



# 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

Anabel Estévez-Carrillo<sup>1</sup> · Sarah Dewilde<sup>2</sup> · Mark Oppe<sup>1</sup> · Juan M. Ramos-Goñi<sup>1</sup>

Accepted: 16 January 2022 / Published online: 16 February 2022  
© The Author(s) 2022

## Abstract

**Background** Face-to-face interviews are recommended for the collection of composite time trade-off (cTTO) data. The coronavirus disease 2019 (COVID-19) social distancing measures made researchers consider videoconferencing interviews as an alternative mode of administration, but little evidence about the implementation of videoconferencing in valuation studies is available. This study provides insights into the effect of videoconferencing on the quality of data, evaluating interviewers' and respondents' engagement level in videoconferences compared with face-to-face interviews.

**Methods** We used cTTO data collected in Belgium and Spain following the EQ-5D-Y-3L valuation protocol. Due to the COVID-19 pandemic, both projects interrupted the face-to-face data collection before reaching the target sample. The remaining interviews were conducted by videoconference. We compared both modes of administration in terms of interviewers' engagement (task duration and number of moves in each example) and respondents' engagement (task duration and proportions of specific response values, in half-year units). To minimise interviewers' learning effects, we split our sample into three groups: (1) first 20 interviews conducted face-to-face; (2) subsequent interviews conducted face-to-face; and (3) videoconferencing interviews.

**Results** The comparison between videoconferencing and subsequent face-to-face interviews showed the interviewer's engagement was not affected by the mode of administration as almost no significant results were found either in the task duration or the numbers of moves shown in the examples. Similarly, none of the respondents' task duration or proportion of specific responses or half-year units were affected by the mode of administration in either of the two countries.

**Conclusions** No evidence suggested that the quality of cTTO data is reduced when using videoconferencing compared with face-to-face interviews.

## Key Points for Decision Makers

The coronavirus disease 2019 (COVID-19) pandemic caused researchers to consider the use of videoconferencing interviews to collect cTTO data. There is little evidence about the quality of cTTO data from videoconferencing interviews in comparison with face-to-face interviews.

We have provided insights into the performance of videoconferencing interviews in comparison with face-to-face interviews, in terms of interviewer and respondent engagement.

No evidence suggests that the quality of cTTO data decreases when using videoconferencing compared with face-to-face interviews.

✉ Juan M. Ramos-Goñi  
juanmanuel.ramosgoni@gmail.com

<sup>1</sup> Maths in Health, Rotterdam, The Netherlands

<sup>2</sup> Services in Health Economics, Brussels, Belgium

## 1 Introduction

Time trade-off (TTO) is one of the most widely used preference elicitation methods for valuing health states. It is a matching task for obtaining utility values, which can be operationalised in numerous ways. Some of the main versions of TTO that are being used have been well-described elsewhere [1–3]. In general, in a TTO task, respondents are asked to make trade-offs between the length of life and quality of life. The goal of the TTO is to obtain answers from respondents on how much time they would be willing to give up (i.e., trade off) to avoid having health impairments; the worse the health impairments become, the more time respondents are willing to give up. As such, the amount of time given up can be used as a measure for the severity of the health impairment. In TTO tasks, respondents are typically asked to answer a series of questions wherein they are forced to choose between a longer life with health impairments and a shorter life in full health. An iterative sequence is used to vary the amount of time shown in the life with full health until a point of indifference is reached. Respondents might be inclined to view very severe health impairments as ‘worse-than-dead’. TTO typically includes a worse-than-dead adapted task.

It is assumed that TTO tasks are not easy to understand for respondents. This is also the case for the composite TTO (cTTO), the version adopted by the EuroQol Group to elicit values for the EQ-5D-5L and EQ-5D-Y-3L instruments [4–6]. cTTO tasks can be abstract and confrontational, given the trade-offs between the length of life and quality of life. In addition, the task itself is complex, with the iterative procedure and separate tasks for ‘better-than-dead’ and ‘worse-than-dead’, which is why face-to-face interviews have traditionally been used for the collection of cTTO data and why the EuroQol Group has developed interviewer scripts used in all EQ-5D-5L valuation studies [7]. Face-to-face cTTO interviews have been shown to be reliable and to yield high-quality data [5, 7]. However, online self-completed TTO experiments have already been conducted. The results of one study have shown that online self-completed cTTO interviews induce a systematic downward bias [8]. Another study reported that TTO tasks conducted online and read by an automated voice had more inconsistencies and decreased engagement [9]. Thus, given the complexity of a cTTO task, the role of the interviewer is crucial. More specifically, the interviewer’s role consists of demonstrating the task to the respondent by using examples (being in a wheelchair in the EQ-5D valuation studies). The interviewer should explain all elements of the task and show how to perform it in terms of the iterative procedure used to vary the time given up when the respondent makes their own choices. In addition, the interviewer should show both the ‘better-than-dead’

and ‘worse-than-dead’ sides of the task. For these reasons, in a cTTO valuation study, it is important that both interviewers and respondents are properly engaged to gain high-quality data [7].

The coronavirus disease 2019 (COVID-19) pandemic has elicited a need to innovate in terms of how to conduct these interviews, as the face-to-face approach became infeasible due to social restrictions imposed in many jurisdictions. The social distancing measures made researchers consider using videoconferencing software (e.g., Skype or Zoom) as an alternative mode of administration to collect the necessary cTTO data. The videoconference-based administration approach was seen as having the potential to produce cTTO data with similar quality to that from traditional face-to-face interview methods, as an interviewer would be present to explain and guide the cTTO exercises. As far as we are aware, only Lipman [10] has discussed the advantages and disadvantages of cTTO interviews conducted by video conference, in comparison with face-to-face interviews, using empirical data from a study in The Netherlands. Lipman reported that a videoconferencing approach would have several advantages over a face-to-face approach, including greater geographical reach, additional convenience to respondents, lower study cost, and more rapid data collection. However, video conference-based collection may be affected by selection bias, such as the differential internet access among segments of the population [11], as well as the experience and ease of speaking over a computer. The conclusions of Lipman are limited to one interviewer and one country, thus limited evidence about the performance of videoconferencing interviews is available.

The aim of this manuscript was to provide insights into the performance of videoconferencing interviews in comparison with the performance of face-to-face interviews, based on interviewer and respondent engagement in cTTO task under each approach.

## 2 Methods

### 2.1 The EQ-5D-Y-3L Valuation Protocol

The data used for this study were collected as part of the national EQ-5D-Y-3L valuation studies in Belgium and Spain. Both studies used the EQ-5D-Y-3L instrument, which consists of five dimensions (mobility, looking after oneself, usual activities, pain/discomfort, and worry/sadness/unhappiness) with three levels each: ‘no problems’, ‘some/a bit’ and ‘a lot/very much’, usually coded from 1 to 3. In this paper, a health state ‘profile’ describes a health state by the levels of its five dimensions. For

example, ‘22222’ means some problems or a bit of problems occur in all five dimensions, and ‘12222’ means no problems occur in mobility and some problems or a bit of problems occur in the other four dimensions. The protocol used in both countries was the cTTO component of the international EQ-5D-Y-3L valuation protocol [6]. This protocol also included an online DCE component; however, the focus of our analysis is only the cTTO side. Briefly, the protocol recommends the use of a face-to-face cTTO to obtain preferences for EQ-5D-Y-3L states in a sample of 200 respondents. The main cTTO tasks included 10 health states: 3 mild, 2 moderate, and 5 severe, as suggested by the EQ-5D-Y-3L valuation protocol (Mild: 11112, 11121, 21111; Moderate: 22223, 22232; and Severe: 31133, 32223, 33233, 33323, 33333). All health states were valued from the perspective of a 10-year-old child.

A cTTO task involves a series of choices between two alternative lives: life A, describing full health for  $t$  years; and life B, describing an impaired health state for 10 years, where  $t$  is  $\leq 10$  years (Fig. 1a). To allow for worse-than-dead preferences, the task is adapted so that life B now consists of 10 years in full health (so-called lead time: LT) followed by 10 years in the impaired health state. The trade-offs are then made between  $t$  years in life A (where  $t$  is again  $\leq 10$  years) and life B with a total of 20 years, 10 of which are spent in the impaired health state (Fig. 1b). The cTTO uses an iterative sequence to guide the respondent to his point of indifference between living life A or life B (i.e., valuing life A and life B the same) by varying  $t$  based on the answers provided by the respondent (Fig. 1c).

## 2.2 Composite Time Trade-Off (cTTO) Interview Structure

The cTTO interviews were structured as follows:

- (1) The information sheet stated the aims of the study and requested a participant’s consent. If a respondent did not provide consent, the survey was terminated and only showed a message thanking the respondent for considering participation.
- (2) Demographic questions asked about their geographical area, age, and sex to delimit quotas and ensure sample representativeness.
- (3) The self-reported EQ-5D-Y-3L instrument asked about their health and was administered as a warm-up task.
- (4) Three questions asked about their experience with illness.
- (5) Two adaptive wheelchair examples were presented to allow the interviewer to explain the cTTO task.

Depending on whether the first wheelchair example was used to explain the better-than-dead or the worse-than-dead side of the task, the second wheelchair example was adapted to explain the other side.

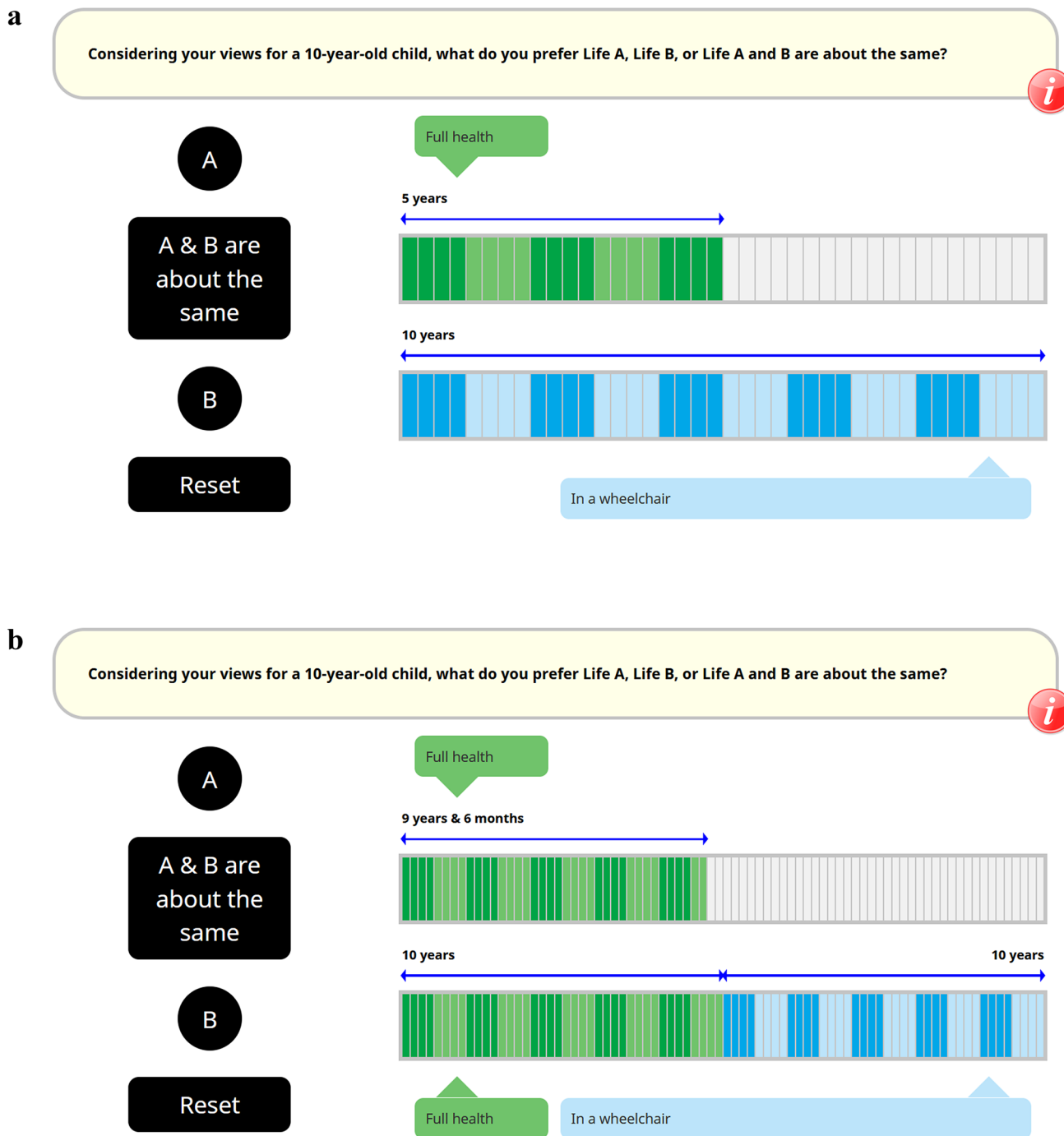
- (6) Three practice states were provided, to allow participants to practice alone before the main tasks.
- (7) The main 10 cTTO states (the same for all participants) were presented in random order.
- (8) The standard feedback module (FBM) was used as in prior EQ-5D-5L valuation studies [13]. Participants were presented with the rank ordering of the health states based on their 10 cTTO valuations and asked to exclude health state valuations they felt did not have the appropriate location in the ranking.
- (9) Additional background questions were asked about their educational level and whether respondents had children.

This interview was electronically implemented in the Euro-QoL Valuation Technology (EQ-VT), a web-based technology allowing for storage of all clicks and times between clicks that the interviewer used to demonstrate the cTTO wheelchair examples, or within the respondent’s choices in the practice and main health state tasks across the whole interview.

## 2.3 Sampling and Data Collection

The studies in Belgium and Spain were conducted independently, and both aimed to collect 200 cTTO interviews each, as suggested by the EQ-5D-Y-3L valuation protocol [6]. The face-to-face data collection was conducted in March 2020 in Spain and between August and October 2020 in Belgium, while the videoconferencing data collection was conducted between July and August 2020 in Spain and between October and December 2020 in Belgium. Both projects were set up to involve only face-to-face interviews, and recruitment was handled via two different market agencies.

Due to COVID-19 social distancing measures, the face-to-face data collection was interrupted at 123 interviews in Spain and 121 interviews in Belgium. Since respondents’ and interviewers’ health was still at risk, the research team agreed to try collecting the remaining interviews needed to reach the target sample size using videoconferencing software. In Spain, interviewers were inactive for 5 months (March 2020–July 2020) before restarting the data collection via Skype, while in Belgium, the remaining interviews were conducted via Zoom without stopping the data collection. Videoconferencing interviews were completed with audio and video connection and sharing of the interviewer’s screen. In both modes of administration, the interviewer operated the cTTO software, and the participants stated their preferences verbally.



**Fig. 1** cTTO task example. **a** Better than dead side; **b** worse than dead side; **c** iteration procedure. The iteration procedure used to vary  $t$  is described elsewhere [12], but briefly, it uses a ping-pong approach starting with  $t$  in 10 years ( $t = 10$  years in full health = ‘Life A’ and 10 years in the impaired health state = ‘Life B’) and moving to  $t = 0$  if a respondent chooses A. If the respondent then chooses B,  $t$  is increased to  $t = 5$  (a) followed by 1-year increments/decrements or 6-month increments/decrements depending on respondent’s

choices. If at  $t = 0$  the respondent chooses A, the worse-than-dead side of the task is shown (b), where  $t = 10$ . If the respondent chooses A again,  $t$  is decreased to  $t = 5$  followed by 1-year increments/decrements or 6-month increments/decrements depending on the respondent’s choices. Utilities shown in Fig. 1c for the impaired health states are calculated using  $t$  of the point of indifference:  $U = t/10$  for states considered better-than-dead, and  $U = (t-10)/10$  for states considered worse-than-dead. cTTO composite time trade-off

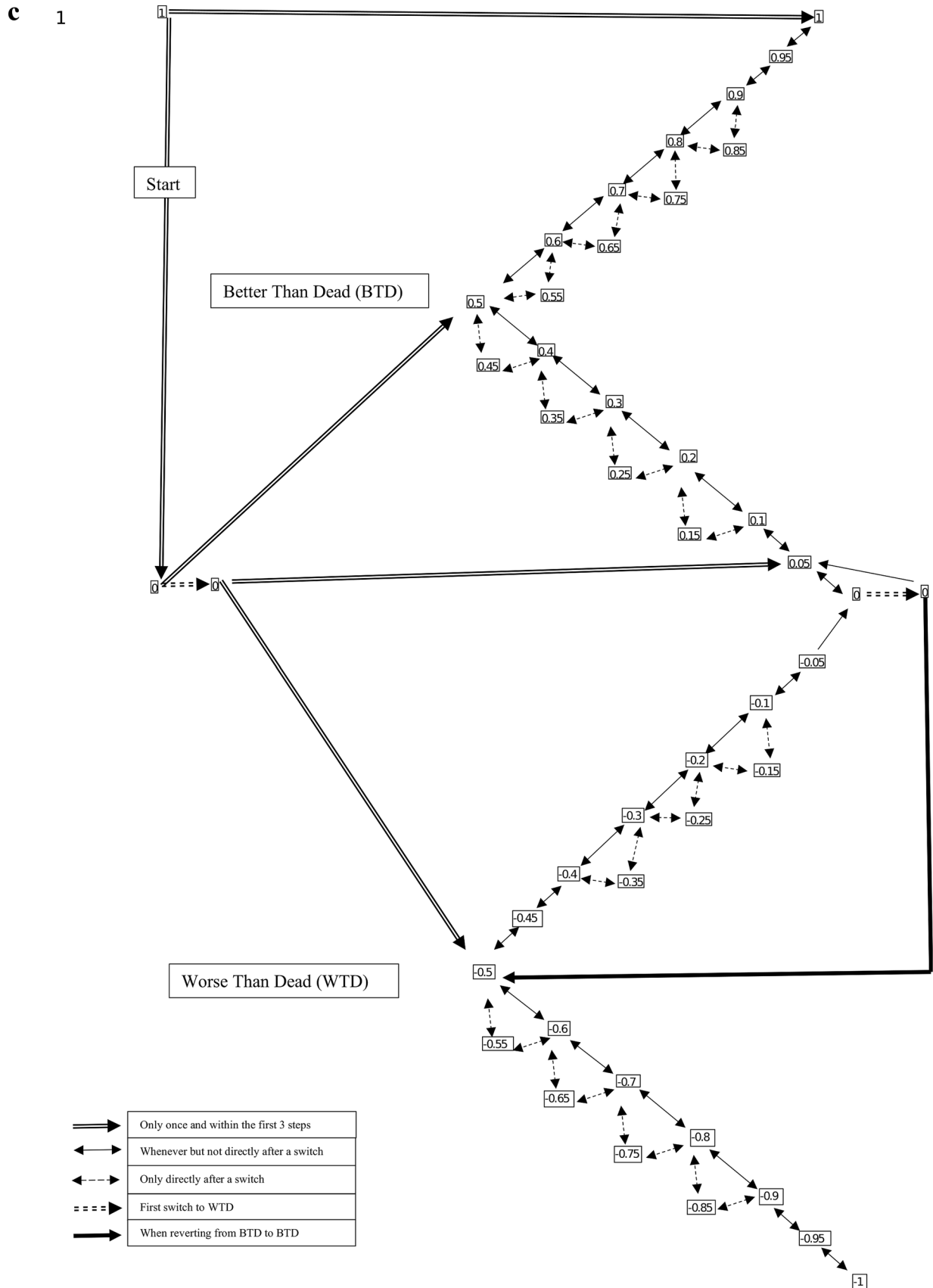


Fig. 1 (continued)

All interviewers were trained for this specific study and monitored on a weekly basis, following the standard quality control procedure of the EuroQol Group [7]. A total of four interviewers with previous experience conducting cTTO interviews participated in the Spanish data collection; however, one interviewer (interviewer 4) did not participate in the initial face-to-face data collection. The interviews conducted by interviewer 4 were therefore withdrawn from our analysis. In Belgium, the same three interviewers conducted both face-to-face and Zoom-based interviews.

## 2.4 Metrics Definition

As described by the EuroQol standard quality control methodology [7], and more recently by the quality assurance programme [14], we measured both interviewer and respondent engagement as follows:

Interviewer engagement pertains to how well they explain the cTTO task to respondents, which is measured to ascertain whether the quality control procedure also worked in the videoconferencing environment. Therefore, the following were measured:

1. The amount of time (in seconds) on each cTTO wheelchair example. Short task duration explaining the task indicates poor engagement, as the proper explanation of a cTTO task requires some time. Based on previous studies [7], evidence suggests that about 5 min is necessary for this task.
2. The number of moves from the iterative procedure used to complete each cTTO wheelchair example. Few moves would mean poor engagement, as the way that  $t$  (i.e., years in life A) is varied by the EuroQol cTTO iteration procedure may be difficult to understand for respondents. Thus, several moves of the procedure must be shown to properly explain the task. Based on previous studies, evidence suggests that 30–40 moves are required.
3. The moves performed in the better-than-dead and worse-than-dead elements of the wheelchair examples. In order to learn whether both sides of the task were explained, we split the analysis of moves. Evidence suggests that the better-than-dead side is usually explained first, and that the split between moves shown for the better-than-dead task and worse-than-dead task should be about 75% and 25%, respectively [7].

To measure respondent engagement, we focused on the following parameters:

1. The amount of time taken in seconds to complete each TTO task (excluding wheelchair examples and practice

cTTO tasks). Respondents require time to make multiple choices to reach their indifference point on the iteration procedure. Short task duration is associated with poor engagement, as speeding through the task or stopping early in the iteration procedure tends to generate imprecise responses. Evidence suggests that about 1 min is necessary for conducting a single cTTO task [7].

2. The proportion of values on specific responses of the iteration procedure. The specific responses are defined as those that only required a few moves before ending the task. More specific responses therefore mean lower engagement. Those specific responses corresponded to the following numbers: 1, 0.5, 0, – 0.5.
3. The proportion of responses in half-year units. To produce half-year-unit responses, respondents must make more of an effort than they must make to reach year responses; therefore, more half-year units were associated with higher engagement.
4. The proportion of negative values. Since the worse-than-dead side of the task was more difficult to understand than the better-than-dead side and also required more steps in order to arrive at a final response, respondents may be reluctant to value a health state as worse-than-dead. Thus, a lower proportion of worse-than-dead responses may be associated with poorer engagement.

Note that the respondent engagement was compared at the aggregate (i.e., interviewer) level and not at the individual respondent level. This is because a degree of variability between individual respondents due to differences in characteristics such as age is to be expected. However, when comparing at the interviewer level, it highlights whether interviewers were able to properly engage their respondents. To add clarity for points 2–4 of respondent engagement, we also looked at the value distribution for all responses. Additionally, a combined measure of both interviewer and respondent engagement is the duration of the whole interview. Shorter interview durations were associated with poorer engagement.

Furthermore, we looked at the face validity of the cTTO data and the interviewer's effects on values. To determine face validity of the cTTO data, we used the proportion of respondents per interviewer that provided inconsistent values (e.g., providing a higher cTTO value for a logically worse health state). The proportion of respondents whose cTTO data contained at least one inconsistency before and after the FBM (removing responses that were flagged by respondents as not valid) was examined across all health states and for the worst EQ-5D-Y-3L health state. We used strict and weak criteria for calculating inconsistencies. To determine interviewer effects, we examined the value distribution of overall health states.

## 2.5 Statistical Comparison between Modes of Administration

The mode of administration of the interview (videoconferencing vs. face-to-face) was not randomised. Videoconferencing cTTO interviews were conducted later than face-to-face interviews in both countries; thus, interviewers were more experienced during the videoconferencing interviews. We assumed that the first 20 interviews conducted by each interviewer were the most affected by potential interviewers' learning effects, based on previous experience in EQ-5D valuation studies [7]. To reduce the influence of potential interviewer learning effects in the mode of administration comparison, we split our sample into the following three groups: (1) the first 20 face-to-face interviews conducted by each interviewer; (2) the subsequent face-to-face interviews; and (3) the videoconferencing interviews.

Sample characteristics were reported using proportions in each subpopulation group categorised by the comparison groups defined above. Task durations and moves were reported using means and standard deviations. Histograms were used to report value distributions, and the proportions of specific values, half-year values, and negative values are reported in the corresponding figures. Finally, inconsistent respondents were reported using proportions. We used a Z-test for all proportion comparisons, while the unequal variance unpaired *t*-test was used for comparing task durations and number of moves on our statistical analysis. We reported statistically significant results with and without Bonferroni correction at a 95% confidence level.

## 3 Results

### 3.1 Descriptive Statistics

In Belgium, 218 cTTO interviews were conducted in total: 120 face-to-face and 98 via videoconferencing. In Spain, 16 interviews conducted by one interviewer who conducted only videoconferencing interviews were excluded. This resulted in 184 interviews in total, of which 123 were face-to-face and 61 were conducted via videoconferencing. The face-to-face interviews were divided into two groups: the first 20 interviews per interviewer and their subsequent interviews (60 and 63, respectively, in Spain; 60 and 60, respectively, in Belgium). The videoconference interviews were not split between first and subsequent interviews as the sample size was too small (61 in Spain and 98 in Belgium).

In Spain, videoconference interviews included a lower proportion of respondents with a university degree compared

with the subsequently conducted face-to-face interviews (63.5% vs. 42.6%), whereas in Belgium there was an imbalance in the proportion of females, which was higher among the videoconference respondents (40% vs. 70.4%) (Table 1).

### 3.2 Interviewer Engagement

Both task duration and moves spent on the wheelchair example appeared higher for videoconferencing interviews in both countries or were not affected by mode of administration, as almost no significant results were found between videoconferencing interviews and subsequent face-to-face interviews (Table 2). As expected, learning effects were present in both studies, as shown by the significant difference in results when comparing first face-to-face interviews with subsequent face-to-face interviews (Table 2).

### 3.3 Respondent Engagement

The proportion of values on specific responses or in half-year units were not affected by the mode of administration in either of the two countries. The observed time per each main cTTO task was higher in videoconferencing interviews than in subsequent face-to-face interviews in Spain. In Belgium, the task durations and proportions of values on specific responses or in half-year units remained relatively unchanged between subsequent face-to-face interviews and videoconferencing interviews (Table 2). There was an observed increase of worse-than-dead values in Spain and Belgium in the videoconferencing interviews (Fig. 2), although neither was significant. Both the proportion of responses ending on specific responses and the proportion of responses ending in half-year responses remained stable in Belgium between the modes of administration, whereas this varied more in Spain (Fig. 2). Further analyses of respondent values by interviewer showed insignificant differences between interviewers in Spain (Fig. 3). Interviewer 1 increased the observed proportion of half-year units and interviewer 3 remained stable when comparing their results on subsequent face-to-face versus videoconferencing interviews. Respondents from interviewer 2 produced the lowest observed proportion of half-year units in the videoconferencing interviews. The results by mode of administration were similar across all interviewers in Belgium (Fig. 3).

Regarding interviewer/respondent engagement, the observed interview duration was higher for videoconferencing interviews in both countries when comparing subsequent face-to-face and videoconferencing interviews (Table 2); however, these results were significant only in Spain.

**Table 1** Respondents' demographic by mode of administration.

	Spain [ <i>N</i> = 184]			Belgium [ <i>N</i> = 218]		
	First 20 F2F [ <i>n</i> = 60]	Subsequent F2F [ <i>n</i> = 63]	Videoconfer-ence [ <i>n</i> = 61]	First 20 F2F [ <i>n</i> = 60]	Subsequent F2F [ <i>n</i> = 60]	Videoconference [ <i>n</i> = 98]
Age, years [mean (SE)]	41.5 (1.7)	42.6 (1.9)	47.9 (2.0)	44.7 (1.9)	44.8 (2.1)	40.7 (1.3)
Age groups, years						
18–24	6 (10)	10 (15.9)	6 (9.8)	5 (8.3)	3 (5)	7 (7.1)
25–29	7 (11.7)	6 (9.5)	3 (4.9)	5 (8.3)	4 (6.7)	14 (14.3)
30–39	15 (25)	10 (15.9)	6 (9.8)	15 (25)	13 (21.7)	32 (32.7)
40–49	13 (21.7)	15 (23.8)	24 (39.3)	12 (20)	10 (16.7)	20 (20.4)
50–59	14 (23.3)	11 (17.5)	7 (11.5)	10 (16.7)	9 (15)	14 (14.3)
60–69	4 (6.7)	8 (12.7)	7 (11.5)	10 (16.7)	11 (18.3)	9 (9.2)
70+	1 (1.7)	3 (4.8)	8 (13.1)	3 (5)	10 (16.7)	2 (2)
Sex						
Male	23 (38.3)	31 (49.2)	27 (44.3)	30 (50)	36 (60)	29 (29.6)
Female	37 (61.7)	32 (50.8)	34 (55.7)	30 (50)	<b>24 (40)</b>	<b>69 (70.4)</b>
Education						
University degree	41 (68.3)	<b>40 (63.5)</b>	<b>26 (42.6)</b>	47 (78.3)	46 (76.7)	82 (83.7)
Have children						
Yes	32 (53.3)	29 (46)	37 (60.7)	39 (65)	34 (56.7)	59 (60.2)
Experience with illness						
Personal (% yes)	20 (33.3)	22 (34.9)	20 (32.8)	13(21.7)	15(25.0)	13 (13.3)
Relatives (% yes)	48 (80)	<b>46 (73)</b>	<b>54 (88.5)</b>	35 (58.3)	45 (75.0)	75 (76.5)
Others (% yes)	19 (31.7)	26 (41.3)	20 (32.8)	22 (36.7)	<b>28 (46.7)</b>	<b>24 (24.5)</b>
EQ-5D-Y-reported problems						
Mobility						
No problems	55 (91.7)	53 (84.1)	52 (85.2)	53 (88.3)	54 (90)	89 (90.8)
Some problems	4 (6.7)	9 (14.3)	9 (14.8)	7 (11.7)	6 (10)	7 (7.1)
A lot of problems	1 (1.7)	1 (1.6)	0 (0)	0 (0)	0 (0)	2 (2)
Looking after yourself						
No problems	60 (100)	60 (95.2)	57 (93.4)	59 (98.3)	58 (96.7)	97 (99)
Some problems	0 (0)	2 (3.2)	4 (6.6)	1 (1.7)	2 (3.3)	0 (0)
A lot of problems	0 (0)	1 (1.6)	0 (0)	0 (0)	0 (0)	1 (1)
Usual activities						
No problems	57 (95)	52 (82.5)	55 (90.2)	55 (91.7)	53 (88.3)	87 (88.8)
Some problems	3 (5)	<b>10 (15.9)</b>	6 (9.8)	5 (8.3)	7 (11.7)	10 (10.2)
A lot of problems	0 (0)	1 (1.6)	0 (0)	0 (0)	0 (0)	1 (1)
Pain/discomfort						
No pain or discomfort	48 (80)	48 (76.2)	40 (65.6)	41 (68.3)	38 (63.3)	61 (62.2)
Some pain or discomfort	11 (18.3)	14 (22.2)	21 (34.4)	15 (25)	20 (33.3)	34 (34.7)
A lot of pain or discomfort	1 (1.7)	1 (1.6)	0 (0)	4 (6.7)	2 (3.3)	3 (3.1)
Worry/unhappy						
Not worried, sad, or unhappy	37 (61.7)	42 (66.7)	35 (57.4)	38 (63.3)	32 (53.3)	45 (45.9)
A bit worried, sad, or unhappy	22 (36.7)	20 (31.7)	25 (41)	20 (33.3)	26 (43.3)	50 (51)
Very worried, sad, or unhappy	1 (1.7)	1 (1.6)	1 (1.6)	2 (3.3)	2 (3.3)	3 (3.1)

We compared only the first 20 F2F interviews vs. subsequent F2F interviews, and subsequent F2F vs. videoconferencing interviews

F2F face-to-face, SE standard error

Statistically significant results at the 95% confidence interval from the proportions test are italicized, and bold italicized font indicates significant results after Bonferroni correction applied



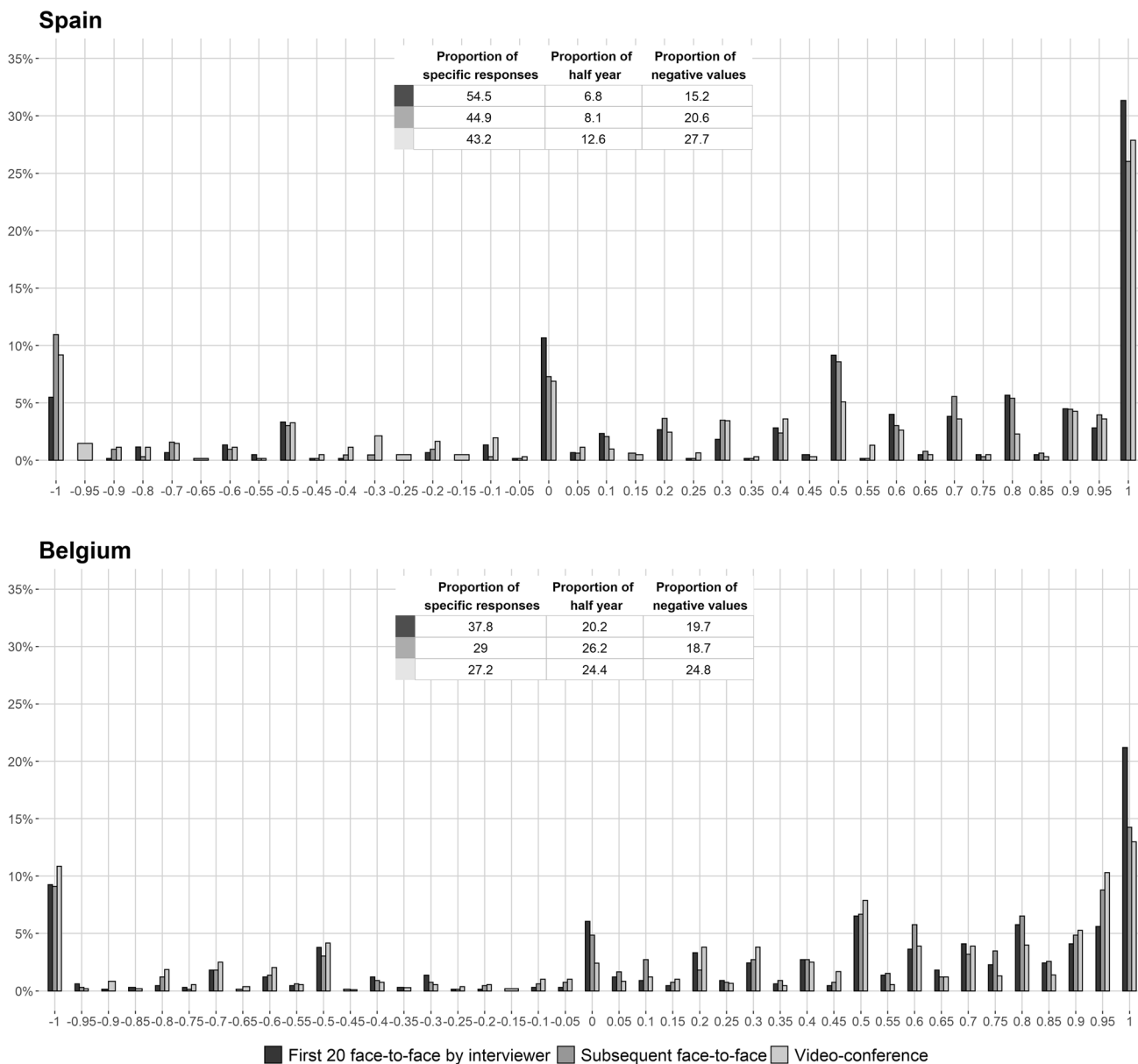
**Table 2** Interviewer and respondent engagement

	Spain			Belgium		
	First 20 F2F	Subsequent F2F	Videoconference	First 20 F2F	Subsequent F2F	Videoconference
Time(s) on wheelchair, example 1						
Interviewer 1	311.6 (112.3)	244.2 (89.3)	289.3 (79.4)	390 (131.0)	343.9 (99.4)	363.9 (93.1)
Interviewer 2	286.2 (67.9)	306.2 (67.1)	304.4 (134.3)	332.2 (62.8)	370.3 (164.4)	368.1 (83.4)
Interviewer 3	251.7 (56.8)	<b>238.6 (35.7)</b>	<b>293.9 (96.3)</b>	<i>234.6 (65.0)</i>	<i>324.8 (164.3)</i>	300 (122.2)
Overall	283.1 (84.9)	275.8 (69)	295.8 (104.2)	320.4 (111.2)	347.3 (148.8)	341.7 (106.3)
Time(s) on wheelchair, example 2						
Interviewer 1	<i>250.2 (84.9)</i>	<i>184.4 (39.2)</i>	233.2 (111.8)	304.7 (109.2)	267.4 (61.8)	288.3 (92.1)
Interviewer 2	<i>236.7 (94.4)</i>	<b>192.3 (51.7)</b>	<b>258.2 (95.6)</b>	384.1 (97.12)	364.7 (84.5)	375 (70.0)
Interviewer 3	297.2 (55.2)	286.4 (39)	314.7 (76.1)	231.9 (193.5)	303.8 (122.1)	246.9 (114.2)
Overall	<b>261.4 (84.0)</b>	<b>222.6 (64.2)</b>	<b>265.9 (100.6)</b>	308.2 (150.1)	317.4 (102.1)	297.9 (108.0)
Moves on wheelchair, example 1						
Interviewer 1	18.4 (5.7)	<i>14.1 (3.2)</i>	<i>20.1 (7.8)</i>	28 (9.7)	<b>29.3 (9.5)</b>	<b>39.3 (14.9)</b>
Interviewer 2	29.0 (11.0)	30.5 (12.4)	26.8 (8.8)	<b>22.6 (10.9)</b>	<b>40 (12.0)</b>	<b>46.2 (6.1)</b>
Interviewer 3	27.9 (8.7)	25 (5.3)	23.6 (8.7)	11.6 (5.3)	15 (5.6)	14.1 (5.8)
Overall	25.1 (9.9)	26.6 (10.9)	23.4 (8.8)	<b>20.9 (11.2)</b>	<b>28.4 (14.2)</b>	32 (17.1)
Moves on wheelchair, example 2						
Interviewer 1	21.8 (7.2)	25.5 (8.1)	22.3 (8.6)	31.4 (20.8)	33.2 (16.8)	34.4 (15.1)
Interviewer 2	30.9 (17.6)	32.2 (8.5)	32.0 (6.7)	<b>27.2 (9.7)</b>	<b>41 (15.6)</b>	43.4 (8.2)
Interviewer 3	<b>28.2 (9.6)</b>	<b>17.5 (6.1)</b>	<b>23 (7.7)</b>	<i>13.6 (10.0)</i>	<i>19.95 (8.8)</i>	20 (9.9)
Overall	27 (12.7)	26.4 (10.1)	25.8 (8.8)	<b>24.2 (16.2)</b>	<b>31.5 (16.6)</b>	31.7 (14.9)
Moves on the better than dead element						
Interviewer 1	23 (20.3)	18 (15.2)	24.4 (22.2)	42.4 (35.6)	42.9 (35.6)	51.1 (43.6)
Interviewer 2	38.7 (32.2)	34.5 (31.4)	29.1 (28.1)	29.3 (26.6)	50.6 (44.4)	52.9 (47.2)
Interviewer 3	33.1 (29.9)	28 (25.8)	29.1 (26)	20.7 (17)	25.2 (20.6)	22.1 (19.1)
Overall	31.6 (28)	30.3 (28)	27.4 (25.5)	30.8 (27.4)	39.7 (35.3)	40.9 (37.8)
Moves on the worse than dead element						
Interviewer 1	17.2 (18.9)	21.6 (22.4)	18 (19)	17 (19.4)	19.6 (22.5)	22.6 (23.9)
Interviewer 2	21.2 (24.1)	28.1 (28.4)	29.7 (30.2)	20.5 (22.1)	30.5 (31.8)	36.7 (37.4)
Interviewer 3	23 (24.3)	14.6 (15.4)	17.5 (18.4)	5.1 (7.3)	9.8 (11.8)	12 (13.8)
Overall	20.5 (22.6)	22.8 (24.1)	21.9 (23.3)	14.2 (17.5)	20.3 (23.9)	22.8 (25.9)
Interview duration (minutes)						
Interviewer 1	37.2 (11.3)	<i>30.5 (7.5)</i>	<i>36.8 (7.0)</i>	40.4 (13.5)	40.1 (8.2)	40.1 (5.0)
Interviewer 2	38.1 (7.2)	<b>38.3 (6.3)</b>	<b>47.6 (12.5)</b>	47.4 (6.7)	46 (13.6)	47.2 (7.4)
Interviewer 3	32.1 (3.9)	<b>31.3 (3.5)</b>	<b>38.9 (6.3)</b>	37.2 (14.6)	39.4 (14.6)	43.8 (31.4)
Overall	35.8 (8.4)	<b>35 (6.6)</b>	<b>41.1 (10.1)</b>	41.7 (12.6)	42.1 (13.0)	43.5 (19.7)
Time per each TTO task						
Interviewer 1	54.7 (46.8)	39.9 (20.9)	64.6 (43.0)	65.9 (45.5)	66.2 (42.4)	57.1 (32.1)
Interviewer 2	60.8 (72.1)	61.7 (55.1)	82.9 (94.7)	85.3 (43.1)	75.4 (56.7)	73.8 (48.2)
Interviewer 3	64.4 (41.3)	59.4 (38.4)	72.5 (39.6)	48.8 (41.1)	65.2 (45.1)	62.4 (48.3)
Overall	59.9 (55.2)	58.1 (47.2)	73.2 (65.3)	(45.7)	67.8 (49.3)	63.8 (43.8)

We compared only the first 20 F2F vs. subsequent F2F, and subsequent F2F vs. videoconferencing interviews

*F2F* face-to-face, *TTO* time trade-off

Statistically significant results at the 95% confidence interval from the proportions test are italicized, and bold italicized font indicates significant results after Bonferroni correction applied



**Fig. 2** cTTO value distribution for all health states. \*There were not significant results at the 95% confidence interval from the proportions test. *cTTO* composite time trade-off

### 3.4 Face Validity and Value Distribution Interviewer's Effects

The observed proportion of inconsistent respondents was lower for videoconferencing interviews compared with the subsequent face-to-face interviews in Spain. In contrast, this proportion was higher for videoconferencing interviews in Belgium. However, these observed differences were not statistically significant (Table 3).

## 4 Discussion

This manuscript reports on the quality of cTTO data, in terms of interviewer and respondent engagement, collected via videoconferencing interviews. More specifically, we report the insights from two EQ-5D-Y-3L valuation studies in Belgium and Spain. We have defined the metrics to measure engagement and we have reported the results of these metrics. None of the defined outcomes for measuring

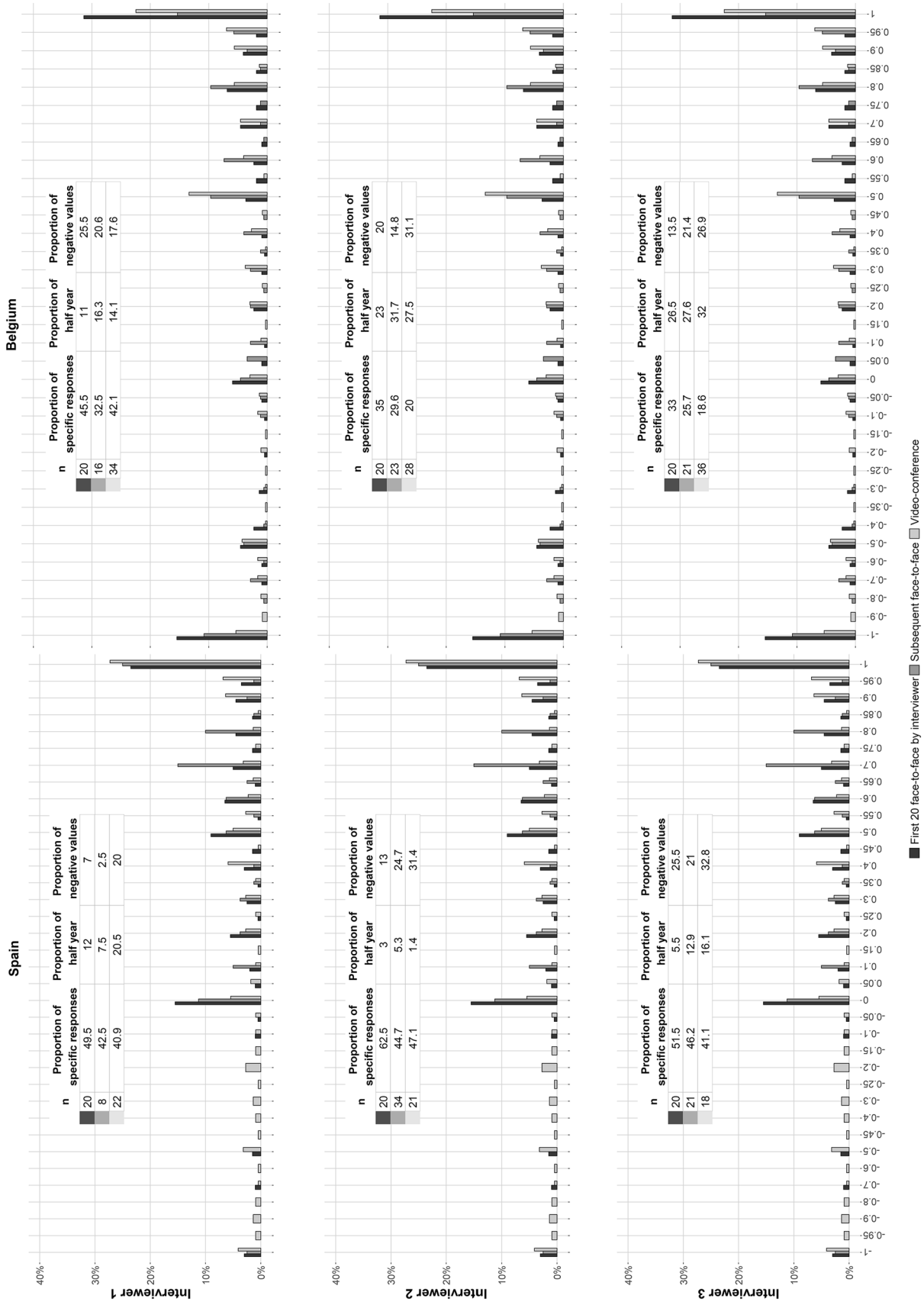


Fig. 3 cTTO value distribution for each interviewer. \*There were not significant results at the 95% confidence interval from the proportions test. cTTO composite time trade-off

**Table 3** Proportion of flagged health states and inconsistent respondents by interviewer and group

	Spain			Belgium		
	First 20 F2F	Subsequent F2F	Video conf.	First 20 F2F	Subsequent F2F	Video conf.
Proportion of flagged health states						
Interviewer 1	4.5	5.0	15.0	6.5	5.0	3.5
Interviewer 2	7.0	6.5	8.6	1.5	4.8	4.6
Interviewer 3	5.5	11.9	15.0	0.5	0.9	0.6
Overall	5.7	8.1	12.8	2.8	3.5	2.8
Proportion of respondents with inconsistent responses						
Interviewer 1						
Strict criterion						
Before FBM	40	50	45.5	55	43.75	32.4
After FBM	35	37.5	18.2	35	25	17.7
Weak criterion						
Before FBM	100	87.5	81.8	95	81.25	85.3
After FBM	95	75	45.5	90	75	73.5
Interviewer 2						
Strict criterion						
Before FBM	35	23.5	19.1	35	34.8	35.7
After FBM	10	11.8	0	30	21.7	25
Weak criterion						
Before FBM	100	94.12	90.5	70	65.2	75
After FBM	100	85.3	85.7	65	52.2	60.7
Interviewer 3						
Strict criterion						
Before FBM	35	<b>71.4</b>	61.1	25	19.1	36.1
After FBM	25	28.6	27.8	25	19.1	31.3
Weak criterion						
Before FBM	85	100	100	85	76.2	77.8
After FBM	80	<b>95.2</b>	<b>61.1</b>	85	76.2	77.8
Overall						
Strict criterion						
Before FBM	36.7	42.9	41.0	38.3	31.7	34.7
After FBM	23.3	20.6	14.8	30	21.7	25.5
Weak criterion						
Before FBM	95	95.24	90.2	83.3	73.3	79.6
After FBM	91.7	<b>87.3</b>	<b>63.9</b>	80	66.7	71.4
Proportion of respondents with inconsistent responses related to 33333						
Interviewer 1						
Strict criterion						
Before FBM	5	12.5	13.6	10	0	0
After FBM	5	0	4.5	0	0	0
Weak criterion						
Before FBM	55	37.5	45.5	55	43.8	26.5
After FBM	55	25	27.3	55	43.8	26.5
Interviewer 2						
Strict criterion						
Before FBM	0	0	0	10	0	3.5
After FBM	0	0	0	5	0	3.5
Weak criterion						

**Table 3** (continued)

	Spain			Belgium		
	First 20 F2F	Subsequent F2F	Video conf.	First 20 F2F	Subsequent F2F	Video conf.
Before FBM	70	70.6	76.2	35	17.4	39.3
After FBM	70	70.6	71.4	30	17.4	35.7
Interviewer 3						
Strict criterion						
Before FBM	<i>10</i>	<i>38.1</i>	<i>22.2</i>	<b>5</b>	<b>4.7</b>	<b>5.5</b>
After FBM	10	14.3	5.6	5	4.7	5.5
Weak criterion						
Before FBM	55	66.7	66.7	45	47.7	44.4
After FBM	55	38.1	50	45	47.7	44.4
Overall						
Strict criterion						
Before FBM	5.0	14.3	11.5	8.3	1.7	3.1
After FBM	5.0	4.8	3.3	3.3	1.7	3.1
Weak criterion						
Before FBM	60	65.1	62.3	45	35	36.7
After FBM	60	54	49.2	43.3	35	35.7

Strict criterion means providing a higher cTTO value for a logically worse health state (e.g., the cTTO value assigned to the health state 32223 is 0.2 and to 33233 is 0.3). Weak criterion means providing a higher or equal cTTO value for a logically worse health state (e.g., the cTTO value assigned to the health state 32223 is 0.2 and to 33233 is 0.3, or the cTTO value assigned to both health states is 0.3). We compared only the first 20 F2F vs. subsequent F2F, and subsequent F2F vs. videoconferencing interviews

F2F face-to-face, FBM feedback module, cTTO composite time trade-off

Statistically significant results at the 95% confidence interval from the proportions test are italicized, and bold italicized font indicates significant results after Bonferroni correction applied

interviewer and respondent engagement has suggested that worse results occur for videoconferencing interviews when compared with face-to-face interviews. All observed results show similar or higher engagement (i.e., higher data quality) for videoconferencing interviews. As a result, the quality of cTTO data in these valuation studies were not affected by introducing videoconferencing interviews.

Other researchers have conducted similar research. Lipman [10] has reported findings in line with our results, suggesting that videoconferencing interviews are feasible and produce values that are similar to those obtained from face-to-face interviews. In our analysis, we have focused on other aspects of the comparison; namely, differences with respect to engagement instead of differences of elicited values. Since our study was not designed as an experiment to compare face-to-face and videoconferencing modes of administration, but rather as national EQ-5D-Y-3L valuation studies, the study design put limitations on this comparison. Given our sample size and the nature of our cTTO data, it was more appropriate to look at engagement, as our non-controlled factors together with the potential effect of the pandemic on the study population's preferences could induce an unfair comparison of elicited values.

We did not control at the designing stage for other potential sources of variability; namely, interviewer's learning effects, respondent demographics, and potential changes to health preferences due to the COVID-19 pandemic. All of these factors may limit the generalisation of our findings. We tried to reduce the influence of these factors by dividing face-to-face interviews (the one affected by learning effects) into two groups (the first 20 and subsequent interviews). As shown in the analysis, this adjustment in sample subgroups was important but its cost is a reduction of the power of the comparison, which may have further limited our capacity to detect potentially real differences in the mode of administration.

As expected, interviewer learning effects were present in the two countries but at a lower level in Belgium than in Spain. In Spain, interviewers were inactive for 5 months before restarting their videoconferencing interviews, and one interviewer was retrained before resuming the videoconferencing interviews; therefore, they may have lost practice during this period, meaning some learning effects could have occurred during videoconferencing interviews as well as during the face-to-face interviews. This may explain the small variations in results between the

countries. However, testing specific learning effects in our cTTO data was not possible due to the small sample size. A recent EQ-5D-5L valuation study conducted by Finch et al. in Italy [15] was entirely completed using videoconferencing settings. Finch et al. reported initial quality control issues due to interviewers' learning effects; however, it cannot be disentangled whether those learning effects stem from the cTTO task learning or the videoconferencing environment.

Other sources of potential variability stem from the fact that the sampling recruitment was designed for gaining a representative sample of the Belgian and Spanish population in the EQ-5D-Y-3L valuation studies, but not between our comparison subgroups. This means that the sampling age/sex/location cells were not homogeneously filled out for face-to-face and videoconferencing interviews, which has produced several differences in the characteristics of the comparison samples. Samples homogeneity may also be affected by respondent's IT skills as recruitment for videoconferencing was limited by its feasibility. Given our sample size, further subgroup analysis to obtain any meaningful results was not possible. While this may be a significant issue for comparing elicited values, it is far less of an issue for comparing engagement. Indeed, we found a higher proportion of negative values for videoconferencing interviews. These results could indicate that respondents are more willing to share views that may not be socially desirable in videoconferencing interviews, as suggested by Lipman [10]. In our case, we cannot disentangle whether this results from participants being more honest in revealing their preferences in online interviews; from differences in population health preferences caused by the impact of the COVID-19 pandemic, as suggested by Webb et al. [16], or simply from differences between samples' characteristics.

cTTO videoconferencing interviews appear promising and have potential advantages, such as reducing costs or reaching population subgroups that are typically excluded from valuation studies. However, further research is needed on different aspects of how they compare with face-to-face interviews, as pointed out by Lipman [10]. This further research should use controlled experiments that assign the mode of administration randomly and use the respondent's own perspective rather than a 10-year-old-child perspective. They can be conducted (1) between subjects (requiring large sample sizes); or (2) within-subject, using test-retest designs where the order of the mode of administration is randomised. In addition, collecting information on interview cancellations, which mode of administration respondents would have preferred if they had a choice, interviewers' and respondents' experiences, or the videoconferencing software used would be of interest.

## 5 Conclusion

When looking at interviewer and respondent engagement, no evidence suggested that the quality of cTTO data is reduced when using videoconferencing compared with face-to-face interviews.

**Acknowledgements** The authors would like to express their gratitude to their funder, EuroQol Research Foundation; to Elly Stolk and Bernhard Slaap for supporting the funding of this study and collaborating when the authors were writing the paper; to Carlos Herrera-Carballo for his contribution in developing the surveys used to collect the data; and to our cTTO interviewers, who conducted all interviews with the greatest rigor to produce high-quality cTTO data. Finally, our greatest debt of gratitude goes to the respondents who participated in the cTTO experiments.

## Declarations

**Funding** This research was funded by the EuroQol Research Foundation (EQ project numbers 20180510 and 124-2020VS).

**Conflicts of interest/competing interests** JMRG, MO, and SD are members of EuroQol Group. The views expressed are those of the authors and not necessarily those of the EuroQol Group. Anabel Estévez-Carrillo has no conflicts of interest to declare.

**Ethics approval** The ethical approvals for the study were obtained from the Comité Ético de Investigación Clínica, Hospital Nuestra Señora del Prado, No. Dictamen 32/19, and from the University of Ghent, Belgium (ref B6702020000574).

**Consent to participate** Informed consent was obtained from all individual participants included in the study.

**Consent for publication** Not applicable.

**Code availability** All code for analysis is available from the authors upon a reasonable request.

**Availability of data and material** Data are available from the corresponding author upon reasonable request.

**Authors' contributions** JMRG, SD, and MO conceptualised and designed the original study. JMRG, SD, and AEC contributed to the data collection process. JMRG, SD, and AEC provided the study materials as well as technical and logistical support. JMRG and AEC conducted the statistical analysis and the interpretation of results. All authors contributed to writing the first draft of the manuscript. JMRG and SD secured funding for the study. AEC, SD, and JMRG supervised the work. All authors read and approved the final manuscript for submission.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative

Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

## References

- Attema AE, Edelaar-Peters Y, Versteegh MM, Stolk EA. Time trade-off: one methodology, different methods. *Eur J Health Econ.* 2013;14(Suppl 1):S53–64. <https://doi.org/10.1007/s10198-013-0508-x>.
- Oppe M, Rand-Hendriksen K, Shah K, Ramos-Goñi JM, Luo N. EuroQol protocols for time trade-off valuation of health outcomes. *Pharmacoeconomics.* 2016;34(10):993–1004. <https://doi.org/10.1007/s40273-016-0404-1>.
- Lugnér AK, Krabbe PFM. 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.* 2020;20(4):331–42. <https://doi.org/10.1080/14737167.2020.1779062>.
- Oppe M, Devlin NJ, van Hout B, Krabbe PF, de Charro F. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. *Value Health.* 2014;17(4):445–53. <https://doi.org/10.1016/j.jval.2014.04.002>.
- Stolk E, Ludwig K, Rand K, van Hout B, Ramos-Goñi JM. Overview, update, and lessons learned from the international EQ-5D-5L valuation work: version 2 of the EQ-5D-5L valuation protocol. *Value Health.* 2019;22(1):23–30. <https://doi.org/10.1016/j.jval.2018.05.010>.
- Ramos-Goñi JM, Oppe M, Stolk E, Shah K, Kreimeier S, Rivero-Arias O, et al. International valuation protocol for the EQ-5D-Y-3L. *Pharmacoeconomics.* 2020;38(7):653–63. <https://doi.org/10.1007/s40273-020-00909-3>.
- Ramos-Goñi JM, Oppe M, Slaap B, Busschbach JJ, Stolk E. Quality control process for EQ-5D-5L valuation studies. *Value Health.* 2017;20(3):466–73. <https://doi.org/10.1016/j.jval.2016.10.012>.
- Norman R, King MT, Clarke D, Viney R, Cronin P, Street D. Does mode of administration matter? Comparison of online and face-to-face administration of a time trade-off task. *Qual Life Res.* 2010;19(4):499–508. <https://doi.org/10.1007/s11136-010-9609-5>.
- Jiang R, Shaw J, Mühlbacher A, Lee TA, Walton S, Kohlmann T, et al. Comparison of online and face-to-face valuation of the EQ-5D-5L using composite time trade-off. *Qual Life Res.* 2021;30(5):1433–44. <https://doi.org/10.1007/s11136-020-02712-1>.
- Lipman SA. Time for tele-TTO? Lessons learned from digital interviewer-assisted time trade-off data collection. *Patient.* 2021;14(5):459–69. <https://doi.org/10.1007/s40271-020-00490-z>.
- Duffy B, Smith K, Terhanian G, Bremer J. Comparing data from online and face-to-face surveys. *Int J Mark Res.* 2005;47:615–39. <https://doi.org/10.1177/147078530504700602>.
- Ramos-Goñi JM, Craig BM, Oppe M, Ramallo-Fariña Y, Pinto-Prades JL, Luo N, et al. Handling data quality issues to estimate the Spanish EQ-5D-5L value set using a hybrid interval regression approach. *Value Health.* 2018;21(5):596–604. <https://doi.org/10.1016/j.jval.2017.10.023>.
- Wong ELY, Ramos-Goñi JM, Cheung AWL, Wong AYK, Rivero-Arias O. Assessing the use of a feedback module to model EQ-5D-5L health states values in Hong Kong. *Patient.* 2018;11(2):235–47. <https://doi.org/10.1007/s40271-017-0278-0>.
- Hernandez Alava M, Pudney S, Wailoo A. The EQ-5D-5L value set for England: findings of a quality assurance program. *Value Health.* 2020;23(5):642–8. <https://doi.org/10.1016/j.jval.2019.10.017>.
- Finch AP, Mereaglia M, Ciani O, Roudijk B, Jommi C. An EQ-5D-5L value set for Italy using videoconferencing interviews and feasibility of a new mode of administration. *Soc Sci Med.* 2022;292: 114519. <https://doi.org/10.1016/j.socscimed.2021.114519>.
- Webb EJD, Kind P, Meads D, Martin A. Does a health crisis change how we value health? *Health Econ.* 2021;30(10):2547–60. <https://doi.org/10.1002/hec.4399>.