The European Journal of Health Economics (2021) 22:17–33
https://doi.org/10.1007/s10198-020-01224-6

ORIGINAL PAPER

Increasing respondent engagement in composite time trade-off tasks by imposing three minimum trade-offs to improve data quality

Ruixuan Jiang[1] · Thomas Kohlmann[2] · Todd A. Lee[3] · Axel Mühlbacher[4] · James Shaw[5] · Surrey Walton[3] · A. Simon Pickard[3]

Received: 16 August 2019 / Accepted: 11 August 2020 / Published online: 28 August 2020

Abstract

Background Web-based surveys are increasingly utilized for health valuation studies but may be more prone to lack of engagement and, therefore, poor data validity. The objective of this study was to evaluate the effect of imposed engagement (i.e., at least three trade-offs) in the composite time trade-off (cTTO) task.

Methods The EQ-5D-5L valuation study protocol and study design were adapted for online, unsupervised completion in two arms: base case and engagement. Validity of preferences was assessed using the prevalence of inconsistent valuations and expected patterns of TTO values. Respondent task engagement was measured using time per task. Value sets were generated using linear regression with a random intercept (RILR).

Results The base case ($n=501$) and engagement arms ($n=504$) clustered at different TTO values: [base case] 0, 1; [engagement] -0.5, 0.45, 0.6. Mean TTO values were lower for the engagement arm. Engagement respondents did not spend more time per TTO task: [base case] 63.3 s (SD 77.9 s); [engagement] 64.7 s (SD 73.3 s); $p=0.36$. No significant difference was found between arms for prevalence of respondents with at least one inconsistent TTO value: [base case] 61.1%; [engagement] 63.5%; $p=0.43$. Both value sets had significant intercepts far from 1: [base case] 0.846; [engagement] 0.783. The relative importance of the EQ-5D dimensions also differed between arms.

Conclusions Both online arms had poor quality data. A minimum trade-off threshold did not improve engagement nor face validity of the data, indicating that modifications to the number of iterations are insufficient alone to improve data quality/validity of online TTO studies.

Introduction

The time trade-off (TTO) task is a direct preference elicitation method which determines the utility of a suboptimal health state by comparing time in the suboptimal health state to time in full health [1]. The time spent in full health is changed until the respondent feels the two lives are about the same. At this equivalence point, the respondent's choice can be interpreted as trading time in full health to avoid the suboptimal health state. The TTO is often used to elicit preferences for health states in the development of value sets for measures of health like the EQ-5D [2, 3]. Because the TTO can be difficult to understand, valuation studies using the TTO have historically been conducted using resource- and time-intensive face-to-face studies to ensure respondent comprehension [4, 5].

In recent years, an alternative approach to conducting valuation studies has emerged with the convenience of web surveys and online panels [6–8]. Online valuation studies can be conducted at a fraction of the cost and time compared to face-to-face studies. However, in the absence of an interviewer, respondents may not understand the task or be inattentive at the point of data collection [8–11]. Additionally, online panel respondents may feel incentivized to complete a greater number of surveys and speed through the tasks considering

✉ A. Simon Pickard
Pickard1@uic.edu

1 Center for Observational and Real-world Evidence, Merck & Co., 2000 Galloping Hill Rd., Kenilworth, NJ 07033, USA

2 Institute for Community Medicine, Medical University Greifswald, Ellernholzstraße 1/2, 17489 Greifswald, Germany

3 Department of Pharmacy Systems, Outcomes, and Policy, University of Illinois at Chicago College of Pharmacy, 833 S Wood St, Chicago, IL 60612, USA

4 Health Economics and Healthcare Management, Hochschule Neubrandenburg, Brodaer Str. 2, 17033 Neubrandenburg, Germany

5 Patient-Reported Outcomes Assessment, Bristol-Myers Squibb, 3551 Lawrenceville Rd, Princeton, NJ 08540, USA

payment is tied to each completed survey. Thought intensive tasks like the TTO may be particularly affected by these aspects of online data collection. In previous evaluations, online, unsupervised respondents were less engaged with TTO tasks and provided preferences that were of poorer quality and less valid than face-to-face respondents. Online respondents were more prone to providing TTO values at − 1, 0, and 1, as well as using fewer trade-offs per TTO task, on average [7, 12, 13]. A value set estimated from online data had a short range of scale (0.446) and the estimated utility for perfect health was 0.846, demonstrating inadequate face validity [13].

As online administration of the TTO was known to be susceptible to task short-cutting, data quality and thus validity may be improved by using approaches that enhance respondent attentiveness to or engagement with the task. One such strategy is the requirement of a minimum number of trade-offs to be completed in the iterative procedure prior to task completion [14]. Such a requirement may cause the respondent to more carefully consider the task and thus enhance data validity as compared to tasks without it. However, implementing the requirement may also significantly shift the observed preferences while respondents' underlying preference functions remained constant [15]. Preference patterns and the validity of elicited preferences should be compared to determine the impact of this requirement. The objective of this study was to test whether requiring a minimum of three trade-offs in a composite time trade-off task improved respondent task engagement and data quality and validity.

Methods

Measure of health

The EQ-5D-5L is a measure of health that covers five dimensions—mobility, self-care, usual activities, pain/discomfort, and anxiety/depression—using five levels of severity—no, slight, moderate, severe, and extreme problems/unable to [16]. The measure describes 3125 health states ($5^5 = 3125$) and is used across many healthcare settings. A misery score can be calculated by summing the dimension-level responses for a crude estimate of health state severity ranging from 5 to 25.

Time trade-off task and data source

The Online US Valuation Study of the EQ-5D-5L served as the data source for these analyses. Respondents from online panels were recruited to complete preference tasks such as composite time trade-off (cTTO) and/or discrete choice experiment tasks to provide valuations for health states described by the EQ-5D-5L. The cTTO task was composed of the conventional TTO and the lead-time TTO for better-than-dead (BTD) and worse-than-dead (WTD) preference elicitation, respectively. The conventional TTO was shown at the beginning of each cTTO (Appendix A); it compared 10 years of full health (Life A) to 10 years of suboptimal health (Life B). The time in Life A changed via a predetermined ping-pong/titration pattern as the respondent made a series of "trade-offs" by indicating a preference (i.e., made a choice) between Life A and Life B. If the respondent indicated that he/she may believe the health state being valued to be worse-than-death (WTD), the TTO task can be modified to elicit worse-than-death values using the lead-time TTO (Appendix B). In this method, 10 additional years were added to both Life A and Life B. These years can be additionally traded off via a ping-pong/titration pattern to elicit WTD TTO values (i.e., less than 0) as negative as − 1. (Appendices A and B) [17, 18]. In either portion of the cTTO, the respondent pressed an equivalence button labeled "Life A and Life B are about the same" to complete the task when he/she felt time in Life A was approximately equal to Life B.

SurveyEngine, a company specializing in online choice experiments, hosted the survey platform and conducted respondent recruitment. Respondents from online panels were quota sampled on age, gender, race, and ethnicity to match the US general population [19]. For online purposes, visual presentation (Appendices A and B), cTTO task specification, and automated iteration pattern of the online platform replicated the EuroQol Valuation Technology (EQ-VT), the official software developed by the EuroQol Group for EQ-5D-5L valuation studies [17, 20–22]. The EQ-VT TTO experimental design was also employed for the online study [23]. The experimental design blocked 86 total EQ-5D-5L health states into 10 blocks of 10 health states each. Every block included 55555 (the worst EQ-5D-5L health state) and a mild health state with only slight problems on a single dimension. Additional notifications were added to those present in the EQ-VT to simulate the role of an interviewer as recommended in EuroQol Standard Valuation Protocol 2.0, such as reminders for respondents to carefully consider each task if the respondent attempted to complete a task too quickly [19, 21].

Each respondent completed five practice health states in the same order to learn health valuation using the cTTO and become familiar with the EQ-5D-5L [21]. The first two tasks illustrated the cTTO through interactive instruction steps. The first task requested respondents to imagine life in a wheelchair to introduce the conventional TTO. The second task asked respondents to imagine "a health state that is much worse than being in a wheelchair" to demonstrate the LT-TTO. The next 3 practice states were predetermined mild, severe, and seemingly implausible EQ-5D-5L health states to show the range and types of health states that can be described by the measure. The respondent was then randomly assigned a block of TTO tasks and valued 10 EQ-5D-5L health states using the cTTO. These tasks were also presented to the respondent in random order.

Two arms of the US Online Valuation Study included TTO tasks—the “base case” and “engagement” arms. The cTTO of the base case arm was as previously described, and respondents could end the task at any time by choosing the equivalence button, including at the beginning of the task when no trade-offs had been made. The cTTO in the engagement arm retained the visual and trade-off sequence aspects of the base case arm. However, at least three trade-offs/choices per task were required in this arm before the respondent could indicate approximate equivalence between Lives A and B to end the task. This imposed engagement was enforced by hiding the equivalence button from respondents until three trade-offs were completed. The task-ending equivalence button could not be accidentally actuated while hidden. If the respondent ended the task immediately following these three trade-offs, four TTO values were possible: 0.4, − 0.4, 0.6, or − 0.6 [20]. If the respondent wished to indicate a TTO value for a health state that was bypassed within the first three trade-offs, he/she must choose to continue trading using the algorithm until the desired TTO value is indicated again. The planned TTO value routing is demonstrated in appendix D.

Although face-to-face data collection may be considered as the gold standard for TTO studies, the base case and engagement arms were both recruited from online panels. Face-to-face and online panel respondents may be systematically different due to varying selection pressures such as access to internet and poor health/mobility precluding participation in face-to-face studies [24–27]. These divergent participant characteristics may also affect underlying preference functions. Therefore, use of online panel participants was determined to be most efficient to isolate the effect of increased engagement. Study arm allocation was not randomized as the arms were not conducted in parallel.

Analyses

Respondent characteristics

Respondent socio-demographics and other relevant characteristics, such as self-reported health, were descriptively summarized, compared between arms, and evaluated for similarity to the US general population.

Elicited TTO values, preference patterns, and meta-data comparison

Distributions of raw elicited TTO values, including number and scale of local maximums (“spikes”), were descriptively reviewed. These values were also used to construct within-respondent preference patterns. Meta-data (e.g., time spent per task) were captured by the online platform with regard to the choices and respondent behavior on each task. The elicited TTO values, preference patterns, and meta-data were analyzed to characterize and compare trading behaviors, data validity, and engagement with the task.

Face validity of the TTO values was assessed using descriptive statistics of the elicited TTO values. The mean and standard deviations of the TTO values were expected to decrease and increase, respectively, when worse health states (e.g., larger misery score) were valued. This pattern would show the decreased desirability and increased disagreement regarding the utility of unfavorable health states.

The number and prevalence of TTO inconsistencies could be used to evaluate data validity. Each TTO block had potential dominant/dominated health state pairs, and each pair corresponded to a possible TTO inconsistency if the dominated health state was given a higher TTO value. Possible dominant/dominated pairs were identified by comparing all the health states within each TTO block. For example, 55555 dominated the other 9 health states in each block. Therefore, each respondent could provide up to 9 TTO inconsistencies involving 55555 if the value given to 55555 was higher than all other health states valued. Other dominant/dominated health state pairs by each block of the experimental design are summarized in Appendix E. The percentage of respondents with at least 1 inconsistency in TTO values (involving any health state and 55555) and the mean number of inconsistent values were compared between arms. Respondent self-reported understanding of TTO tasks was additionally evaluated for assessment of data validity. Data validity was hypothesized to be greater in the engagement arm.

Respondent engagement was evaluated using time spent, mean number of trade-offs per TTO task, and prevalence of minimum effort trading (i.e., task completion in fewest number of trade-offs allowed). Trading behaviors compared between arms included (1) percentage of elicited values that were worse-than-dead (TTO value <0), (2) prevalence of non-traders (all TTO values $=1$ for a single respondent), and (3) prevalence of better-than-dead only traders (all TTO values >0 for a single respondent). Respondent engagement was hypothesized to be lower in the base case arm, and trading behaviors were hypothesized to differ between arms.

Value set modeling and comparison

As the purpose of TTO preferences is often to create a value set, the characteristics of the resulting value sets were important considerations. Value sets were estimated from TTO-based preferences using a linear regression model with a respondent-level random intercept (RILR) to account for repeated observations (Eq. 1). Each arm of the survey was separately modeled using the same model type and parameters to compare characteristics of the resulting value sets (Models 1 and 2). The dependent variable was elicited cTTO values. Models 1 and 2 were specified with 20 dummy variables as the independent variables: 4 decrements from "no problems" for each of the 5 dimensions of the EQ-5D-5L ($4 \times 5 = 20$).

$$\begin{aligned} cTTO_{ij} =& \beta_0 + \beta_1(MO2)_{ij} + \beta_2(MO3)_{ij} + \beta_3(MO4)_{ij} \\ &+ \beta_4(MO5)_{ij} + \beta_5(SC2)_{ij} + \beta_6(SC3)_{ij} \\ &+ \beta_7(SC4)_{ij} + \beta_8(SC5)_{ij} + \beta_9(UA2)_{ij} \\ &+ \beta_{10}(UA3)_{ij} + \beta_{11}(UA4)_{ij} + \beta_{12}(UA5)_{ij} \\ &+ \beta_{13}(PD2)_{ij} + \beta_{14}(PD3)_{ij} + +\beta_{15}(PD4)_{ij} \\ &+ \beta_{16}(PD5)_{ij} + \beta_{17}(AD2)_{ij} + \beta_{18}(AD3)_{ij} \\ &+ \beta_{19}(AD4)_{ij} + \beta_{20}(AD5)_{ij} + u_i + \varepsilon_{ij}, \end{aligned}$$

MO is mobility; SC is self-care; UA is usual activities; PD is pain/discomfort; AD is anxiety/depression; number following dimension indicates level of severity (e.g., MO2 is mobility level 2); *i* is respondent; *j* is accounts for the multiple cTTO tasks completed; $u_i \sim N(0, \sigma^2)$ is the random intercept for the *i*th subject; ε_{ij} is normally distributed error term.

Desirable value set characteristics were minimal number of preference inversions and number of insignificant beta parameters. Other evaluated descriptive value set characteristics included relative importance of dimensions, percentage of modeled health states WTD, and range of scale. Average valuation difference between arms was assessed using a RILR using the 20-dummy specification for dimension levels with an additional variable for study arm (Model 3). The value set for the engagement arm was hypothesized to be more valid.

Results

Five hundred and one and 504 respondents completed the base case and engagement surveys, respectively (Table 1). No notable respondent characteristics differed between arms. Respondents in both arms were generally similar to the US general population except for educational attainment; a slightly higher proportion of survey respondents completed any education after high school/GED (data not shown).

Elicited TTO values, preference patterns, and meta-data

Local maximum ("spikes") occurred at different values and/or scales for the study arms (Fig. 1). The base case arm had 3 large spikes at TTO values of 0 (15.2%), 0.5 (8.0%), and 1 (32.0%) whereas the engagement arm had smaller local maximums at 1 (15.3%), 0.6 (11.0%), 0.45 (9.7%), and − 0.5 (6.2%). In review of face validity, poorer health states were less preferred (lower means) and were assigned a larger range of TTO values (higher standard deviation) in both arms, matching the anticipated pattern (Fig. 2). When compared by severity/misery score, means for the engagement arm were typically lower than the base case arm [misery score 15: (base case) 0.648; (engagement) 0.514; 55555: (base case) 0.409; (engagement) 0.258; Fig. 2]. Notably, for very mild health states (misery score = 6), the mean TTO values were 0.838 and 0.753 for the base case and engagement arms, respectively.

Additional assessments of data validity, such as number of respondents with any inconsistency and mean number of inconsistencies per respondent, did not significantly differ between arms (Table 2). On average, each respondent had 2.77 (SD 3.68) inconsistencies in the base case arm and 3.14 (SD 4.11) in the engagement arms ($p = 0.14$); 61.1% of base case respondents and 63.5% of engagement respondents produced at least 1 inconsistent TTO value ($p = 0.14$). Self-reported TTO task comprehension also did not differ between arms; 80.8% and 82.9% of base case and engagement arm respondents, respectively, reported that the TTO task was somewhat or very easy to understand ($p = 0.40$).

Trading behavior also differed between arms. In comparison to the base case arm, the engagement arm yielded more worse-than-dead (TTO value < 0) valuations: [base case] 2.8%; [engagement] 10.6% (Fig. 1). Despite having more WTD values, the engagement arm had a greater percentage of respondents who only traded within positive TTO values: [base case] 46.3%; [engagement] 56.9% $p = 0.0007$ (Table 2). Thus, respondents who gave any negative values in the engagement arm provided more negative health state valuations than the base case arm. A smaller portion of engagement arm respondents were non-traders (all TTO values = 1): [base case] 7.2%; [engagement] 2.0%; $p < 0.0001$.

Measures of respondent task engagement generally did not differ between arms unless it was an artifact of the task design. For example, respondents in the engagement arm used more moves than those in the base case arm (trade-offs + 1 for actuating equivalence button): [base case] 5.6 (SD 5.8); [engagement] 7.7 (SD 5.3) $p < 0.0001$ (Table 3). However, respondents in both arms spent about the same time per task: [base case] 63.3 s (SD 77.9 s); [engagement] 64.7 s (SD 73.3 s) $p = 0.36$. The proportion of respondents that completed all TTO tasks in the minimum number of trade-offs (0 for base case and 3 for engagement) was similar between arms: [base case] 5.4%; [engagement] 4.6%; $p = 0.55$ (Table 2). However, a larger proportion of tasks ended with the minimum number of trade-offs in the engagement arm: [base case] 24.6%; [engagement] 28.3%; $p < 0.0001$ (Table 3).

Table 1 Respondent characteristics

Characteristic	Base case arm [1] $N=501$	Engagement arm [2] $N=504$	*p* value for [1] vs [2]
Age, mean (SD), n (%)	45.9 (15.1)	45.0 (15.3)	0.35
18–34	149 (29.7)	163 (32.3)	0.61
35–54	180 (35.9)	180 (35.7)	
55 +	172 (34.3)	161 (31.9)	
Range	17–80	16–85	
Gender, *n* (%)			
Male	251 (50.1)	238 (47.2)	0.36
Female	250 (49.9)	266 (52.8)	
Gender, other	–	–	
Race, *n* (%)			
White	387 (77.3)	410 (81.4)	0.60 (black vs. others)
Black	63 (12.6)	58 (11.5)	
Hispanic ethnicity, *n* (%)	75 (15.0)	67 (13.3)	0.45
Education level greater than secondary, *n* (%)	344 (68.7)	364 (72.2)	0.22
Child dependents			
None	338 (67.5)	350 (69.4)	0.06
Child(ren), ≤ 5 years old	65 (13.0)	52 (10.5)	0.19
Child(ren), 6 to 17 years old	138 (27.5)	131 (26.0)	0.58
Primary health insurance			
None	49 (9.8)	54 (10.7)	0.22
Public	204 (40.5)	178 (35.3)	
Private	249 (49.7)	272 (54.0)	
Country of birth, United States	475 (94.8)	472 (93.7)	0.43
History of illness, *n* (%) (45)			
Hypertension	141 (28.1)	116 (23.0)	0.06
Arthritis	120 (24.0)	106 (21.0)	0.27
Diabetes	71 (14.2)	56 (11.1)	0.14
Heart failure	11 (2.2)	6 (1.0)	0.22
Stroke	11 (2.2)	9 (1.8)	0.64
Bronchitis	18 (3.6)	14 (2.8)	0.46
Asthma	52 (10.4)	46 (9.1)	0.50
Depression	117 (23.4)	112 (22.2)	0.67
Migraine	58 (11.6)	53 (10.5)	0.59
Cancer	12 (2.4)	8 (1.6)	0.36
None	157 (31.3)	189 (37.5)	0.04
Health status, *n* (%) (44)			
Excellent/very good/good	411 (82.0)	416 (82.5)	0.83
Fair/poor	90 (18.0)	88 (17.5)	
Self-reported EQ-VAS			
Mean (SD)	73.6 (20.4)	73.9 (19.0)	0.81
Median (IQR)	46 (26)	44.5 (26)	
Mobility			
No problems	333 (66.5)	352 (69.8)	0.21
Slight problems	98 (19.6)	89 (17.7)	
Some/moderate problems	55 (11.0)	41 (8.1)	
Severe problems	13 (2.6)	15 (3.0)	
Unable to walk about	2 (0.4)	7 (1.4)	

Table 1 (continued)

Characteristic	Base case arm [1] $N=501$	Engagement arm [2] $N=504$	p value for [1] vs [2]
Self-care			
No problems	420 (82.9)	441 (87.5)	0.47
Slight problems	50 (10.0)	36 (7.1)	
Some/moderate problems	24 (4.8)	19 (3.8)	
Severe problems	5 (1.0)	5 (1.0)	
Unable to wash or dress	2 (0.4)	3 (0.6)	
Usual activities			
No problems	328 (65.5)	338 (67.1)	0.96
Slight problems	103 (20.6)	103 (20.4)	
Some/moderate problems	54 (10.8)	49 (9.2)	
Severe problems	13 (2.6)	12 (2.4)	
Unable to do usual activities	3 (0.6)	2 (0.4)	
Pain/discomfort			
No pain or discomfort	185 (36.9)	183 (36.3)	0.20
Slight pain or discomfort	172 (34.3)	189 (37.5)	
Moderate pain or discomfort	104 (20.8	98 (19.4)	
Severe pain or discomfort	36 (7.2)	24 (4.8)	
Extreme pain or discomfort	4 (0.8)	10 (2.0)	
Anxiety/depression			
Not anxious or depressed	228 (45.5)	248 (49.2)	0.44
Slightly anxious or depressed	138 (27.5)	130 (25.8)	
Moderately anxious or depressed	90 (18.0)	93 (18.5)	
Severely anxious or depressed	29 (5.8)	18 (3.6)	
Extremely anxious or depressed	16 (3.2)	15 (3.0)	

SD standard deviation

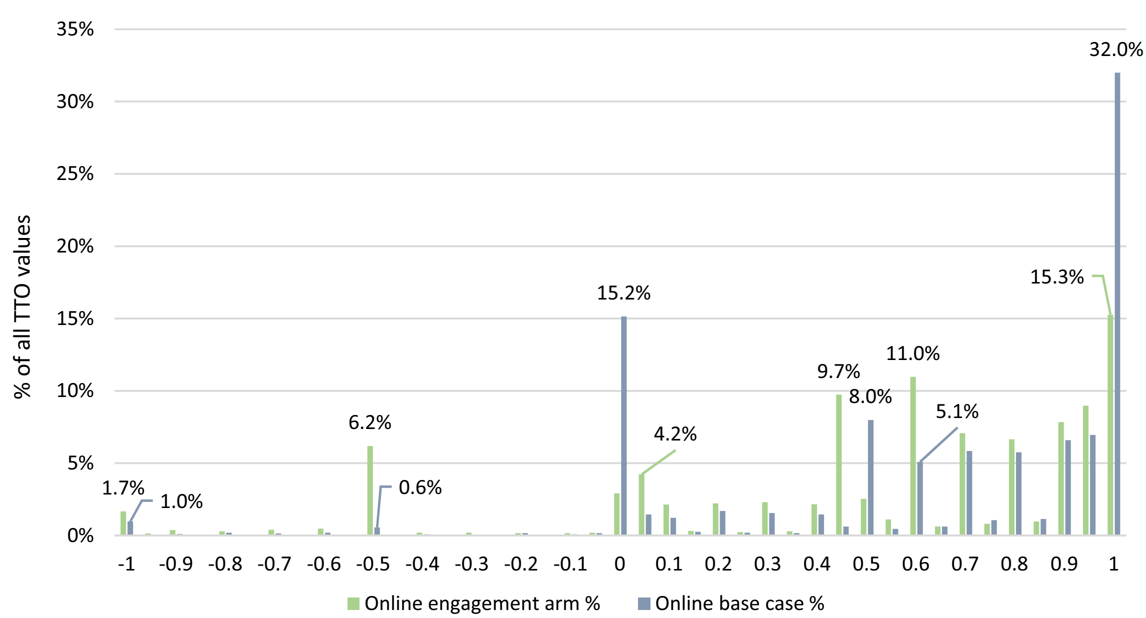

Fig. 1 Distribution of cTTO values by online arm

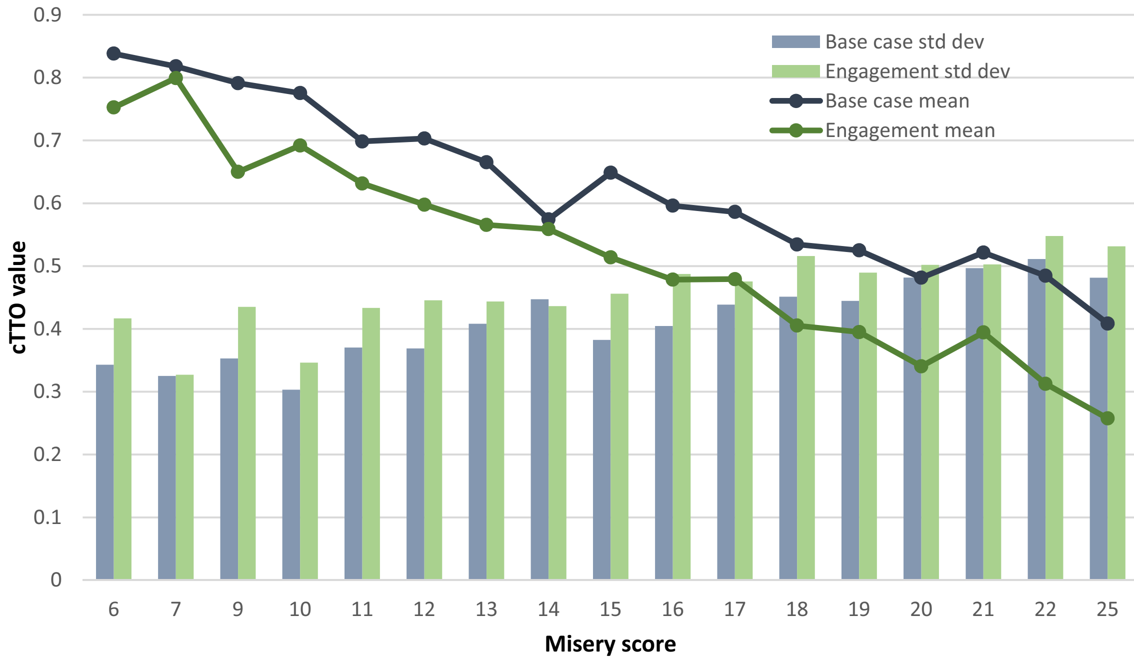

Fig. 2 Mean and standard deviation of elicited cTTO values by 5L health state severity

Table 2 Summary of respondent-level preferences and meta-data

Category	Pattern	Base case arm $N = 501$		Engagement arm $N = 504$		*p* value
		n	%	*n*	%	
Trading behavior	Better-than-dead only traders (all TTO values > 0)	232	46.3%	287	56.9%	0.0007
	Non-trader (all TTO values = 1)	36	7.2%	10	2.0%	< .0001
Data validity	Trader with at least 1 inconsistency	306	61.1%	320	63.5%	0.43
	Trader with at least 1 inconsistency involving 55555	207	41.3%	217	43.1%	0.58
	Inconsistencies per respondent (mean, SD)	2.77	3.68	3.14	4.11	0.14
	55555 inconsistencies per respondent (mean, SD)	1.26	2.04	1.50	2.37	0.09
	TTO task understanding	405	80.8%	418	82.9%	0.40
Task engagement	Time spent on survey in minutes (mean, SD)	38.7	16.7	37.5	16.3	0.24
	Ended all TTO tasks with minimum trade-offs	27	5.4%	23	4.6%	0.55

TTO time trade-off, *SD* standard deviation

Modeled value set

The modeled value sets were dissimilar between arms (Models 1 and 2; Table 4). Compared to base case, the engagement arm had fewer parameters that were not significantly different from the reference level of “no problems” (6 versus 8). Each arm had 2 preference inversions, but none were significant (Appendix C). For example, the incremental disutility of the move from self-care levels 4 to 5 was not statistically significantly different from 0 in both arms, indicating that self-care levels 4 and 5 had approximately the same disutility for both value sets. Few estimated incremental disutilities were statistically significantly different from 0 in both Models 1 and 2 (Appendix C). The relative dimension importance was different between arms. No dimension was in the same importance ranking when compared between arms. As would be projected from the mean TTO values, the compulsory engagement caused a downward shift of predicted health state utility values. On average, valuations were 0.107 lower on the utility scale in the engagement arm (Table 4).

The kernel density plots demonstrated that both value sets (Models 1 and 2) had generally positive distributional characteristics, including normality and unimodality (Fig. 3). The engagement arm had a longer range of scale and fewer

Table 3 Summary of task-level preferences and meta-data

	Base case arm N = 5010		Engagement arm N = 5040		p value
	Mean	SD	Mean	SD	
Moves (trade-offs + 1 move for task completion)	5.6	5.8	7.7	5.3	< 0.0001
Time spent per task in seconds	63.3	77.9	64.7	73.3	0.36
cTTO value	0.63	0.43	0.52	0.49	< 0.0001
Task ended with minimal trade-offs (n, %)	1234	24.6%	1425	28.3%	< 0.0001

cTTO composite time trade-off, *SD* standard deviation

Table 4 Modeled linear regression results

	Model 1—Online cTTO base case (n = 501)			Model 2—Online cTTO engagement arm (n = 504)			Model 3—Online cTTO joint model		
	Estimate	SE	p value	Estimate	SE	p value	Estimate	SE	p value
Intercept	0.846	0.021	< .0001	0.783	0.023	< .0001	0.868	0.018	< .0001
MO2	− 0.026	0.016	0.114*	− 0.029	0.016	0.079*	− 0.027	0.012	0.019
MO3	− 0.043	0.017	0.011	− 0.04	0.017	0.019	− 0.041	0.012	0.0006
MO4	− 0.067	0.019	< .0001	− 0.08	0.018	< .0001	− 0.074	0.013	< .0001
MO5	− 0.112	0.017	< .0001	− 0.096	0.017	< .0001	− 0.104	0.012	< .0001
SC2	− 0.003	0.016	0.874*	− 0.019	0.016	0.249*	− 0.011	0.011	0.349*
SC3	− 0.035	0.018	0.055*	− 0.056	0.018	0.002	− 0.045	0.013	0.0004
SC4	− 0.098	0.018	< .0001	− 0.077	0.018	< .0001	− 0.088	0.013	< .0001
SC5	− 0.077	0.016	< .0001	− 0.105	0.016	< .0001	− 0.090	0.012	< .0001
UA2	− 0.030	0.017	0.075*	0.006	0.017	0.728	− 0.012	0.012	0.304*
UA3	− 0.067	0.018	< .0001	− 0.034	0.018	0.061*	− 0.051	0.013	< .0001
UA4	− 0.059	0.018	0.001	− 0.074	0.018	< .0001	− 0.066	0.013	< .0001
UA5	− 0.075	0.016	< .0001	− 0.085	0.016	< .0001	− 0.080	0.012	< .0001
PD2	− 0.020	0.015	0.187*	− 0.023	0.015	0.121*	− 0.022	0.011	0.043*
PD3	− 0.023	0.018	0.210*	− 0.060	0.018	0.001	− 0.042	0.013	0.001
PD4	− 0.090	0.016	< .0001	− 0.143	0.016	< .0001	− 0.116	0.012	< .0001
PD5	− 0.108	0.018	< .0001	− 0.146	0.017	< .0001	− 0.127	0.012	< .0001
AD2	− 0.010	0.018	0.586*	− 0.016	0.018	0.348*	− 0.013	0.012	0.280*
AD3	− 0.031	0.02	0.114*	− 0.046	0.02	0.019	− 0.039	0.014	0.005
AD4	− 0.066	0.018	< .0001	− 0.111	0.018	< .0001	− 0.088	0.013	< .0001
AD5	− 0.067	0.017	< .0001	− 0.099	0.017	< .0001	− 0.084	0.012	< .0001
Engagement arm utility (ref = base case)	N/A			N/A			− 0.107	0.020	< .0001
Dimension ranking	MO-PD-SC-UA-AD			PD-SC-AD-MO-UA			PD-MO-SC-AD-UA		
Estimated utility values by health state							N/A		
21111	0.820			0.754					
12111	0.844			0.764					
11211	0.816			0.789					
11121	0.826			0.759					
11112	0.837			0.766					
55555	0.400			0.253					
Health States WTD, n (%)	0 (0.0%)			0 (0.0%)			N/A		

WTD worse-than-death

*Parameter estimate that is not significantly different from the reference level health states with values close to 1 (Fig. 3). A linear regression fit through the scatterplot of estimated beta values showed the absolute value of base case betas were about 37% smaller than those from the engagement arm, on average (Fig. 4). There was no clear pattern discerned for the relationship between paired beta estimates by dimension.

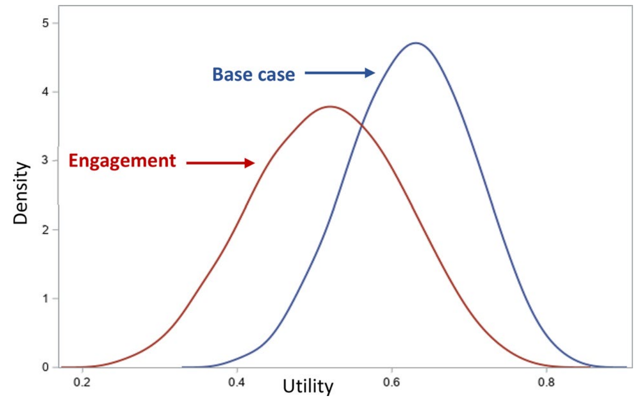

Fig. 3 Kernel density plots of base case and engagement value sets

Discussion

The results suggest that mandating three trade-offs prior to task completion in the composite TTO tasks did not meaningfully improve engagement and data validity in online respondents. The intervention was intended to induce more careful respondent consideration of the tasks and prevent task short-cutting, but the time spent per task and overall survey time were similar between arms. Although time spent was an imprecise measure of engagement, more sophisticated measurement such as eye-tracking would be challenging to implement in online studies meant for respondent self-completion without specialized technology. The differences in mean moves per task were induced by the task design and did not appreciably contribute to the evidence of increased engagement. The prevalence and mean number of inconsistent valuations also did not improve with mandatory task engagement.

Comparing the distribution of local maximums offered some suggestive evidence that three required trade-offs led to more considered preferences as the engagement arm had fewer large "spikes" and more evenly distributed TTO values, which can be taken as evidence of emulating valid preferences in that they were less clustered around a few TTO values. The size of local maximums in the engagement arm was likely diminished due to the engagement requirement, which redistributed the preferences of respondents who simply wished to complete a task in the fewest steps. If a task was completed using the minimum number of trade-offs per task (i.e., 0 trade-offs in the base case arm and 3 trade-offs in the engagement arm), only 1 TTO value was possible in the base case arm whereas multiple TTO values were possible for the engagement arm. The largest spike for the base case arm was at the first possible equivalence point (TTO value = 1), followed by the second (0) and third (0.5) equivalence points. Local maximums were observed in the engagement arm at 1, 0.6, 0.45, and − 0.5; due to a programming error in the trade-off mechanism that was discovered during analysis, 0.6, 0.45, and − 0.5 were the

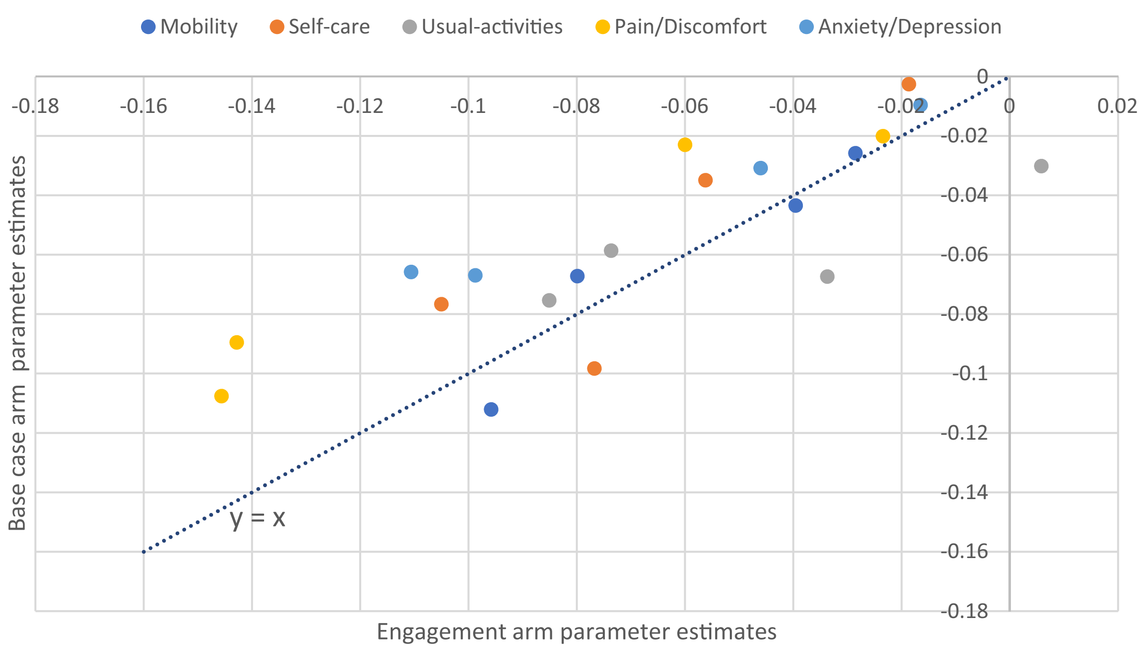

Fig. 4 Scatterplot of cTTO parameter estimates for both base case and engagement arms

TTO values following 3 trade-offs instead of the expected ± 0.6 and ± 0.4. However, as the data quality was evidently quite poor, this change to three values that are numerically close to the originally planned four values plausibly did not affect data quality. Unexpectedly, the largest local maximum within the engagement arm was at 1. To reach the TTO value of 1, the respondent needed to continue trading-off following fulfillment of the engagement requirement until the TTO value was back at 1. Respondents may have continued to indicate that they "prefer" Life A because that was the most reliable way to make the equivalence button appear, and because the TTO value cannot be greater than 1, there were many health states assigned a value of 1. Additional qualitative debriefing with respondents could provide additional insights into this phenomenon.

The value set constructed from the engagement arm TTO values actually had poorer face validity than the one derived from preferences collected in the base case arm. The value set intercepts, which can be interpreted as the utility for 11111 (no problems on any dimension/perfect health), were 0.846 and 0.783 for base case and engagement arms, respectively. The difference resulted from the lower mean TTO value of very mild health states (severity = 6), which may have been a downstream effect of the aforementioned TTO value redistribution (Fig. 2). The TTO value redistribution likely also contributed to the lower mean TTO value in the engagement arm by diminishing the proportion of TTO values at 1.

Several causes could have contributed to the observed differences between arms. Respondents in the engagement arm may not have understood that a conscious effort was needed to indicate their true preference if it was bypassed within the first 3 trade-offs. There were no differences between arms in respondent self-reported understanding of the task, suggesting that at minimum, respondents were not confused by the constraint. However, this finding may be indicative of respondents' general lack of investment in the successful completion of the tasks rather than evidence of no additional confusion. Further, respondents could have been swayed in their preferences by the task specification as evidenced by the potential dispersal of observed preferences by the mandatory trade-offs. Previous evidence also suggested that TTO-based preferences are vulnerable to influences from task design. For example, Lenert and colleagues previously found that ping-pong and titration-based iterative algorithms elicited different TTO values [15].

To better contextualize the data quality of both online arms, measures of engagement, data validity, and value set characteristics can be descriptively compared between TTO tasks from face-to-face (F2F), online base case, and online engagement arms [12, 13, 19]. Online base case tasks were completed using the fewest moves with 5.6 (SD 5.8) average moves per task, followed by face-to-face with 6.6 moves (SD 4.8), and online engagement with 7.7 moves (SD 5.3) (Table 3). If the mean number of inconsistencies involving any valued health state per respondent is used to assess data validity, the complete sample of face-to-face respondents had 1.43 inconsistencies while online base case and engagement arms each had 2.77 and 3.14 inconsistencies (Table 2) [13]. Further, 31.8% of face-to-face respondents provided at least 1 inconsistency in contrast to the 61.1% of online base case and 63.5% of engagement respondents who had the same data quality issue. When all samples were modeled using RILR, the range of scale was 1.27, 0.446, 0.530 for F2F, online base case, and online engagement arms, respectively. (Table 4) For the F2F data, its value for perfect health was closest to 1 at 0.963, and the values for perfect health both online arms were quite far from 1: [online base case] 0.846 [online engagement] 0.783. If face-to-face TTO administration is considered standard practice and the gold standard, neither the online base case nor engagement arm could achieve similar results. These results suggest that online, unsupervised TTO-based valuation of multi-attribute utility instruments as currently conceptualized and implemented are unable to reliably collect valid preferences with or without imposed task engagement. Without significant reimagining of the task, online, unsupervised TTO should not be the researchers' first choice to value multi-attribute health states. If the TTO must be used in such contexts due to monetary or time restraints, the results should be interpreted with caution based on these analyses.

Despite such discouraging results, respondent task shortcutting showed some signs of improvement compared to previous studies. The EuroQol group previously investigated the implementation of 3 minimum trade-offs using Dutch online panels when investigating TTO task specification for valuing the EQ-5D-5L [17]. In this pilot study, nearly half (46.7%) of all tasks were completed using the minimum number of trade-offs in the engagement arm as compared to 26.1% of tasks in the base case arm [14]. In both arms of the present study, only about a quarter of all TTO tasks were completed using the fewest possible trade-offs. Therefore, present-study respondents were less inclined to end the task as soon as possible and engage more with the task, even in the engagement arm when they may be irritated by the requirement. The research question was revisited in the present study in part due to improved understanding of the cTTO, accumulated researcher experience in optimizing online data collection platforms, and technology advances in visual task presentation. These factors may have contributed to a more intuitive preference elicitation task and more attentiveness than earlier platforms.

Other researchers have also contemplated ways to improve data validity of TTO-based preferences. In "repair of errors", if the respondent committed a logical inconsistency, where a lower value was assigned to an objectively better health state, the inconsistency was re-presented to the respondent [28]. The respondent could then change his/her answer, or "repair the error". Preferences elicited using this approach were similar to the preferences in respondents that did not commit a logical inconsistency, suggesting improved data validity. Additionally, Edelaars-Peeters and colleagues evaluated respondent behavior on TTO tasks using qualitative methods and recommend incorporation of automated features for online TTO administration, including checks similar to "repair the error" and accessibility to answers to frequently asked questions [8]. The face-to-face TTO feedback module allowed for a reconsideration of preference validity, but respondents could not re-value the health state. This feedback module was previously shown to lower the number of inconsistencies [20]. The evidence suggests that implementation of an opportunity for the respondent to re-evaluate the first set of TTO values may continue to improve the quality of preferences elicited online. However, this checkpoint may need to be implemented differently than face-to-face as online respondents may be more prone to being significantly influenced by its presence, particularly if the grasp of the TTO task is already tenuous.

In summary, imposing three trade-offs as the minimum bar for respondent engagement was ineffective alone in achieving more valid TTO values and TTO-based value sets, but online TTO data are of poor quality in general compared to face-to-face. Survey data collection is transitioning toward use of online panels and unsupervised methods as researchers look to harness their benefits. However, these research methods can carry risks, and this evidence contributes toward the debate on the strengths and weaknesses of online data collection.

Acknowledgements Funding for data collection was supported by Bristol-Myers Squibb. R Jiang was supported by fellowships from the PhRMA Foundation and a Dean's Scholar Award from the University of Illinois at Chicago during her PhD, during the time of study conduct and analysis.

Appendix

See Tables 5 and 6 and Figs. 5, 6, and 7.

Table 5 Table 4 with incremental dummies

	Model 1—Online cTTO base case			Model 2—Online cTTO engagement arm			Model 3—Online cTTO joint model		
	Estimate	SE	*p* value	Estimate	SE	*p* value	Estimate	SE	*p* value
Intercept	0.846	0.021	< .0001	0.783	0.023	< .0001	0.868	0.018	< .0001
MO2	− 0.026	0.016	0.114	− 0.029	0.016	0.079	− 0.027	0.012	0.019
MO3	− 0.018	0.018	0.323	− 0.011	0.018	0.530	− 0.014	0.013	0.262
MO4	− 0.024	0.020	0.224	− 0.040	0.019	0.038	− 0.032	0.014	0.018
MO5	− 0.045	0.019	0.017	− 0.016	0.019	0.397	− 0.030	0.013	0.024
SC2	− 0.003	0.016	0.874	− 0.019	0.016	0.249	− 0.011	0.011	0.349
SC3	− 0.032	0.020	0.097	− 0.038	0.019	0.051	− 0.035	0.014	0.012
SC4	− 0.063	0.019	0.001	− 0.021	0.019	0.278	− 0.042	0.013	0.002
SC5	0.022	0.017	0.194	− 0.028	0.017	0.092	− 0.003	0.012	0.805
UA2	− 0.030	0.017	0.075	0.006	0.017	0.728	− 0.012	0.012	0.304
UA3	− 0.037	0.018	0.038	− 0.040	0.018	0.026	− 0.038	0.013	0.002
UA4	0.009	0.019	0.652	− 0.040	0.019	0.037	− 0.016	0.014	0.250
UA5	− 0.017	0.019	0.374	− 0.011	0.019	0.541	− 0.014	0.013	0.301
PD2	− 0.020	0.015	0.187	− 0.023	0.015	0.121	− 0.022	0.011	0.043
PD3	− 0.003	0.020	0.886	− 0.037	0.020	0.069	− 0.020	0.014	0.166
PD4	− 0.067	0.019	0.0003	− 0.083	0.018	< .0001	− 0.074	0.013	< .0001
PD5	− 0.018	0.021	0.382	− 0.003	0.020	0.890	− 0.011	0.015	0.439
AD2	− 0.010	0.018	0.586	− 0.016	0.018	0.348	− 0.013	0.012	0.280
AD3	− 0.021	0.019	0.257	− 0.030	0.019	0.113	− 0.026	0.013	0.053
AD4	− 0.035	0.018	0.049	− 0.065	0.018	0.0003	− 0.049	0.013	< .0001
AD5	− 0.001	0.017	0.944	0.012	0.017	0.473	0.005	0.012	0.683
Engagement arm	N/A			N/A			− 0.107	0.020	< .0001

Yellow highlighting indicates preference inversion (all insignificant); blue highlighting indicates parameter estimate that is insignificantly different from the level beneath it

Table 6 Dominated health state pairs within each TTO block

Block	Health state	Dominated health states within the same block
1	11221	None
	11235	11211
	54231	12111; 11221
	51451	11221
	34515	12111; 12514
	35245	12111; 11221; 11235
	12514	12111
	45144	12111
	12111	None
	55555	11221; 11235; 54231; 51451; 34515; 35245; 12514; 45144; 12111
Total number of possible inconsistencies in block		19
2	12543	11211; 12121
	12121	None
	43542	11211; 12121
	34155	12121
	52215	11211
	45133	12121
	32443	11211; 12121
	23514	11211
	11211	None
	55555	12543; 12121; 43542; 34155; 52215; 45133; 32443; 23514; 11211
Total number of possible inconsistencies in block		19
3	45233	21111; 25122
	55233	21111; 25122; 45233
	31525	21111; 11421
	52455	21111; 12244; 11421
	12244	None
	13313	None
	25122	21111
	11421	None
	21111	None
	5555	45233; 55233; 31525; 52455; 12244; 13313; 25122; 11421; 21111
Total number of possible inconsistencies in block		20
4	21112	None
	14554	11121; 12513; 12344
	12513	None
	44345	11121; 21112; 12344; 44125
	12344	11121
	53221	11121
	54342	11121; 21112; 53221
	44125	11121; 21112
	11121	None
	55555	21112; 14554; 12513; 44345; 12344; 53221; 54342; 44125; 11121
Total number of possible inconsistencies in block		22
5	43315	11112; 21315
	54153	11112; 14113
	52431	None

Table 6 (continued)

Block	Health state	Dominated health states within the same block
	24443	11112; 14113
	14113	11112
	31524	11112
	15151	None
	21315	11112
	11112	None
	55555	43315; 54153; 52431; 24443; 14113; 31524; 15151; 21315; 11112
Total number of possible inconsistencies in block		18
6	12112	None
	11212	None
	44553	21111; 11212; 12112; 23152
	21345	21111; 11212
	34244	21111; 11212; 12112
	23152	21111; 12112
	43514	21111; 11212; 12112
	55424	21111; 11212; 12112
	21111	None
	55555	12112; 11212; 44553; 21345; 34244; 23152; 43514; 55424; 21111
Total number of possible inconsistencies in block		26
7	13122	None
	24553	11211; 13122
	51152	None
	11425	11211
	22434	11211
	42115	None
	35332	11211; 13122
	45413	11211
	11211	None
	55555	13122; 24553; 51152; 11425; 22434; 42115; 35332; 45413; 11211
Total number of possible inconsistencies in block		16
8	33253	11112; 23242
	23242	11112
	24342	11112; 23242
	32314	11112
	12334	11112
	21334	11112
	55225	11112
	53412	11112
	11112	None
	55555	33253; 23242; 24342; 32314; 12334; 21334; 55225; 53412; 11112
Total number of possible inconsistencies in block		19
9	11414	None
	25331	11121
	25222	11121;
	21444	11121

Table 6 (continued)

Block	Health state	Dominated health states within the same block
	31514	11414
	53243	11121
	53244	11121; 53243
	35143	11121
	11121	None
	55555	11414; 25331; 25222; 21444; 31514; 53243; 53244; 35143; 11121
Total number of possible inconsistencies in block		17
10	11122	None
	52335	12111; 11122; 42321
	35311	12111
	43555	12111; 11122; 42321; 13224
	24445	12111; 11122; 13224
	13224	12111; 11122
	34232	12111; 11122
	42321	12111
	12111	None
	55555	11122; 52335; 35311; 43555; 24445; 13224; 34232; 42321; 12111
Total number of possible inconsistencies in block		24

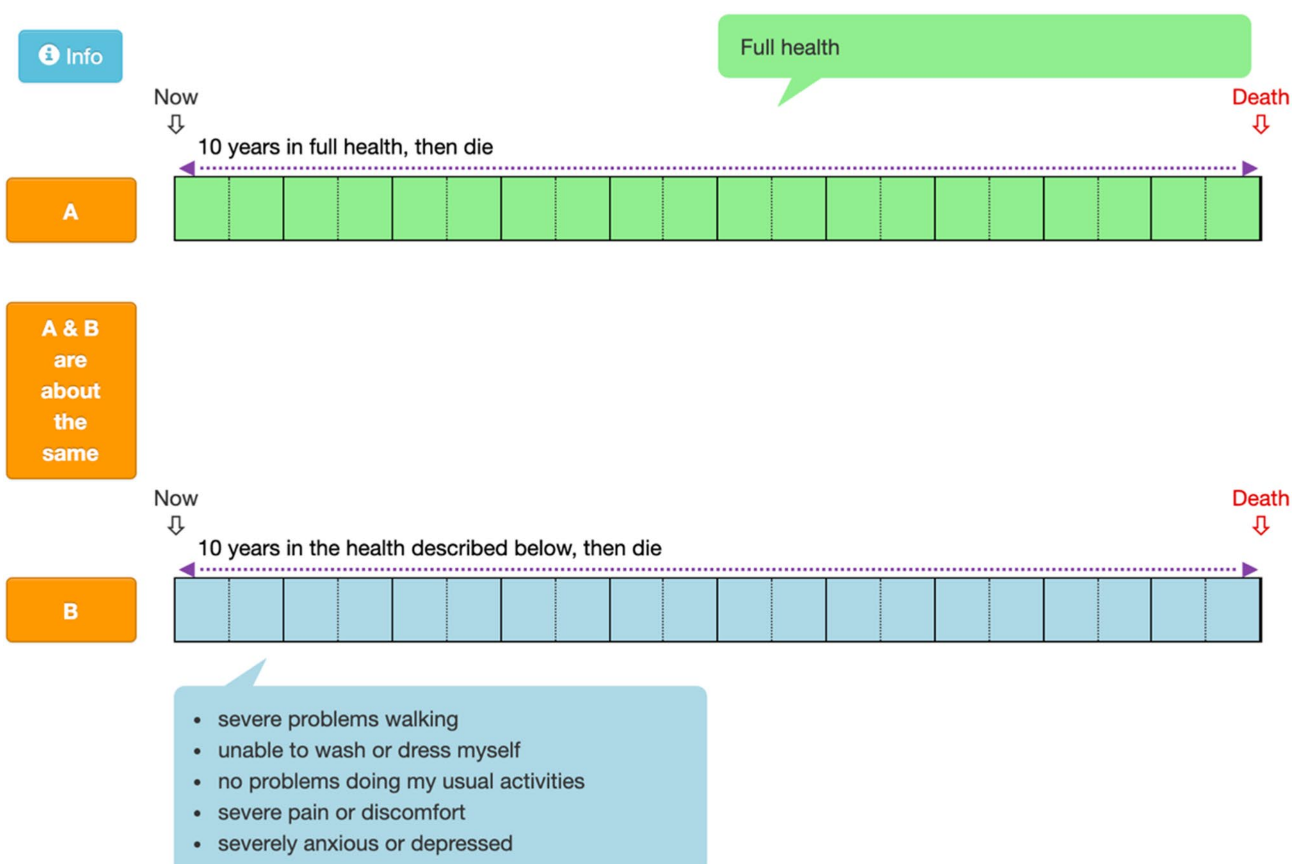

Fig. 5 composite time trade-off task: conventional TTO for better-than-dead preference elicitation

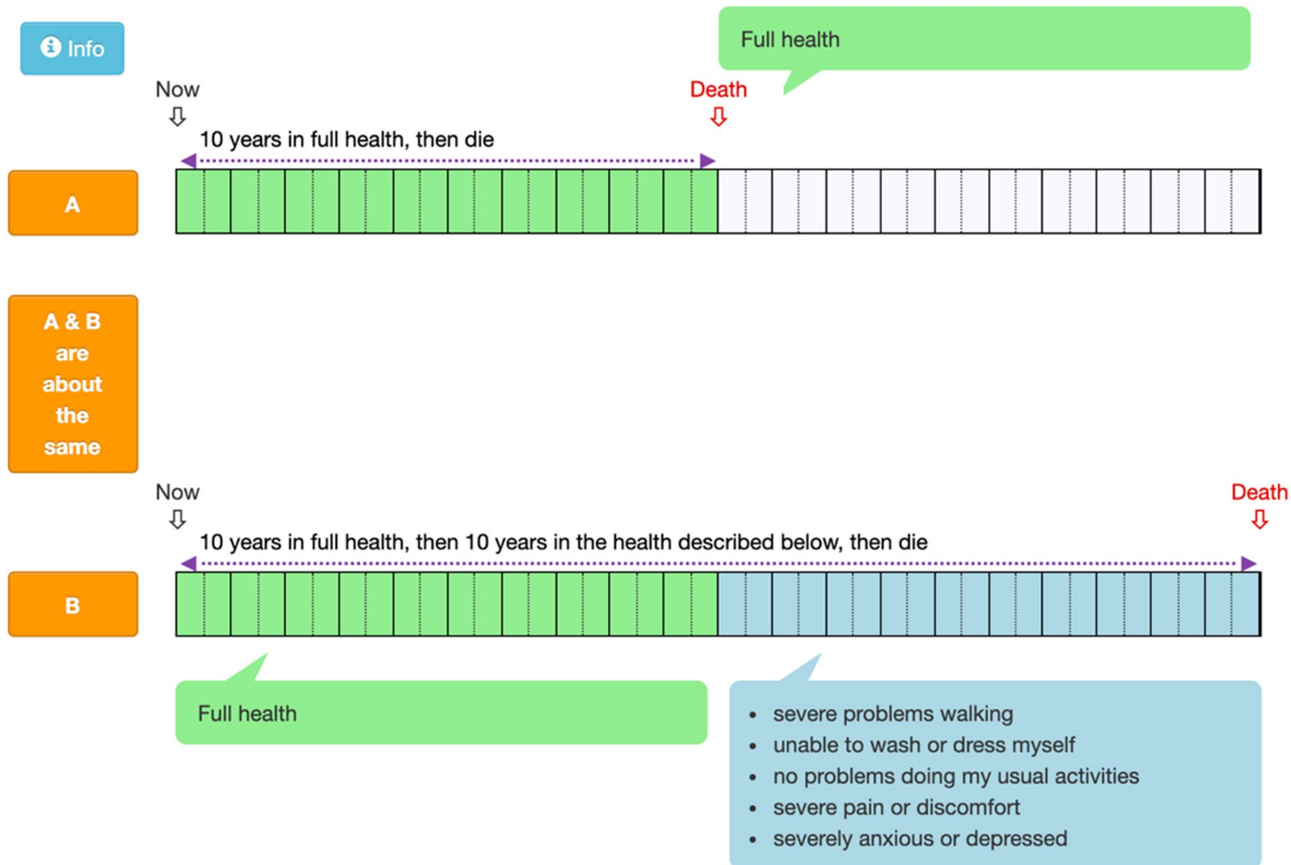

Fig. 6 Composite time trade-off task: lead-time TTO for worse-than-dead preference elicitation

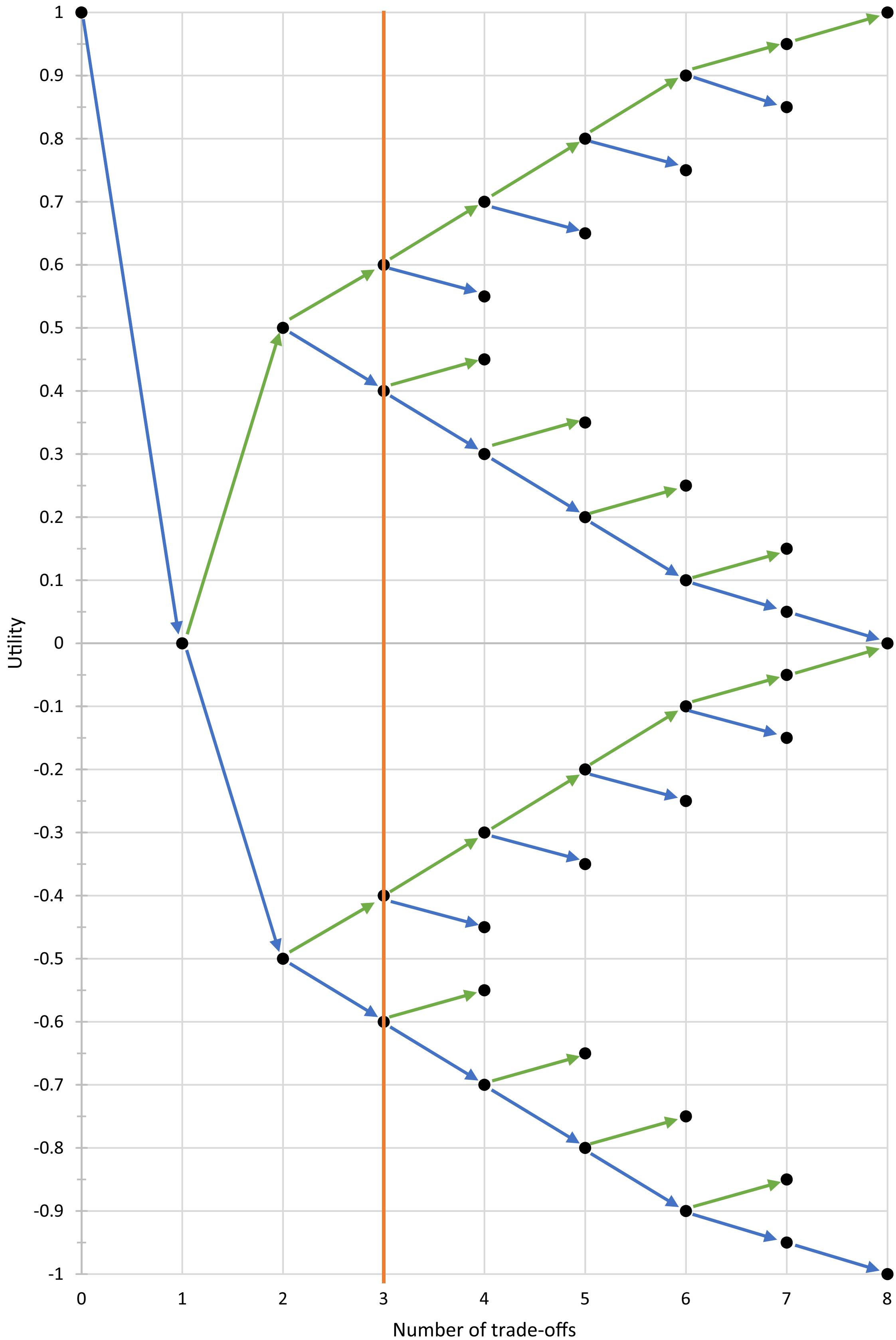

Fig. 7 Planned TTO routing adapted from (Stolk 2019)

References

1. Whitehead, S.J., Ali, S.: Health outcomes in economic evaluation: the QALY and utilities. Br. Med. Bull. **96**, 5–21 (2010). https://doi.org/10.1093/bmb/ldq033
2. Oppe, M., Devlin, N.J., van Hout, B., Krabbe, P.F., de Charro, F.: A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. Val. Health. **17**(4), 445–453 (2014). https://doi.org/10.1016/j.jval.2014.04.002
3. Boye, K.S., Matza, L.S., Feeny, D.H., Johnston, J.A., Bowman, L., Jordan, J.B.: Challenges to time trade-off utility assessment methods: when should you consider alternative approaches? Expert. Rev. Pharmacoecon. Outcomes. Res. **14**(3), 437–450 (2014). https://doi.org/10.1586/14737167.2014.912562
4. Xie, F., Pullenayegum, E., Pickard, A.S., Ramos Goni, J.M., Jo, M.W., Igarashi, A.: Transforming latent utilities to health utilities: east does not meet west. Health. Econ. **26**(12), 1524–1533 (2017). https://doi.org/10.1002/hec.3444
5. Hadi, M., Swinburn, P., Nalysnyk, L., Hamed, A., Mehta, A.: A health state utility valuation study to assess the impact of treatment mode of administration in Gaucher disease. Orphanet. J. Rare. Dis. **13**(1), 159 (2018). https://doi.org/10.1186/s13023-018-0903-6
6. King, M.T., Viney, R., Simon Pickard, A., Rowen, D., Aaronson, N.K., Brazier, J.E., Cella, D., Costa, D.S.J., Fayers, P.M., Kemmler, G., McTaggart-Cowen, H., Mercieca-Bebber, R., Peacock, S., Street, D.J., Young, T.A., Norman, R., Consortium, M.A.: Australian utility weights for the EORTC QLU-C10D, a multi-attribute utility instrument derived from the cancer-specific quality of life questionnaire, EORTC QLQ-C30. Pharmacoeconomics. (2017). https://doi.org/10.1007/s40273-017-0582-5
7. Norman, R., King, M.T., 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. **19**(4), 499–508 (2010). https://doi.org/10.1007/s11136-010-9609-5
8. Edelaar-Peeters, Y., Stiggelbout, A.M., Van Den Hout, W.B.: Qualitative and quantitative analysis of interviewer help answering the time tradeoff. Med. Decis. Making. **34**(5), 655–665 (2014). https://doi.org/10.1177/0272989X14524989
9. Maniaci, M.R., Rogge, R.D.: Caring about carelessness: participant inattention and its effects on research. J. Res. Pers. **48**, 61–83 (2014). https://doi.org/10.1016/j.jrp.2013.09.008
10. Paas, L.J., Morren, M.: PLease do not answer if you are reading this: respondent attention in online panels. Mark. Lett. **29**(1), 13–21 (2018). https://doi.org/10.1007/s11002-018-9448-7
11. Bansback, N., Tsuchiya, A., Brazier, J., Anis, A.: Canadian valuation of EQ-5D health states: preliminary value set and considerations for future valuation studies. PLoS ONE **7**(2), e31115 (2012). https://doi.org/10.1371/journal.pone.0031115
12. Jiang, R., Shaw, J.W., Pickard, A.S.: PS2-37 Comparison of online and face-to-face valuation of the EQ-5D-5L using composite time trade-off and discrete choice tasks. In: 40th Annual Meeting of the Society for Medical Decision Making, Montréal, Québec, Canada, vol. 39, pp. E1–E343. SAGE Publications Inc STM (2018). https://doi.org/10.1177/0272989X18821191
13. Jiang, R., Muhlbacher, A., Shaw, J.W., Lee, T.A., Walton, S.M., Pickard, A.S. Comparing online and face-to-face data quality and preferences in a health valuation study. In: 10th Meeting "International Academy of Health Preference Research", Basel, Switzerland, vol. 12, pp. 429–435 (2019). https://doi.org/10.1007/s40271-019-00368-9
14. Oppe, M.: Personal Communication. In. (2018)
15. Lenert, L.A., Cher, D.J., Goldstein, M.K., Bergen, M.R., Garber, A.: The effect of search procedures on utility elicitations. Med. Decis. Making. **18**(1), 76–83 (1998). https://doi.org/10.1177/0272989X9801800115
16. van Reenen, M., Janssen, B.: EQ-5D-5L user guide-basic information on how to use the EQ-5D-5 L instrument. EuroQol Group, Rotterdam (2013)
17. Versteegh, M.M., Attema, A.E., Oppe, M., Devlin, N.J., Stolk, E.A.: Time to tweak the TTO: results from a comparison of alternative specifications of the TTO. Eur. J. Health. Econ. **14**(Suppl 1), S43–51 (2013). https://doi.org/10.1007/s10198-013-0507-y
18. Devlin, N., Buckingham, K., Shah, K., Tsuchiya, A., Tilling, C., Wilkinson, G., van Hout, B.: A comparison of alternative variants of the lead and lag time TTO. Health. Econ. **22**(5), 517–532 (2013). https://doi.org/10.1002/hec.2819
19. Pickard, A.S., Law, E.H., Jiang, R., Pullenayegum, E., Shaw, J.W., Xie, F., Oppe, M., Boye, K.S., Chapman, R.H., Gong, C.L., Balch, A., Busschbach, J.J.V.: United states valuation of EQ-5D-5L health states using an international protocol. Val. Health. (2019). https://doi.org/10.1016/j.jval.2019.02.009
20. Stolk, E., Ludwig, K., Rand, K., van Hout, B., Ramos-Goni, J.M.: Overview, update, and lessons learned from the international EQ-5D-5L valuation work: version 2 of the EQ-5D-5L valuation protocol. Val. Health. **22**(1), 23–30 (2019). https://doi.org/10.1016/j.jval.2018.05.010
21. Oppe, M., Rand-Hendriksen, K., Shah, K., Ramos-Goni, J.M., Luo, N.: EuroQol protocols for time trade-off valuation of health outcomes. Pharmacoeconomics. **34**(10), 993–1004 (2016). https://doi.org/10.1007/s40273-016-0404-1
22. Luo, N., Li, M., Stolk, E.A., Devlin, N.J.: The effects of lead time and visual aids in TTO valuation: a study of the EQ-VT framework. Eur. J. Health. Econ. **14**(Suppl 1), S15–24 (2013). https://doi.org/10.1007/s10198-013-0504-1
23. Oppe, M., van Hout, B.: The "power" of eliciting EQ-5D-5L values: the experimental design of the EQ-VT (2017). https://euroqol.org/wp-content/uploads/2016/10/EuroQol-Working-Paper-Series-Manuscript-17003-Mark-Oppe.pdf
24. Pew Research Center: Internet/Broadband Factsheet. https://www.pewinternet.org/fact-sheet/internet-broadband/ (2018). Accessed June 6 2019
25. Anderson, M., Perrin, A., Jiang, J., Kumar, M.: 10% of Americans don't use the internet. Who are they? (2018). http://www.pewresearch.org/fact-tank/2018/03/05/some-americans-dont-use-the-internet-who-are-they/
26. Hanmer, J., Hays, R.D., Fryback, D.G.: Mode of administration is important in US national estimates of health-related quality of life. Med. Care. **45**(12), 1171–1179 (2007). https://doi.org/10.1097/MLR.0b013e3181354828
27. Duffy, B., Smith, K., Terhanian, G., Bremer, J.: Comparing data from online and face-to-face surveys. Int. J. Mark. Res. **47**(6), 615–639 (2005). https://doi.org/10.1177/147078530504700602
28. Lenert, L.A., Sturley, A., Rupnow, M.: Toward improved methods for measurement of utility: automated repair of errors in elicitations. Med. Decis. Making. **23**(1), 67–75 (2003). https://doi.org/10.1177/0272989X02239649

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.