard two sample t-tests to test the difference in the means between groups were also conducted. Thirdly, the modified Levene's robust test of standard deviations from the medians (Brown-Forsythe Test) was used to compare statistical difference in standard deviations by mode of adminstration, robust estimates were used to allow for the non-normal distribution of the cTTO values. Finally, the non-parametric Wilcoxon rank sum test was used to compare the medians across the two modes of administration.

2.6.6. Regression analysis

Regression analysis was used to explore the impact of mode whilst controlling for health states, respondent sociodemographic characteristics and potential interviewer effects. As each participant values 10 health states, collinearity between their own values was anticipated. Additionally, the data is censored at −1 due to the nature of the lead-time component of the task which does not allow respondents to record a value below −1. Finally, as standard deviation was anticipated to

Table 2
Mean TTO values for each health state by mode, with tests of significance.

	F2F			Online			Tests of Difference					
	mean	SD	median	mean	SD	median	Differences in mean TTO values	P-value (*t*-test of mean)	P-value (Wilcoxon's *U* test of median	P-value (Brown-Forsythe test of variances)	P-Value (Two one-sided t tests)	
21111	0.947	0.140	1.00	0.951	0.097	1.00	−0.004	0.713	0.409	0.504	0.000	0.000
11212	0.924	0.136	1.00	0.912	0.190	1.00	0.011	0.504	0.589	0.552	0.000	0.010
12112	0.897	0.201	0.95	0.881	0.225	0.95	0.016	0.462	0.310	0.713	0.001	0.055
23152	0.132	0.653	0.30	0.135	0.669	0.40	−0.003	0.964	0.858	0.068	0.239	0.211
21345	−0.011	0.701	0.10	0.032	0.653	0.20	−0.043	0.524	0.740	0.824	0.459	0.084
34244	−0.141	0.674	0.00	−0.118	0.641	0.00	−0.023	0.723	0.860	0.158	0.342	0.132
43514	0.116	0.637	0.35	0.083	0.648	0.30	0.033	0.606	0.573	0.826	0.098	0.397
55424	−0.148	0.632	0.00	−0.195	0.614	0.00	0.047	0.453	0.408	0.163	0.060	0.479
44553	−0.332	0.627	−0.50	−0.331	0.587	−0.50	−0.001	0.991	0.759	0.684	0.208	0.201
55555	−0.575	0.534	−0.90	−0.549	0.502	−0.70	−0.026	0.614	0.269	0.806	0.322	0.071
All States	0.199	0.741	0.45	0.203	0.724	0.45	−0.004	0.842	0.747	0.344	0.007	0.002

increase with health state severity, heteroscedasticity was also predicted. To deal with these complicating factors the dependent variable was set to the amount of time that individuals traded off and was thus equal to 1-cTTO (2 means the participant traded off all the time; 0 meant they traded off no time). The model completed was a tobit multiplicative heteroscedasticity regression that was robust to clustering of standard errors of each participant (estimated in stata using '*tobithettm*') (Shehata, 2011).

A secondary logistic regression was conducted to assess the impact of mode and socio-demographic characteristics on initial acceptance of the interview after randomisation and of those same characteristics on attendance once the interview had been accepted. A final logistic regression was conducted to assess whether any sociodemographic characteristics were predictive of preference for mode.

3. Results

The flow of participants from randomisation to complete interview is shown in Fig. 1. 403 interviews were conducted in total, of which 47% were face-to-face. The four interviewers conducted between 44% and 50% of their interviews as face-to-face. Results disaggregated by interviewer can be found in the standard EQ-VT QC report available in Supplementary Appendix 1.

3.1. Sample by mode and sample representativeness

Two samples were analysed separately for this study. The first sample included only those who completed their interviews (Table 1). The second sample included all participants who volunteered for the research project and were subsequently randomised. This sample had three groups: i) those who refused an interview once they were randomised, ii) those who booked an interview but did not attend and iii) those who attended their interview (Supplementary Appendix: Table S1).

The sample size of those who attended their interviews in the second sample ($n = 405$) includes two participants who were not included in the first sample ($n = 403$). These two participants attended their interviews at the correct time and location, however due to organisational errors did not complete their interviews.

In the sample of those who were randomised, the proportion of individuals who did not accept their interview in the face-to face group (21.6%) was significantly larger than the proportion in the online group (1.9%) ($p < 0.0001$) (Supplementary Appendix 2: Table S1). While the cancellation rate was higher in the online group the difference was not significant ($p = 0.09$) and thus the proportion of those who actually attended their interviews was larger in the online group (80.3%) compared to the face-to-face group (65.6%) ($p = 0.0001$) (Supplementary Appendix 2: Table S1). This relationship remained when controlling for socio-demographic characteristics in a logistic regression (Supplementary Appendix 2: Table S2), with the odds ratio of acceptance when randomised to face-to-face being 0.060 ($p < 0.001$). However, once those who did not accept the interview were removed from the sample, mode was not predictive of interview attendance (Supplementary Appendix 2: Table S2).

The demographic characteristics of face-to-face and online samples were compared to the most recent Australian census data (Table 1). The sample had a higher proportion of females 56.8% compared to the national average of 50.7% and was more highly educated, 45.9% with a bachelor's degree or higher compared to 35% in the general population (25–74 years). The sample was representative for age and approximately representative for household income based on median income for Victoria. No statistically significant demographic differences were identified between the online and face-to-face groups of those who completed their interviews (Table 1).

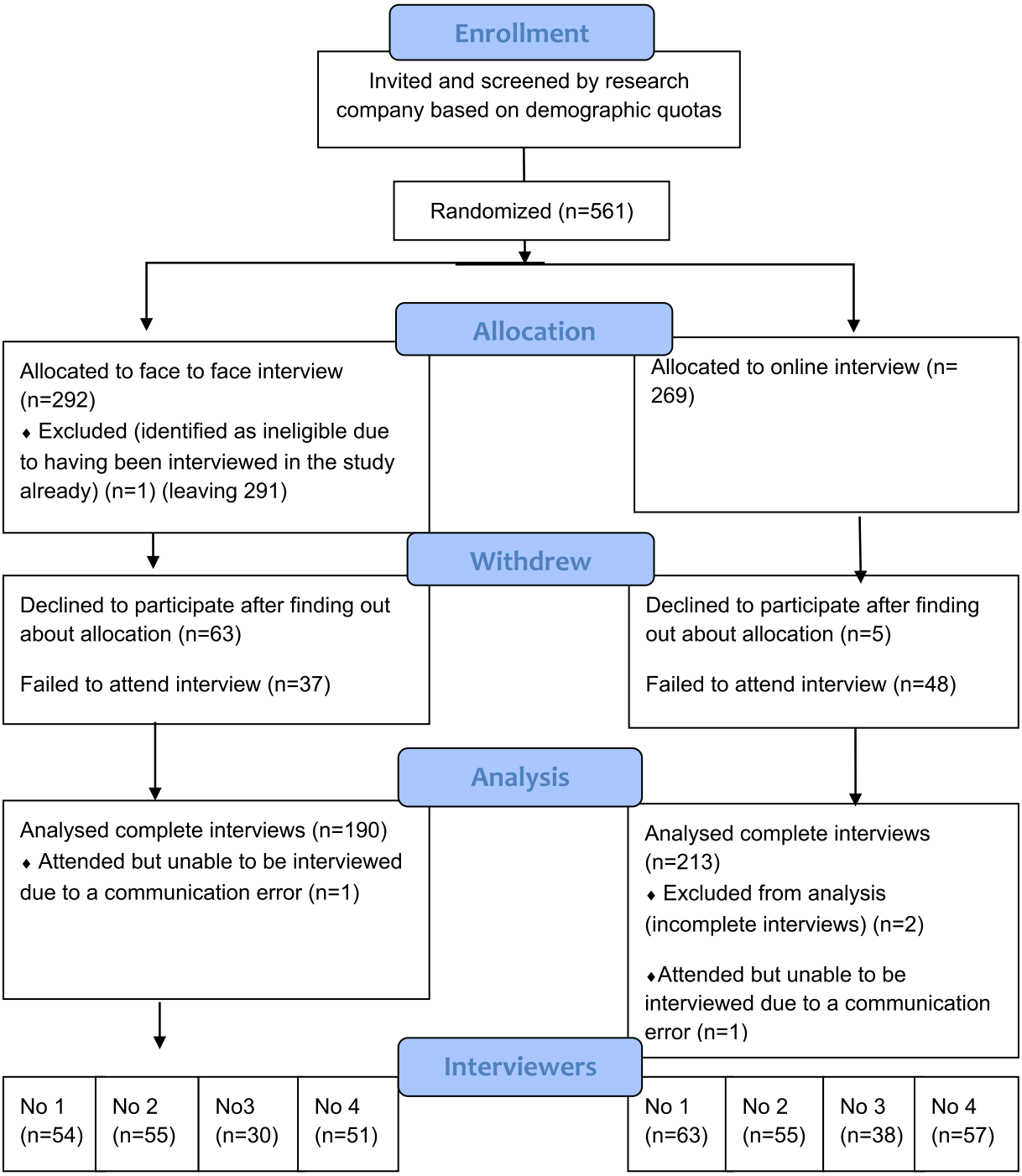

Fig. 1. Flow diagram of participant progress through the study.

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

Participant understanding of the task was strong in both arms of the study with 75.2% of the online group and 71.4% of the face-to-face group strongly agreeing with the statement "it was easy to understand the questions I was asked" ($p = 0.664$) (Supplementary Appendix 2: Table S3). There was no statistically significant difference in the two other self-reported understanding questions "I found it easy to tell the difference between the lives I was asked to think about" ($p = 0.598$) and "I found it difficult to decide on the exact points were Life A and Life B were about the same" ($p = 0.661$).

Self-reported participant engagement showed no difference between modes. 93.2% in the face-to-face group and 88.3% in the online group reported strongly disagreeing with the statement "I got bored during the interview" ($p = 0.207$) (Supplementary Appendix 2: Table S4). No significant between group differences were found for any of the subsequent feedback questions (Supplementary Appendix 2: Table S4). Interviewers reported that participants "concentrated very hard and put a great deal of effort into it" for 81% of face-to-face interviews and 73% of online interviews (Supplementary Appendix 2: Table S5). Interviewers reported that participants "didn't concentrate very hard and put little effort into it" in 9 (2.2%) interviews. No significant interviewer reported difference in concentration was noted between groups ($p = 0.301$). Furthermore, interviewers reported it was "doubtful whether the respondent understood the exercise" in only 18 (4.5%) of the interviews

and there was no difference between arms of the study ($p = 0.814$). Interviewer perception of understanding showed no significant difference between groups ($p = 0.434$).

Participant preference for mode was significantly different between groups ($p < 0.0001$) (Supplementary Appendix: Table S6). Overall, 151 (37.4%) participants reported that they would have chosen to be interviewed online if given the choice, 51 (12.7%) would have preferred to be interviewed face-to-face and 201 (49.9%) did not mind. The most common reason given for preferring online was "more convenient" (94.7%) whereas for face-to-face it was "would feel most at ease being interviewed that way" (80.4%).

While online interviews were preferred to face-to-face interviews in both groups, the mode by which a participant was interviewed was predictive of their preference for that mode. Logistic regression confirmed this preference by mode with participants who completed face-to-face interviews significantly more likely to report a preference for face-to face interviews (OR 26.39, $p < 0.001$) (Supplementary Appendix 2: Table S7). Being employed made participants more likely to prefer online interviews than those who were not employed (OR 4.624, $p = 0.004$) and participants in age group 50–64 preferred face-to-face compared to online interviews (OR 3.175, $p = 0.042$).

3.3. Data quality

Overall, the quality of the data was good, and no statistically significant differences were seen between groups. There was no significant difference in the proportion of values at 1, 0.5, 0, −0.5 or −1 between groups for any of the health states (Supplementary Appendix Table S8). There was no difference in the proportion of any of the types of potentially problematic responders (Supplementary Appendix: Table S9) which includes participants who used less than five distinct health values overall (11.58% F2F vs 9.39% Online, $p = 0.473$), participants who did not use half-years (35.26% F2F vs 33.8% Online, $p = 0.758$), participants who only used positive values (22.63% F2F vs 17.37% Online, $p = 0.186$), participants who only used only values 1, 0.5, 0, −0.5 or −1 (11.58% F2F vs 10.8% Online, $p = 0.804$) and those who valued at least two health states at zero without valuing any health states below zero (2.11% F2F vs 1.41% Online, $p = 0.593$). No participants gave the same value for every health state. The mean number of trades completed per health state was 6.63 for online group and 6.82 for the face-to-face group ($p = 0.12$) for all health states. There were no significant differences in the mean number of trades in any individual health state (Supplementary Appendix Table S10).

3.4. Mean cTTO value and distributions for each health state by mode

The study found that cTTO values were similar across both arms for mean, SD and median (Table 2). The difference in means does not exceed 0.05 for any of the health states, which is the smallest difference possible within the individual cTTO task.

The difference in the standard deviations of each health state between the two modes were not significant for any of the health states at the 5% level. There was no statistically significant difference in median value between mode of administration for any health states. Standard two-sided t-tests for difference in means between online and face-to-face showed no significant difference overall or for any of the individual health states.

Statistical equivalence of mean cTTO value between groups was demonstrated when all health states were considered. The two one sided t-tests gave p values of $p = 0.007$ and $p = 0.002$, therefore we accept the alternative hypothesis that the difference between the group means is inside our equivalence interval (+0.05, −0.05)). However, means were equivalent between modes for only two of the individual health states at the 5% level (state 11,212 and state 2111). One further state (12,112) was equivalent at the 10% significance level. The difference in means for the eight health states that did not show equivalence at the 5% level did not show a dominant pattern, with larger mean cTTO values for face-to face interviews in five of the health states and smaller mean cTTO values for face-to face interviews in three of the health states. There was also no relationship between increasing health state severity and direction of mean difference. The overall mean difference between groups was 0.004 larger in the online group.

3.5. Regressions

Mode of administration did not have a statistically significant impact on cTTO values when all states were considered ($p = 0.817$) and this was consistent with the regression models for each of the individual health states (Table 3). Being a parent or guardian to a child under the age of 18 had a significant effect on mean cTTO value overall, with parents being less likely to trade off time. Interviewer effects on cTTO values were statistically significant for two interviewers overall and for all moderate and severe health states. Compared to interviewer 1, values were on average 0.109 lower ($p < 0.026$) for interviewer 3 and 0.118 lower ($p = 0.008$) for interviewer 4 overall. Supplementary regression (Supplementary Appendix: Table S17) investigated the interactions between interviewer and mode and found no significant interactions when combining the health states overall. When analysed separately, interviewer 4 showed significant interactions for 2 health states, state 21,345 and state 43,514, and no other significant interactions were discovered. There was no significant interaction between mode and age group, or between mode and gender (Supplementary Appendix: Table S17). The interaction between interview mode and holding a university degree was statistically significant for four moderate-to-severe health states with those in this group trading more time when undertaking face-to-face interviews than online interviews. However, there was no statistically significant interaction effect when all states were combined ($p = 0.095$).

3.6. Comparison to the UK sister study

Comparison to the UK sister study showed mean values of one of the ten states valued face-to-face and three of the states valued online to be significantly different at the 5% level (Supplementary Appendix 2: Table S16). The three most severe states were valued higher in the UK when valued online.

4. Discussion

The results from this study indicate that cTTO values obtained from face-to-face and online interviews are statistically equivalent for mild health states and when all the ten health states were combined. The finding of equivalence for mild health states only is likely to be explained by the sample size that was calculated for these milder health states with smaller standard deviations. However, we do not wish to rely too heavily on these equivalence tests given the uncertainty around the appropriate equivalence limits.

Differences in mean cTTO value for moderate and severe health states were extremely small and standard statistical tests for difference in mean, standard deviation and median found no significant differences by mode of administration. This finding was supported by exploratory regression analysis which found no statistically significant impact of mode on cTTO value once adjusted for demographic characteristics and interviewer effects.

The data from the interviews did not demonstrate any significant between-group difference in data quality, as measured by the mean number of cTTO moves, proportion of problematic responders and interviewer reported participant engagement and understanding. Historically a limitation of online valuation methods administered using a survey without an interviewer present has been lower data quality (Shah et al., 2013; Norman et al., 2010; Jiang et al., 2021). However, these results add weight to more recent literature which demonstrated the

Table 3
Regressions of total time traded, overall and by state, adjusting for sociodemographic characteristics (Tobit).

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	All States	All States (het)	state 21111	state 11212	state 12112	state 23152	state 21345	state 34244	state 43514	state 55,424	state 44,553	state 55,555
Mode f2f	−0.007		0.005	−0.009	−0.015	−0.007	0.056	0.030	−0.038	−0.057	0.035	0.088
	(0.817)		(0.662)	(0.571)	(0.471)	(0.921)	(0.476)	(0.704)	(0.602)	(0.460)	(0.655)	(0.274)
Age 35 to 49	0.000		0.009	0.009	0.020	−0.026	−0.029	0.124	0.016	−0.050	−0.008	−0.019
	(0.990)		(0.581)	(0.683)	(0.503)	(0.798)	(0.790)	(0.257)	(0.871)	(0.641)	(0.940)	(0.864)
Age 50 to 64	−0.020		0.011	−0.022	0.016	0.115	0.084	0.197*	0.009	0.058	0.004	−0.129
	(0.620)		(0.514)	(0.351)	(0.602)	(0.262)	(0.442)	(0.076)	(0.928)	(0.592)	(0.970)	(0.250)
Age 65 plus	0.023		0.023	0.007	−0.003	0.213	0.060	0.330**	0.078	0.104	0.042	−0.029
	(0.633)		(0.278)	(0.827)	(0.939)	(0.104)	(0.668)	(0.021)	(0.551)	(0.454)	(0.763)	(0.841)
Interviewer 2	0.024		−0.004	0.010	−0.034	−0.134	−0.216**	−0.182*	−0.062	−0.091	−0.078	0.078
	(0.553)		(0.817)	(0.653)	(0.227)	(0.168)	(0.038)	(0.082)	(0.522)	(0.374)	(0.447)	(0.452)
Interviewer 3	0.109**		−0.031**	0.002	−0.020	0.317***	0.315***	0.344***	0.315***	0.328***	0.440***	0.462***
	(0.026)		(0.047)	(0.939)	(0.482)	(0.001)	(0.003)	(0.001)	(0.001)	(0.002)	(0.000)	(0.000)
Interviewer 4	0.118***		−0.013	−0.003	−0.050	0.313***	0.157	0.398***	0.203*	0.170	0.419***	0.363***
	(0.008)		(0.472)	(0.893)	(0.125)	(0.005)	(0.192)	(0.001)	(0.070)	(0.150)	(0.001)	(0.003)
Female	0.009		−0.016	−0.028*	−0.017	0.239***	0.121	0.092	0.001	0.080	0.070	0.131
	(0.755)		(0.179)	(0.098)	(0.436)	(0.001)	(0.130)	(0.259)	(0.992)	(0.310)	(0.383)	(0.111)
VAS (5L)	0.001		0.000	0.001	0.001	0.000	0.000	−0.001	0.001	0.002	0.002	−0.002
	(0.324)		(0.298)	(0.130)	(0.318)	(0.982)	(0.875)	(0.741)	(0.588)	(0.426)	(0.459)	(0.448)
Parent or guardian of an under 18	−0.072**		−0.000	−0.026	−0.059**	−0.038	−0.120	0.006	−0.119	−0.063	−0.088	−0.152
	(0.045)		(0.985)	(0.205)	(0.022)	(0.665)	(0.206)	(0.951)	(0.175)	(0.498)	(0.353)	(0.119)
Employed	0.012		−0.003	0.016	0.023	0.006	0.027	−0.007	0.042	−0.034	−0.007	−0.059
	(0.731)		(0.866)	(0.452)	(0.404)	(0.946)	(0.789)	(0.947)	(0.655)	(0.728)	(0.947)	(0.567)
State_11212	0.076***	0.054										
	(0.002)	(0.728)										
State_12112	0.116***	0.190										
	(0.000)	(0.219)										
State_21345	1.099***	0.924***										
	(0.000)	(0.000)										
State_23152	0.975***	0.901***										
	(0.000)	(0.000)										
State_34244	1.249***	0.874***										
	(0.000)	(0.000)										
State_43514	1.016***	0.865***										
	(0.000)	(0.000)										
State_44553	1.456***	0.774***										
	(0.000)	(0.000)										
State_55424	1.296***	0.814***										
	(0.000)	(0.000)										
State _55555	1.688***	0.615***										
	(0.000)	(0.000)										
var (e.total-time-traded)			0.014***	0.027***	0.045***	0.514***	0.582***	0.587***	0.587***	0.557***	0.562***	0.551***
			(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Constant	−0.253**	0.276***	0.028	0.030	0.087	0.622***	0.869***	0.973***	0.973***	1.009***	1.102***	1.757***
	(0.015)	(0.000)	(0.449)	(0.566)	(0.190)	(0.007)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	4030	4030	403	403	403	403	403	403	403	403	403	403

Notes: Robust p value in parentheses (***p < 0.01, **p < 0.05, *p < 0.1). Baseline: Mode online, Age <35, Male or other gender, Interviewer 1, State_21,111, Not a parent/guardian of <18 child, Not currently employed/self-employed, State_21,111. Column (Kennedy-Martin et al., 2020) shows the modelling of the heteroskedasticity of the error term; only the health states were used to model this heterogeneity. var(e.total-time-traded) = variance of the error term.

feasibility of online cTTO interviews using videoconferencing technology (Finch et al., 2022; Lipman, 2021; Estévez-Carrillo et al., 2022; Rowen et al., 2022).

Participants reported a significant preference for online interviews over face-to-face interviews, particularly in those randomised to the online group. It is plausible that participants were more likely to prefer the mode by which they completed their interview, once they had experienced it. Another likely contributing factor for this result, is that those who had a preference for the opposite mode to which they were allocated dropped out prior to accepting their interview thus biasing the reported preferences. Participants who were employed were more likely to prefer online interviews, which is not surprising given their convenience which was the most commonly offered reason for preferring online. The preference for online interviews was consistent with the higher acceptance rate for participants randomised to the online group. However, cancellations were slightly higher in the online group, although not significantly so.

Coefficients for dummy variables assessing interviewer effects in the regression analysis showed statistically significant interviewer effects. These effects were present when all health states were considered and for moderate and severe health states. Interviewer effects have been shown in the literature in valuation studies for the EQ-5D using TTO tasks (Oppe et al., 2016; Stolk et al., 2019; Purba et al., 2017), and do not on their own indicate poor data quality. There was no indication that the mode had any impact on these interviewer effects.

The sample size for this study was calculated for testing the equivalence of the mild health states, which have lower standard deviations. One limitation was that the study may have been insufficiently powered to establish significant equivalence in the moderate and severe health states where standard deviation of cTTO values is higher. While the two

one-sided t-tests did not establish statistical equivalence, standard tests for difference found no statistical difference in mean, SD or median for cTTO values in any of the health states.

Once participants who did not accept the interview were removed from the cohort participants with a TAFE (Technical and Further Education) or high-school level of education compared to Bachelor education were less likely to attend their interview. It is plausible that the type of work these participants do makes completing the 40-min interview more difficult. This higher cancellation rate may partially explain the highly educated final cohort that completed the interviews.

The findings support the conclusions and recommendations of the sister UK study also assessing the acceptability and equivalence of online videoconferencing and face-to-face TTO interviews (Rowen et al., 2022) and has an advantage of an equal sample size by mode (the UK study had a larger proportion of respondents (62%) interviewed using the online mode) and a less highly educated sample. The preference for online interviews, difference in samples preferring online interviews to face-to-face interviews, and reasons provided for preferring online/face-to-face interviews were consistent in both studies. However, reassuringly this study did not replicate the UK findings of a difference in the cTTO values for the more severe states (though the difference in the UK study did not appear to be caused by mode).

Interviewers were asked to share their perceptions of the difference between interview mode, with three of the four interviewers perceiving more engagement and attention when conducting interviews face-to-face, reporting that it allowed more time for 'small talk' and rapport building. One interviewer noted that online some participants seemed to be distracted by their surroundings, whether it be by their pets/family at home, or even by other windows that may have been simultaneously open on their device. In contrast, another interviewer noted online interviews appeared to be less in a rush when conducted online compared to in person. They reported that some face-to-face interviewees were more time conscious throughout the interview due to issues such as parking, other appointments and needing to get back to work. It was noted by one interviewer that compassionate management of respondents was easier in face-to-face interviews where body language was more easily assessable, this was of particular importance for participants who may have found the interview questions triggering.

4.1. Limitations

The sample had a higher proportion of females than the general population, but there was no between group differences in the proportion of females, and regression analysis did not demonstrate a significant interaction between sex and interview mode on mean cTTO value. The men in the sample were more likely to have a university degree than the general male Australian population, as the demographic of males without a degree proved challenging to recruit within the study area. There was some evidence of interaction between holding a university degree and mode, with a significant interaction being present for 4 of the 10 health states. The representation of the population in terms of socio-economic status was good, in part due to the efforts to hold face-to-face interviews in a diverse range of locations including in lower socio-economic geographical areas.

Willingness to attend both a face-to-face interview and an online interview with availability of a computer or large screen device and a camera was stated as an inclusion criterion for potential participants within the initial invite to the study. Furthermore, the recruitment via a market research company limits the sample to those who have previously shown an interest in research. The finding of equivalence of mode of administration may not extend beyond this self-selected group.

Participants were told about the additional $40 reward for face to face interviews (to cover travel costs) only after they had expressed an interest in the study and been randomised. This approach may have influenced respondent's decisions. However, the unexpected difference in financial reward would be expected to encouraged face to face take up and discouraged online take up yet our response rates find higher initial take up of online interviews following randomisation.

5. Conclusion

Including either online interviews or face-to-face interviews in future Australian valuation research using cTTO is appropriate. Online interviews were preferred by a larger proportion of participants than face-to-face interviews and those who were randomised to the online group displayed a lower dropout rate. The use of online interviews did not decrease data quality. While online interviews were preferred by more participants, a proportion still preferred face-to-face interviews. Participants who were employed were more likely to prefer online interviews and participants aged 50–64 were more likely to prefer face-to-face. Using only one mode of administration may have a differential impact on the ability to recruit from particular subsets of the Australian population. The ability to conduct online interviews allows a broader geographical sample to be considered in the research. Face-to-face interviews in this study were substantially more expensive to conduct due to venue hire, interviewer travel time, interviewers' spare capacity whilst at venues and the higher participant reward to compensate for participant travel time and expenses. Offering both online and face-to-face interviews routinely allows those who have a strong preference to conduct their interview in that format and allows all other participants to select the most convenient option.

Mode of administration did not appear to have an impact on mean cTTO values. These values were shown to be statistically equivalent (based on an equivalence limit of 0.05) for the mild health states for which this study was powered. It was not possible to demonstrate equivalence for moderate and severe health states, which have greater variance. To establish equivalence for the more serious health states would require a much larger study. However, no statistically significant difference of mean cTTO values, standard deviations or median values between modes were established, and there was no consistent direction for the small mean differences that did exist by mode. Further equivalence research for more severe health states may therefore be overly costly and unnecessary.

Credit author statement

Tessa Peasgood: Methodology, Investigation, Formal analysis, Writing – original draft, Writing – review & editing, Funding acquisition. Mackenzie Bourke: Investigation, Formal analysis, Writing – original draft, Writing – review & editing. Nancy Devlin: Methodology, Writing – review & editing, Funding acquisition. Donna Rowen: Conceptualization, Methodology, Writing – review & editing, Funding acquisition. Yaling Yang: Conceptualization, Methodology, Writing – review & editing, Funding acquisition. Kim Dalziel: Methodology, Investigation, Writing – review & editing, Funding acquisition.

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 Research Foundation. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

Data availability

Data will be made available on request.

Acknowledgements

We would like to thank Elia Bidgood, Gayanda Madurapperuma and Alex van Heusden for their enormous contribution in conducting the

interviews. We extend thanks to all research participants who took part in the study. We acknowledge the contribution of the EuroQol office and the EQ-VT team for their technical support.

Appendix. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.socscimed.2023.115818.

References

Brazier, J., Ratcliffe, J., Saloman, J., Tsuchiya, A., 2017. Measuring and Valuing Health Benefits for Economic Evaluation. OXFORD university press.

Coretti, S., Ruggeri, M., McNamee, P., 2014. The minimum clinically important difference for EQ-5D index: a critical review. Expert Rev. Pharmacoecon. Outcomes Res. 14 (2), 221–233.

Davies, L., LeClair, K.L., Bagley, P., Blunt, H., Hinton, L., Ryan, S., et al., 2020. Face-to-Face compared with online collected accounts of health and illness experiences: a scoping review. Qual. Health Res. 30 (13), 2092–2102.

Estévez-Carrillo, A., Dewilde, S., Oppe, M., Ramos-Goñi, J.M., 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. Patient 1–15.

Finch, A.P., Meregaglia, M., Ciani, O., Roudijk, B., Jommi, C., 2022. An EQ-5D-5L value set for Italy using videoconferencing interviews and feasibility of a new mode of administration. Soc. Sci. Med. 292, 114519.

Gao, F., Ng, G.Y., Cheung, Y.B., Thumboo, J., Pang, G., Koo, W.H., et al., 2009. The Singaporean English and Chinese versions of the EQ-5D achieved measurement equivalence in cancer patients. J. Clin. Epidemiol. 62 (2), 206–213.

Janssen, B.M.F., Oppe, M., Versteegh, M.M., Stolk, E.A., 2013. Introducing the composite time trade-off: a test of feasibility and face validity. Eur. J. Health Econ. : HEPAC : Health Econ. Prevent. Care 14 (Suppl. 1), S5–S13.

Jensen, C.E., Sørensen, S.S., Gudex, C., Jensen, M.B., Pedersen, K.M., Ehlers, L.H., 2021. The Danish EQ-5D-5L value set: a hybrid model using cTTO and DCE data. Appl. Health Econ. Health Pol. 19 (4), 579–591.

Jiang, R., Shaw, J., Mühlbacher, A., Lee, T.A., Walton, S., Kohlmann, T., 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.

Kennedy-Martin, M., Slaap, B., Herdman, M., van Reenen, M., Kennedy-Martin, T., Greiner, W., et al., 2020. Which multi-attribute utility instruments are recommended for use in cost-utility analysis? A review of national health technology assessment (HTA) guidelines. Eur. J. Health Econ. 21 (8), 1245–1257.

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

Lugnér, A.K., Krabbe, P.F.M., 2020. An overview of the time trade-off method: concept, foundation, and the evaluation of distorting factors in putting a value on health. Expert Rev. Pharmacoecon. Outcomes Res. 20 (4), 331–342.

Luo, N., Chew, L.H., Fong, K.Y., Koh, D.R., Ng, S.C., Yoon, K.H., et al., 2003. Do English and Chinese EQ-5D versions demonstrate measurement equivalence? An exploratory study. Health Qual. Life Outcome 1, 7.

McClure, N.S., Sayah, F.A., Xie, F., Luo, N., Johnson, J.A., 2017. Instrument-defined estimates of the minimally important difference for EQ-5D-5L index scores. Value Health 20 (1524–4733) (Electronic)):644–50.

Mulhern, B., Longworth, L., Brazier, J., Rowen, D., Bansback, N., Devlin, N., et al., 2013. Binary choice health state valuation and mode of administration: head-to-head comparison of online and CAPI. Value Health 16 (1), 104–113.

Norman, R., King, M.T., Clarke, D., Viney, R., Cronin, P., Street, D., 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., Devlin, N.J., van Hout, B., Krabbe, P.F., de Charro, F., 2014. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. Value Health 17 (4), 445–453.

Oppe, M., Rand-Hendriksen, K., Shah, K., Ramos-Goñi, J.M., Luo, N., 2016. EuroQol protocols for time trade-off valuation of health outcomes. Pharmacoeconomics 34 (10), 993–1004.

Pickard, A.S., Law, E.H., Jiang, R., Pullenayegum, E., Shaw, J.W., Xie, F., et al., 2019. United States valuation of EQ-5D-5L health states using an international protocol. Value Health 22 (8), 931–941.

Purba, F.D., Hunfeld, J.A., Iskandarsyah, A., Fitriana, T.S., Sadarjoen, S.S., Passchier, J., et al., 2017. Employing quality control and feedback to the EQ-5D-5L valuation protocol to improve the quality of data collection. Qual. Life Res. 26 (5), 1197–1208.

Ramos-Goñi, J.M., Oppe, M., Slaap, B., Busschbach, J.J., Stolk, E., 2017. Quality control process for EQ-5D-5L valuation studies. Value Health 20 (3), 466–473.

Rowen, D., Mukuria, C., Bray, N., Carlton, J., Longworth, L., Meads, D., et al., 2022. Assessing the comparative feasibility, acceptability and equivalence of videoconference interviews and face-to-face interviews using the time trade-off technique. Soc. Sci. Med. 309, 115227.

Schuirmann, D.J., 1987. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. J. Pharmacokinet. Biopharm. 15 (6), 657–680.

Shah, K.K., Lloyd, A., Oppe, M., Devlin, N.J., 2013. One-to-one versus group setting for conducting computer-assisted TTO studies: findings from pilot studies in England and The Netherlands. Eur. J. Health Econ. 14 (Suppl. 1), S65–S73.

Shehata, E., 2011. TOBITHETM: Stata Module to Estimate Tobit Multiplicative Heteroscedasticity Regression.

Stolk, E., Ludwig, K., Rand, K., van Hout, B., Ramos-Goñi, J.M., 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.

Xie, F., Pickard, A.S., Krabbe, P.F.M., Revicki, D., Viney, R., Devlin, N., et al., 2015. A checklist for reporting valuation studies of multi-attribute utility-based instruments (CREATE). Pharmacoeconomics 33 (8), 867–877.