Valuing Treatments for Parkinson Disease Incorporating Process Utility: Performance of Best-Worst Scaling, Time Trade-Off, and Visual Analogue Scales

Marieke G.M. Weernink, MSc*, Catharina G.M. Groothuis-Oudshoorn, PhD, Maarten J. IJzerman, PhD, Janine A. van Til, PhD

Department of Health Technology and Services Research, MIRA Institute for Biomedical Technology and Technical Medicine, University of Twente, Enschede, The Netherlands

A B S T R A C T

Objective: The objective of this study was to compare treatment profiles including both health outcomes and process characteristics in Parkinson disease using best-worst scaling (BWS), time trade-off (TTO), and visual analogue scales (VAS). **Methods:** From the model comprising of seven attributes with three levels, six unique profiles were selected representing process-related factors and health outcomes in Parkinson disease. A Web-based survey (N = 613) was conducted in a general population to estimate process-related utilities using profile-based BWS (case 2), multiprofile-based BWS (case 3), TTO, and VAS. The rank order of the six profiles was compared, convergent validity among methods was assessed, and individual analysis focused on the differentiation between pairs of profiles with methods used. **Results:** The aggregated health-state utilities for the six treatment profiles were highly comparable for all methods and no rank reversals were identified. On the individual level, the convergent validity between all methods was strong; however, respondents differentiated less in the utility of closely related treatment profiles with a VAS or TTO than with BWS. For TTO and VAS, this resulted in nonsignificant differences in mean utilities for closely related treatment profiles. **Conclusions:** This study suggests that all methods are equally able to measure process-related utility when the aim is to estimate the overall value of treatments. On an individual level, such as in shared decision making, BWS allows for better prioritization of treatment alternatives, especially if they are closely related. The decision-making problem and the need for explicit trade-off between attributes should determine the choice for a method.

Keywords: health-state utility, Parkinson disease, preference, process utility.

© 2016 Published by Elsevier Inc. on behalf of International Society for Pharmacoeconomics and Outcomes Research (ISPOR).

Introduction

Patients with chronic diseases are confronted with reduced quality of life and decreased length of life while the management of chronic disease results in high costs for society [1]. Management of chronic disease is generally aimed at symptom control and improving daily functioning, thus limiting the effects of living with a chronic disease and preventing further functional loss [2].

It is generally accepted that optimizing the process of care in patients with chronic conditions is of paramount importance [3,4]. However, because of the chronicity, current approaches to valuing outcomes may not be sensitive nor do they capture all value components. First, effective disease management requires adherence to guidelines and treatments prescribed. However, treatment adherence is reported to be moderate in many patient groups suffering from chronic diseases [5]. This can be explained because the treatment primarily has a preventive purpose, whereas negative consequences such as adverse events are experienced immediately. Second, the ease of use or convenience while participating in a disease management program generally is not captured, although these elements constitute a value relevant to patients. Therefore, recent studies recommend that the valuation space should include a broader range of patients' experiences such as process of care factors and factors that relate to enabling individuals to the best they can be [6,7].

Even though many people advocate widening the evaluation space, the traditional health economist view does not consider that factors such as the process of receiving care and valuation are restricted to health outcomes alone [8,9]. In this view, the process of care has value only because it is a commodity that can be exchanged to derive health gains [10]. Thus, lack of treatment adherence or a less desirable process of care will be reflected in the health outcomes achieved and do not require explicit valuation.

Despite these different views, the present article assumes that process-related attributes do affect the value of a treatment independent from health outcomes achieved. Hence, to determine the value of a treatment in chronic disease it is important to go beyond health outcomes and include attributes such as the ease of use and process of care.

Standard gamble (SG), time trade-off (TTO), and visual analogue scales (VAS) are common methods to value health outcomes. SG is mostly used in economic evaluation studies and clinical decision analyses. SG is considered to be the "criterion standard" for utility measurement because it involves making choices under conditions of uncertainty and has rigorous foundations in expected utility theory [11]. Alternatively, TTO forces trade-offs between length of life and quality of life under conditions of certainty. TTO is well known for being the standard approach to derive health utility weights for EuroQol five-dimensional questionnaire (EQ-5D) health states (EQ-5D tariff) [12,13]. A VAS asks respondents to rate health states on a scale anchored at "worst imaginable health state" and "best imaginable health state" under conditions of certainty. They are easy and are often used in combination with the EQ-5D [14,15].

Although TTO, SG, and VAS can be used to capture process-related utility in addition to health outcomes, they are not frequently used for that purpose. A review by Brennan and Dixon [4] concluded that direct valuation methods SG and TTO have been used in only 13 studies to assess process-related utility.

Another approach being used to estimate process utility is discrete choice experiment (DCE) and best-worst scaling (BWS) methods [16,17]. In BWS, respondents are asked to choose the best and the worst from a number of options. Essentially, respondents are asked to choose the pair that maximizes the difference in value (on the latent utility scale) between them. BWS case 3 (BWS-3)—which is comparable to a standard DCE choice set—asks respondents to choose the best and the worst from a number of alternative treatment options, which can be composed from multiple criteria. Alternatively, BWS case 2 (BWS-2) lets respondents consider one treatment at a time, asking for the best and worst characteristics within a profile. The indirect valuations are used to estimate the utility of multiple criteria.

DCE and BWS are more often used to explicitly value process factors and nonhealth outcomes than are methods such as VAS and TTO [4,6,18]. Their popularity might be explained by the lower degree of abstract reasoning required to answer a task of DCE and BWS and its methodological rooting in random utility theory [16,18–22]. However, DCE and BWS have outcomes on a latent scale, which means that no reference can be made to the health utility scale (0–1), which is favored in health technology assessment. The potential of BWS and DCE to estimate health utilities from part-worth utilities using anchoring and mapping techniques is currently being studied [20,23–25].

One study was found that compared VAS, TTO, and DCE using vignettes depicting moderate-risk pregnancy at term (including process factors). The authors concluded that DCE was superior to TTO and performed equal to VAS with regard to validity and reliability and that DCE had slightly higher user feasibility [26]. However, because of a lack of head-to-head comparisons, it is unclear which methods differentiate best between health states.

The objective of this article was to compare the health-state utilities incorporating a process-related attribute for treatment profiles in Parkinson disease (PD) using TTO, VAS, and BWS-2 and BWS-3. Several unique treatment profiles were identified in which the process and outcomes of the treatment differed. A head-to-head comparison of aggregated utility scores and the comparison of individual scores were included. Because individuals used all the methods to evaluate treatment profiles, we had the possibility to conduct a within-person comparison analysis to study whether individuals were able to differentiate between the values of treatment profiles with all methods. PD was chosen because of the specific problems in managing the disease over a longer time and the different treatment modalities that are used during this time. Patients on drug treatment eventually may undergo neurosurgery or will receive their medication through pump infusion. Such treatments differ from pharmacological therapy in its process because they require surgery, daily cleaning routines, and use of mechanical equipment.

Methods

BWS Experiments

Relevant attributes of care were identified through literature review and qualitative interviews with 15 patients with PD. Seven attributes were selected: process of care, resting tremor, posture and balance problems, slowness of movement, dizziness, drowsiness, and dyskinesia. Symptoms and adverse effects were assigned the same level scale values (from "seldom to never" to "sometimes" or "often suffer from"). The "process of care" attribute was described as the oral intake of tablets, continuous pump infusion of medication, and neurosurgery. In the BWS-2 experiment, respondents were asked to select the aspects of treatment that they perceived as the most and the least preferable within one single treatment profile. For the BWS-3 experiment, respondents were asked to select the most and the least desirable treatments from three treatment profiles. It was not feasible to provide respondents with a full factorial design: 2187 (3^7) possible profiles. Experimental-design software from Sawtooth Software was used to generate a D-efficient design and four blocks [27,28]. In total, each respondent answered nine profile tasks for BWS-2 and 10 choice tasks for BWS-3. Appendix 1 in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2015.11.011 displays examples of a BWS-2 task and a BWS-3 task.

Treatment Profile Valuations

Six hypothetical core treatment profiles were defined: both extreme profiles and four intermediate (closely related) treatment profiles (Table 1). A full variation of the process attribute was not included in the six core scenarios because we also wanted to test the effect of small changes in health outcomes on utility scores. All respondents valued both the best and worst treatment profiles and were randomly assigned to value two of the four intermediate profiles with both a TTO and a VAS. The end points of the VAS were labeled as "best imaginable treatment" and "worst imaginable treatment."

A computer-based TTO was executed to determine how many years of life lived in perfect health followed by death would be equivalent to 10 years of life lived in the particular PD treatment profile followed by death.

Respondents were asked to put themselves in the hypothetical scenarios and were asked whether they were willing to give up any life-years to return to full health again. If they did, a second question determined whether the respondent was willing to give up 5 life-years. This was followed by an iteration procedure in which the number of life-years was increased or decreased until the point of indifference was reached.

Sampling

A Web-based survey was conducted in June 2013 to measure treatment preferences for PD management in the United Kingdom and the Netherlands. The respondent sample consisted of members of the general population. Survey Sampling International (Rotterdam, the Netherlands) recruited 613 respondents (aged 18–65 years) from large panels, which have gone through rigorous quality controls. The sample size was based on a rule of thumb for conjoint analysis: estimate precision increases quickly at sample sizes of less than 150 and flattens out at ±300 observations [27]. Sample size calculation for TTO and VAS was not applicable, because no expectations were set regarding the desired difference in utility scores between treatment profiles.

Attributes of PD treatment	Worst	Intermediate 1	Intermediate 2	Intermediate 3	Intermediate 4	Best
Process of care	Brain surgery	Medication	Medication	Medication	Medication	Medication
Resting tremor	Often	Often	Seldom to never	Seldom to never	Sometimes	Seldom to never
Posture and balance problems	Often	Seldom to never	Seldom to never	Often	Sometimes	Seldom to never
Slowness of movement	Often	Seldom to never	Seldom to never	Often	Sometimes	Seldom to never
Dizziness	Often	Often	Often	Seldom to never	Seldom to never	Seldom to never
Drowsiness	Often	Sometimes	Often	Often	Seldom to never	Seldom to never
Dyskinesia	Often	Often	Often	Seldom to never	Seldom to never	Seldom to never

Table 1 – Six core treatments selected for valuation with time trade-off and a visual analogue scale.

While completing the survey, the respondents were instructed to consider being a patient with PD, and they were given a standard sheet of clinical background information.

Data Analysis

Respondents who completed the survey within 5 minutes were excluded from analysis. Furthermore, BWS data were checked using the Empirical Scale Parameter (ESP), which shows the extent to which people are consistent in their choices [29]. Respondents were excluded if they either chose randomly (ESP < 2) or used an elimination-by-aspect strategy (ESP > 5.5).

For some respondents the profiles were all of equal value, the so-called nondifferentiators. Nondifferentiators include respondents who did not want to trade any life-years (TTO), respondents who traded the same number of life-years for each of the four evaluated treatment profiles (TTO), and respondents who assigned equal values to each profile on the scale from 0 to 100 (VAS). Inherent to the BWS method, respondents were forced to make choices, which in principle leads to a difference in utilities for the treatment profiles. Nondifferentiators were excluded from the main analysis to investigate whether the remainder of respondents, as a group and as individuals, made significant differences in value between treatment profiles.

The main analysis consisted of two components: 1) comparison of aggregated utility scores and 2) comparison of individual scores (within-person comparison). From the final data set, aggregated TTO and VAS utilities for each of the six treatment profiles were calculated using the mean. BWS-2 and BWS-3 utility weights (β) for each attribute level were estimated using a conditional logit model. These β weights allow for utility estimation of the six selected treatment profiles. Because of the latent scale of BWS, the derived BWS-2 and BWS-3 utility weights could not be directly compared with each other and compared with TTO/VAS utilities. Therefore, a rank order comparison was used to observe whether the same order from best to worst treatment profile existed for each method. The convergent validity between the aggregated utilities of all methods was assessed using the squared Pearson correlation coefficient (across-respondents means) [30]. To study whether the hypothesis that VAS and TTO use a smaller part of their scale to value small changes in treatments was true, it was calculated which part of each method's scale was needed to differentiate between the value of the worst, intermediate, and best profiles. Furthermore, the importance of each attribute according to BWS-2 and BWS-3 was compared by estimating the difference between minimum and maximum β weights within an attribute divided by the total variance summed across all attributes. In addition, the relative size of BWS-2 and BWS-3 differences was adjusted by using TTO and VAS valuations of the best and worst treatment profiles to "anchor" BWS valuations on the VAS/TTO scale of 0 to 1 [31]. Subsequently, scatterplots were used to identify (scaling) differences and outliers.

Individual utilities were analyzed to study whether the respondents were able to differentiate between the values of the six treatment profiles with all methods. Mixed logit models were used to estimate the individual BWS-2 and BWS-3 utility weights for the best and the worst profiles, and for the two intermediate profiles for which the respondent also has provided a TTO and VAS valuation. For each pair of treatment profiles, the number of nondifferentiators was identified. The difference with previously excluded nondifferentiators is that these respondents at least made one differentiation between the four evaluated profiles; for example, respondents might have indicated a difference between the best and the worst profile but not between two intermediate profiles. Individual differences between all pairs of two treatment profiles were calculated and tested with a paired t test. The convergent validity between the individual utilities was studied using the squared Pearson correlation coefficient (within-respondents mean) [30]. Data were analyzed using Stata version 13 (StataCorp, College Station, TX). P values of less than 0.05 were considered statistically significant.

Results

Six hundred thirteen out of 855 respondents completed the survey (response rate 72%). The mean completion time was 25 ± 8.2 minutes. On the basis of time criterion (<5 minutes), 18 respondents were excluded. Sixteen respondents were excluded on the basis of BWS-consistency criterion (ESP <2 or >5.5). Less than 2% of respondents were diagnosed with PD, and 22.7% of the respondents had a friend or family member diagnosed with PD.

Nondifferentiators

One hundred forty-one (23%) respondents were not willing to trade life-years for any of the four treatment profiles. Almost 4% of respondents were willing to trade life-years, but did not vary the number of life-years per profile. In contrast, less than 2% of respondents assigned equal values to each of the four evaluated profiles with VAS. The data of 157 nondifferentiators were excluded from the main analysis to investigate whether the remainder of respondents (N=422) did make significant differences in value between treatment profiles with TTO and VAS.

Aggregated Utility Scores of Hypothetical Profiles

Table 2 presents the aggregated mean utility for each of the treatment profiles. Each of the mean utility scores follows a monotonic relationship: both VAS and TTO utilities increase if BWS utilities increase. There were no rank reversals, and the group of respondents was able to distinguish between the best, the worst, and all intermediate profiles. Squared Pearson correlation coefficients also show that the convergent validity between all methods is very strong (VAS-BWS-2 0.97; VAS-BWS-3 0.98; TTO-BWS-2 0.97; TTO-BWS-3 0.98; TTO-VAS 0.95; BWS-2-BWS-3 0.98; P = 0.000, N = 6). Overall, respondents gave higher TTO

Table 2 – Mean utilities for the six core treatment profiles (mean/standard error).				
Core treatment profiles	TTO	VAS	BWS-2	BWS-3
Worst*	0.34 (0.01)	0.25 (0.01)	−2.41 (0.85)	−1.10 (0.04)
Intermediate 1†	0.59 (0.02)	0.48 (0.01)	11.16 (0.83)	1.75 (0.07)
Intermediate 2†	0.69 (0.02)	0.52 (0.01)	14.5 (0.84)	2.22 (0.07)
Intermediate 3†	0.78 (0.02)	0.55 (0.01)	17.24 (0.97)	2.65 (0.08)
Intermediate 4†	0.88 (0.02)	0.65 (0.01)	19.41 (0.98)	4.17 (0.11)
Best*	0.93 (0.01)	0.78 (0.01)	24.89 (0.99)	4.74 (0.12)

BWS-2, best-worst scaling case 2; BWS-3, best-worst case 3; TTO, time trade-off; VAS, visual analogue scale.
*422 observations.
†200 observations (each respondent valued only two of the four intermediate treatment profiles).

utilities for treatment profiles compared with VAS utilities. In TTO, respondents used 49% [(0.88 − 0.59)/(0.93 − 0.34)] of the scale to differentiate between the four intermediate treatment profiles, whereas with BWS-2 this was only 30%, with BWS-3 41%, and with VAS 32%. The largest difference in utility was found between the utility of the worst profile and the utility of intermediate treatment 1: BWS-2 50%, BWS-3 49%, VAS 43%, and TTO 42%.

According to the attribute importance calculated with both BWS-2 and BWS-3, process of care had the largest impact on overall treatment valuation in patients with PD. When the results of BWS-2 and BWS-3 were compared, there was a difference in ranking of attributes: there were rank reversals between the attributes dizziness and resting tremor and the attributes dyskinesia and drowsiness (Fig. 1). The rescaled BWS data (Fig. 2,A,B) show that the shape of both BWS-2 and BWS-3 utilities is similar, except for the fact that respondents assigned slightly higher utilities to treatment profiles with BWS-3 than with BWS-2. Furthermore, the scatterplots show that TTO utilities for intermediate profiles were more in line with BWS$_{TTO}$ rescaled utilities (2A) than with VAS utilities compared with BWS$_{VAS}$ rescaled utilities (2B).

Individual Scores of Hypothetical Profiles (within-Person Comparison)

Table 3 presents the mean difference in individual utilities for all pairs of treatment profiles. In addition, the number of respondents who did not differentiate the utility of pairs of profiles is presented for TTO and VAS. Overall, when respondents used a TTO to state their preference for the treatment alternatives, 30% of 2477 pairs of treatment profiles did not result in a preference for one of the two profiles. Conversely, for VAS valuations this was only 2.5%. Because of the forced choice, all respondents stated their preference for treatment profiles with BWS-2 and

BWS-3. Most respondents gave a higher preference to the best-case scenario than to the worst-case scenario with all methods. Only a few respondents (TTO 5%; VAS 1%) were indifferent between the extreme profiles, but gave different valuations for at least one other pair of treatment profiles. More respondents were indifferent between the value of pairs of intermediate profiles (TTO 33%–58%; VAS 1%–13%). As a result, only 8% of respondents had a ranking of treatment profiles that was similar to the overall ranking with TTO (VAS 56%; BWS-2 94%; BWS-3 91%). Furthermore, respondents were not able to make a significant difference in value between intermediate profiles 3 and 2 with both TTO and VAS, and intermediate profiles 2 and 1 (only TTO). For all methods, the mean difference in estimated utility between the intermediate profiles and either extreme profile decreased when the intermediate profile was more similar to one of the extreme profiles. At the same time, the rate of nondifferentiators for TTO increased when the intermediate profile was more similar to one of the extreme profiles (not observed for VAS). Squared Pearson mean within-respondent correlation showed that the convergent validity between all methods is strong (VAS-BWS-2 0.67; VAS-BWS-3 0.68; TTO-BWS-2 0.58; TTO-BWS-3 0.59; TTO-VAS 0.56; BWS-2-BWS-3 0.97; P = 0.000, N = 422).

Discussion

The aim of this study was to compare the health-state utilities incorporating a process-related attribute for treatment profiles in general public using TTO, VAS, BWS-2, and BWS-3. The results of the study indicate that all the methods used are able to measure process-related utility: BWS, TTO, and VAS perform equally well with regard to distinguishing treatment alternatives on the basis of aggregated utility scores (no rank reversals and highly comparable). However, on an individual level, respondents differentiated less in the utility of intermediate treatment profiles with a VAS or TTO (nonsignificant differences) than with BWS. On the individual level, BWS allows for better prioritization of treatment alternatives and provides evidence on those characteristics that are most important to respondents. Our results confirm the findings of Bijlenga et al. [26] who have concluded that VAS, TTO, and DCE lead to the same outcomes, but on the basis of their reliability and validity tests, they reported DCE to be superior to TTO and equal to VAS. Based on the results of our study, neither of the methods can be directly favored over the other, but each method has its advantages and disadvantages in estimating process-related utility on different levels of decision making, which will be discussed.

With the VAS, most of the respondents in our study were able to differentiate between the value of treatment profiles. The results of a treatment valuation with VAS allows for prioritization of outcomes and treatment choice on the individual level. A VAS
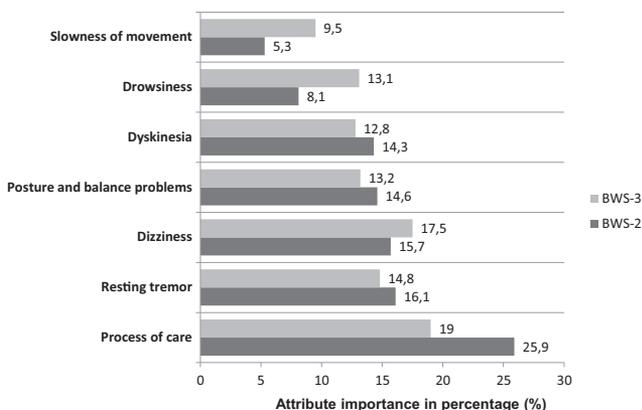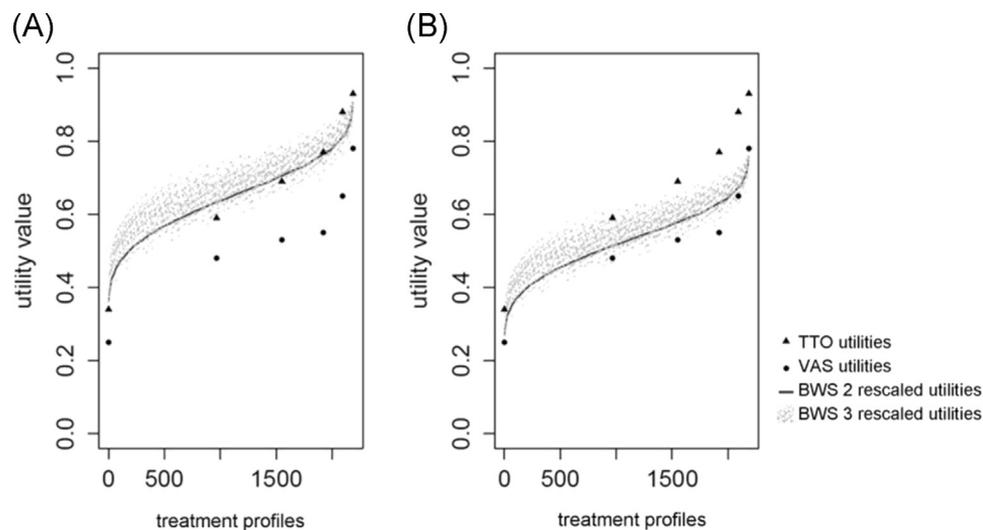

Figure 1 – Attribute importance according to BWS-2 and BWS-3 utility weights (N = 422).

* Treatment scenarios were sorted on the utilities of BWS-2.

Figure 2 – Scatterplots of BWS-2/BWS-3 utilities for 2187 (3⁷) treatment profiles which were rescaled on the quality of life scale (0-1) with the TTO (2A) and VAS (2B) data of the best and worst profile.

can be used in clinical decision making if there is only a need to inform on relative treatment preferences (e.g., choice for treatment A vs. treatment B). VAS is a simple and cognitively undemanding method. However, it does not involve uncertainty or any trade-off between gains and possible harms (i.e., length of life, higher risk of). Because this is not in agreement with utility theory, the use of a VAS as a valuation tool is not useful in economic evaluations [32].

With TTO, 25% of respondents were not willing to trade-off life-years with any of the treatment scenarios. This might have resulted from the unrealistic choice a TTO poses, religious beliefs, or the abstract reasoning of life and death. Yet, 25% of non-differentiators is not seen as a high rate in literature [33]. With TTO, individual respondents were able to differentiate between the extreme profiles, but not always between intermediate profiles. The minor variations within these treatment profiles deter people from trading life-years. These are well-recognized limitations of TTO, which were shown previously in literature [34,35]. It can be concluded that TTO is less suited for measuring the effects of small changes in treatment on individual level, such as clinical decision making. In contrast to our expectations, the respondents (group) were able to use TTO to differentiate between treatments that have minor variations in outcomes or a different process of care. Remarkably, compared with other PD literature, our respondents used large parts of the VAS and the TTO scale to value (intermediate) PD treatments [36–38]. Therefore, TTO is useful in decision making on the societal level. TTO as a valuation method is common practice in economic evaluation, and the method is widely recognized as a valid method for utility measurement [39]. However, if one decides to exclude nondifferentiators, one should be aware that the respondents who do trade life-years might be a selective group and preferences might be surrounded by uncertainty [35].

Inherent to the method, BWS is more able to disaggregate the health-state utility to a single attribute level, providing evidence on those characteristics relevant to respondents. The results of BWS allow for estimation of treatment valuation on the individual level, thereby supporting clinical decision making. Compared with the number of nondifferentiators in TTO, more respondents engaged with the BWS task. Although VAS is a simple method, BWS overcomes the measurement bias of VAS by using discrete choices and trading of characteristics [40]. For the use of BWS as a

valuation tool on a societal level, the method lacks scaling of the treatment utilities on a general, disease-exceeding level (health utility scale 0–1). Several options are available to rescale BWS data on the quality-of-life scale [24,31]. However, as long as a sound methodological framework for mapping DCE/BWS utilities is lacking, SG and TTO will remain the criterion standard. The establishment of an ISPOR task force on mapping utilities with the aim to develop good practice guidelines can be seen as an important step toward the use of BWS and DCE in health technology assessment [25]. However, SG continues to have the advantage of letting respondents make choices under conditions of uncertainty. Furthermore, our results have shown that utilities derived from BWS-2 or BWS-3 are highly comparable. Our recommendation is that the characteristics of the respondents (such as cognitive state) should influence the choice for a specific BWS case.

The results of this study also showed that the process of care is an attribute that was relevant to respondents. Whether or not it is relevant to include it in a measure of health depends on its contexts. We feel it is very relevant to include in clinical decision making. However, in societal decision making it is debatable whether valuations should include process factors [6,8]. Yet, the Belgium Health Care Knowledge Centre has included discomfort/inconvenience of current treatment as an attribute in a DCE to measure societal preferences in reimbursement decisions in Belgium [41].

Strengths and Limitations

A strength of this head-to-head comparison of methods is that besides the aggregated utility analysis we were able to perform a within-person comparison of utility estimates. Another strength was that the BWS exercises allowed for insights into the impact of process utility on the overall value of treatment. In principle, TTO/VAS cannot estimate this; however, if one uses an experimental design to select the TTO/VAS treatment profiles this is possible. Criteria weights can be estimated and compared with BWS indirect utility weights. This gives the possibility to measure the precise influence of process of care on treatment valuation with TTO, VAS, and BWS. In addition, including a full variation of

Table 3 – Mean utility difference between possible combinations of the six core profiles, and the rate of nondifferentiators for TTO and VAS.

Profiles	N	TTO mean difference (SE)	TTO non-differentiators, n (%)	VAS mean difference (SE)	VAS non-differentiators, n (%)	BWS-2 mean difference (SE)	BWS-3 mean difference (SE)
Best & worst	422	0.59 (0.016)*	20 (5)	0.53 (0.014)*	4 (1)	34.58 (0.280)*	7.47 (0.064)*
Best & intermediate 4	215	0.06 (0.015)*	154 (72)	0.14 (0.010)*	7 (3)	6.29 (0.028)*	0.78 (0.029)*
Best & intermediate 3	199	0.14 (0.024)*	107 (54)	0.22 (0.013)*	5 (3)	9.84 (0.150)*	2.71 (0.027)*
Best & intermediate 2	204	0.24 (0.025)*	93 (46)	0.25 (0.012)*	3 (1)	12.65 (0.102)*	3.15 (0.030)*
Best & intermediate 1	204	0.35 (0.024)*	68 (33)	0.30 (0.014)*	3 (1)	16.91 (0.125)*	3.68 (0.034)*
Intermediate 4 & 1	72	0.34 (0.039)*	24 (33)	0.18 (0.018)*	1 (1)	10.40 (0.185)*	2.88 (0.062)*
Intermediate 4 & 2	70	0.18 (0.039)*	34 (49)	0.14 (0.017)*	3 (4)	6.45 (0.180)*	2.40 (0.057)*
Intermediate 4 & 3	62	0.22 (0.044)*	40 (55)	0.06 (0.016)*	5 (7)	6.60 (0.353)*	0.99 (0.060)*
Intermediate 3 & 1	73	0.08 (0.040)*	21 (34)	0.11 (0.018)*	6 (10)	3.41 (0.218)*	1.96 (0.042)*
Intermediate 3 & 2	64	0.05 (0.043)	37 (58)	0.01 (0.018)	5 (8)	2.93 (0.368)*	0.45 (0.067)*
Intermediate 2 & 1	70	0.06 (0.033)	37 (53)	0.07 (0.015)*	9 (13)	4.34 (0.167)*	0.57 (0.030)*
Intermediate 4 & worst	215	0.55 (0.023)*	14 (7)	0.42 (0.016)*	3 (1)	28.13 (0.379)*	6.65 (0.087)*
Intermediate 3 & worst	199	0.45 (0.024)*	21 (11)	0.29 (0.017)*	1 (1)	24.35 (0.383)*	4.86 (0.092)*
Intermediate 2 & worst	204	0.35 (0.025)*	32 (16)	0.27 (0.017)*	4 (2)	21.78 (0.404)*	4.18 (0.083)*
Intermediate 1 & worst	204	0.23 (0.022)*	51 (25)	0.22 (0.014)*	3 (1)	17.95 (0.363)*	3.81 (0.092)*

*P value < 0.05

profiles will lead to more precise estimations of the difference (or nondifference) in value for a change in one of the attributes.

A computerized choice-based version of the TTO was used to simulate the iterative procedure of the interviewer-TTO, but this might have influenced results [42]. Our final sample of 422 respondents affords good precision for estimations of utilities (± 0.1 effect size). However, only 200 respondents were included for the analysis of the intermediate profiles, which resulted in slightly broader confidence bounds for utilities of those profiles and increased uncertainty among the preferences found compared with the best and worst profiles. Because many respondents were reluctant to trade-off life-years with TTO, the remainder of respondents could also have been a selective group with strong opinions. Still, it was a deliberate choice to exclude those respondents from analysis because beyond this study, either at the clinical level or at the societal level, respondents need to differ between the value of treatments to (for example) aid a decision for individual treatment or estimate cost-effectiveness.

Conclusions

The value of treatments in chronic diseases such as PD greatly depends on both clinical outcomes and process characteristics. If one wants to explicitly include process characteristics in scenarios, the decision-making problem, and the need for explicit trade-off between attributes, should determine the choice for a method to measure process-related utility. We conclude that on a societal level, when group preferences are relevant, BWS, TTO, and VAS perform equally well with regard to distinguishing treatment alternatives on the basis of process-related utility. In clinical decision making, BWS allows for better prioritization of treatment alternatives and provides evidence on those characteristics that are most important to respondents.

Acknowledgments

Supplemental Materials

Supplemental material accompanying this article can be found in the online version as a hyperlink at http://dx.doi.org/10.1016/j.jval.2015.11.011 or, if a hard copy of article, at www.valueinhealthjournal.com/issues (select volume, issue, and article).

REFERENCES

[1] Beaglehole R, Ebrahim S, Reddy S, et al. Prevention of chronic diseases: a call to action. Lancet 2007;370:2152–7.

[2] Stewart AL, Greenfield S, Hays RD, et al. Functional status and well-being of patients with chronic conditions: results from the Medical Outcomes Study. JAMA 1989;262:907–13.

[3] Coleman CI, Limone B, Sobieraj DM, et al. Dosing frequency and medication adherence in chronic disease. J Manag Care Pharm 2012;18:527–39.

[4] Brennan VK, Dixon S. Incorporating process utility into quality adjusted life years: a systematic review of empirical studies. Pharmacoeconomics 2013;31:677–91.

[5] Dunbar-Jacob J, Erlen JA, Schlenk EA, et al. Adherence in chronic disease. Annu Rev Nurs Res 2000;18:48–90.

[6] Ryan M, Kinghorn P, Entwistle VA, et al. Valuing patients' experiences of healthcare processes: towards broader applications of existing methods. Soc Sci Med 2014;106:194–203.

[7] Ryan M, Gerard K. Inclusiveness in the health economic evaluation space. Soc Sci Med 2014;108:248–51.

[8] Culyer AJ, Newhouse JP. Handbook of Health Economics. 1st ed. Amsterdam; New York: Elsevier; 2000.

[9] Mooney G. Communitarian claims and community capabilities: furthering priority setting? Soc Sci Med 2005;60:247–55.

[10] Donaldson C, Shackley P. Does "process utility" exist? A case study of willingness to pay for laparoscopic cholecystectomy. Soc Sci Med 1997;44:699–707.

[11] von Neumann J, Morgenstern O. Theory of Games and Economic Behavior. Princeton, NJ: Princeton University Press, 1944.

[12] Viney R, Norman R, King MT, et al. Time trade-off derived EQ-5D weights for Australia. Value Health 2011;14:928–36.

[13] Longworth L, Rowen D. Mapping to obtain EQ-5D utility values for use in NICE health technology assessments. Value Health 2013;16:202–10.

[14] Gafni A. The standard gamble method: what is being measured and how it is interpreted. Health Serv Res 1994;29:207–24.

[15] Weernink MGM, Janus SLM, van Til JA, et al. A systematic review to identify the use of preference elicitation methods in healthcare decision making. Pharm Med 2014;28:175–85.

[16] Flynn TN. Valuing citizen and patient preferences in health: recent developments in three types of best-worst scaling. Expert Rev Pharmacoecon Outcomes Res 2010;10:259–67.

[17] Flynn TN, Louviere JJ, Peters TJ, et al. Best–worst scaling: what it can do for health care research and how to do it. J Health Econ 2007;26:171–89.

[18] Ali S, Ronaldson S. Ordinal preference elicitation methods in health economics and health services research: using discrete choice experiments and ranking methods. Br Med Bull 2012;103:21–44.

[19] Louviere JJ, Woodworth GG. Best-Worst Scaling: A Model for Largest Difference Judgments. University of Alberta, Faculty of Business: Working Paper, 1990.

[20] Rowen D, Brazier J, Van Hout B. A comparison of methods for converting DCE values onto the full health-dead QALY scale. Med Decis Making 2015;35:328–40.

[21] Ratcliffe J, Couzner L, Flynn T, et al. Valuing child health utility 9D health states with a young adolescent sample: a feasibility study to compare best-worst scaling discrete-choice experiment, standard gamble and time trade-off methods. App Health Econ Health Pol 2011;9:15–27.

[22] Salomon JA. Reconsidering the use of rankings in the valuation of health states: a model for estimating cardinal values from ordinal data. Popul Health Metr 2003;1:12.

[23] Brazier J, Rowen D, Yang Y, et al. Comparison of health state utility values derived using time trade-off, rank and discrete choice data anchored on the full health-dead scale. Eur J Health Econ 2012;13:575–87.

[24] Flynn T, Louviere J, Marley A, et al. Rescaling quality of life values from discrete choice experiments for use as QALYs: a cautionary tale. Popul Health Metr 2008;6:6.

[25] International Society for Pharmacoeconomics and Outcomes Research. Mapping to estimate health state utility values from non-preference based outcomes measures for cost per QALY economic analysis - Good Practices Task Force 2014. Available from: http://www.ispor.org/taskforces/mapping-utility-values-from-non-pref-based-outcomes-grp.asp. [Accessed on December 30, 2015].

[26] Bijlenga D, Birnie E, Bonsel GJ. Feasibility, reliability, and validity of three health-state valuation methods using multiple-outcome vignettes on moderate-risk pregnancy at term. Value Health 2009;12:821–7.

[27] Johnson FR, Lancsar E, Marshall D, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. Value Health 2013;16:3–13.

[28] Sawtooth Software. Sawtooth Software. Orem: Sawtooth Software, 2009.

[29] Swait J, Louviere J. The role of the scale parameter in the estimation and comparison of multinomial logit models. J Mark Res 1993:305–14.

[30] Nickerson CAE. Assessing convergent validity of health-state utilities obtained using different scaling methods. Med Decis Making 1999;19:487–98.

[31] Gu NY, Botteman MF, Gerber RA, et al. Eliciting health state utilities for Dupuytren's contracture using a discrete choice experiment. Acta Orthop 2013;84:571–8.

[32] Parkin D, Devlin N. Is there a case for using visual analogue scale valuations in cost-utility analysis? Health Econ 2006;15:653–64.

[33] Arnesen T, Trommald M. Roughly right or precisely wrong? Systematic review of quality-of-life weights elicited with the time trade-off method. J Health Serv Res Policy 2004;9:43–50.

[34] Boye KS, Matza LS, Feeny DH, et al. Challenges to time trade-off utility assessment methods: when should you consider alternative approaches? Expert Rev Pharmacoecon Outcomes Res 2014;14:437–50.

[35] Attema AE, Edelaar-Peeters Y, Versteegh MM, et al. Time trade-off: one methodology, different methods. Eur J Health Econ 2013;14(Suppl): S53–64.

[36] Gray A, McNamara I, Aziz T, et al. Quality of life outcomes following surgical treatment of Parkinson's disease. Mov Disord 2002;17:68–75.

[37] Morimoto T, Shimbo T, Orav JE, et al. Impact of social functioning and vitality on preference for life in patients with Parkinson's disease. Mov Disord 2003;18:171–5.

[38] Palmer CS, Schmier JK, Snyder E, et al. Patient preferences and utilities for 'off-time' outcomes in the treatment of Parkinson's disease. Qual Life Res 2000;9:819–27.

[39] Torrance GW, Feeny D. Utilities and quality-adjusted life years. Int J Technol Assess Health Care 1989;5: 559–75.

[40] Torrance GW, Feeny D, Furlong W. Visual analog scales: do they have a role in the measurement of preferences for health states? Med Decis Making 2001;21:329–34.

[41] Cleemput I, Devriese S, Kohn L, et al. Incorporating Societal Preferences in Reimbursement Decisions—Relative Importance of Decision Criteria According to Belgian Citizens. Brussels: Belgian Health Care Knowledge Centre (KCE), 2014.

[42] Norman R, King MT, Clarke D, et al. Does mode of administration matter? Comparison of online and face-to-face administration of a time trade-off task. Qual Life Res 2010;19:499–508.