MASTER THESIS

Validity and Feasibility of Best Worst Scaling

Using Multiple Treatment Outcomes of Parkinson's Disease

University of Twente, School of Management and Governance

Department of Health Technology and Service Research

Author:

Tina Tin Wan, Leung

Supervisors:

Dr. Janine van Til

Dr. Karin C.G.M. Groothuis-Oudshoorn

MASTER GRADUATION THESIS

ENSCHEDE, AUGUST 2013

TINA TIN WAN, LEUNG

SCHOOL OF HEALTH SCIENCE

UNIVERSITY OF TWENTE

STUDENT NUMBER: S1353403

PRIMARY SUPERVISOR

DR. JANINE VAN TIL

ASSISTANT PROFESSOR

DEPARTMENT OF HEALTH TECHNOLOGY AND SERVICES RESEARCH

SECONDARY SUPERVISOR

DR. KARIN C.G.M. GROOTHUIS-OUDSHOORN

ASSISTANT PROFESSOR

DEPARTMENT OF HEALTH TECHNOLOGY AND SERVICES RESEARCH

POSTAL ADDRESS

P.O. BOX 217

7500 AE ENSCHEDE

ABSTRACT

BACKGROUND Best Worst Scaling (BWS) has been shown to be more superior to traditional valuation methods, including Time Trade Off (TTO) and Visual Analogue Scale (VAS). Additional information with higher discriminative power can be collected by asking respondents to choose the best and the worst option. However, BWS can only provide cardinal utility estimates and cannot be anchored on scale with definite range for cost-utility analysis.

OBJECTIVE To investigate the convergent validity, theoretical validity and feasibility of profile-based (Case 2) BWS and multi-profile-based (Case3) BWS. To visualize and test a BWS choice experiment for measuring treatment preferences of the public towards Parkinson's Disease (PD).

METHOD An online survey was conducted in June 2013 to measure treatment preferences of the public (N=592) using VAS, TTO, Case 2 and Case 3 BWS in the Netherlands and the United Kingdom. PD treatment consists of seven attributes: Treatment methods, extent of tremor, posture and balance problems, slowness in motion, dizziness, drowsiness and rapid uncontrolled movement. To assess the convergent validity, individual treatment preferences of participants were used to model the utility of six treatments using mixed logit regression based on choice data obtained through Case 2 and Case 3 BWS. The utilities these treatments measured through BWS were correlated to those obtained through VAS and TTO. To assess the theoretical validity, conditional logit regression was conducted to model the choice data. The feasibility of BWS was examined based on the quality of response, time required to perform the BWS tasks and the preference of participants to examine the practical issues of using BWS.

Result: The convergent validity of Case 2 and Case 3 BWS were satisfactory (TTO: Case 2 BWS=0.447; TTO: Case 3 BWS=0.438; VAS: Case 2 BWS: 0.593; VAS: Case 3 BWS=0.563; Case 2 BWS: Case 3 BWS= 0.830, p-value <0.00). Treatment methods, posture and balance problem, as well as rapid uncontrolled movement are relatively more important, whereas slowness in motion weighed relatively less. Completion rate is 71.7%. Mean time needed for Case 2 and Case 3 BWS are 5.0 minutes and 5.9 minutes respectively. They are equally preferred by participants (50%: 50%).

CONCLUSION Given the significant correlation with traditional valuation methods and the superiorities of BWS, it can be promisingly a standardized tool in evaluating treatments for chronic diseases in reimbursement decision making. Through demonstrating the use of BWS, this study provides insight about the differences between Case 2 and Case 3 BWS. Future research is needed to standardize the method for converting latent values of BWS onto the full Quality Adjusted Life Years (QALY) scale.

PREFACE

With this thesis I hope to broaden my knowledge about the scientific field of medical decision making. Parkinson's disease has been my research interest ever since I worked as a registered nurse in neurology and neurosurgery unit of hospital in Hong Kong, where I have experienced how the quality of life of patients and their beloved ones are being compromised.

My knowledge regarding health-related utility measurement was limited before working on this research. I have learnt the basic knowledge about health utility in the courses during this Master program. During my research, I realized that there are better approaches to quantify the values of medical treatments from the point of view of different important stakeholders. Through reviewing literatures, designing the choice experiment and analyzing data, I have the chance to learn statistical and computational techniques in measuring the value of medical treatments. This experience is mind broadening and interesting to me. I hope this study will contribute to the literature in improving the quality of life of patients with chronic diseases and increasing public awareness of Parkinson's disease.

CONTENT

1. INTRODUCTION

In view of the medical technology advancement in managing chronic diseases, evaluating the value of ever-growing medical treatments is of essence to make informed decision concerning allocating health care resources. The evaluation process should be transparent and reproducible, as well as being able to include the preferences of key stakeholders.

UTILITY VALUATION METHODS FOR MEASURING TREATMENT PREFERENCE

Utility refers to the desirability of an outcome or process to consumer (Feeny, 2000). Health utility is the cardinal measure of the preference for certain health outcome, which is measured through constructing a set of health states to be valued by individuals (Tolley, 2009). The utilities estimates generated based on the choice of individuals reflect their preferences. Treatment preference can be measured through evaluating the health utility in treatment using various approaches (Tolley, 2009).

Calculating the health utility of treatment outcome is crucial in the determination of the cost-effectiveness of treatments. To perform the cost-utility analysis of treatment, the cost involved are compared to the QALY, which is calculated by combining the perceived utility of HRQoL with the survival estimates in that health state (Atthobari, Bos, Boersma, de Jong-van dan Berg & Postma, 2005; Kind, Lafata, Matuszewski & Raisch, 2009). Traditional valuation methods, including VAS and TTO, involve measuring the utility of HRQoL associated with the treatment outcome by asking respondents to give numerical values towards the health states (Asadi-lari, Tamburini & Gray, 2004; Robinson & Spencer, 2006).

VISUAL ANALOGUE SCALE

In the VAS approach, respondents are asked to indicate their numerical value judgment towards the health state on a fractionated scale with endpoints of the best imaginable health state, valued as 100, and the worst imaginable health state which is valued as 0 (EuroQol Group, 2013). The utility of the health state can be calculated by ($VAS_j \div 100$), where VAS_j refers to the value assigned by respondents (Bennett & Blamey, 2001).

TIME TRADE OFF

On the other hand, the eliciting approach of TTO is different from VAS. The utility of health-related outcome is measured by asking respondents to choose between two given health states, either a duration of life years under certain health state or a shorter life years under perfect health (Cubo, 2010). Respondents can also choose the third option, which is that they cannot choose either one since they are equally desirable (Badia et al., 1999). The amount of time which a respondent is willing to trade for the perfect health state is the TTO value. It is calculated by $(1- X_j \div 10)$, where X_j equals to the number of years willing to be traded off for full health given the health related scenario (Prades & Miguez, 2011).

CONCERNS ASSOCIATED WITH USING TIME TRADE OFF AND VISUAL ANALOGUE SCALE

In spite of the fact that VAS and TTO are not cognitively demanding and the results produced are consistent without statistical variance, they are prone to numerical error since individuals may not have adequate experience in rating health states or different interpretations towards numbers (Kjær, 2005).

The most crucial disadvantage of the above two traditional approaches is that neither one of them mirror the actual decision making process (Flynn, 2010; Bennett & Blamey, 2001). Choosing treatments involves making trade-off between the advantages and disadvantages of treatments instead of rating the outcome of options or trading off life span (Bennett & Blamey, 2001). These approaches do not consider the process by which the end-result was obtained (Bennett & Blamey, 2001). Furthermore, both approaches do not have sufficient discriminative power in capturing the slight differences among the treatments for chronic disease (Arons & Krabbe, 2013).

Chronic diseases refer to the medical diseases, which are slowly progressive with no curative treatment (World Health Organization, 2013). The traditional assessment approach focuses only on the extent of reducing disease progression and the extent of relieving symptoms (Kaplan, 2003). Nevertheless, these criteria may not represent the overall value, because the treatment modalities and side effects can reduce the quality of life of patients significantly and outweigh the above stated effects (Montgomery & Fahey, 2001). Furthermore, treatment alternatives of chronic diseases are ever-growing with small incremental changes, in which these small changes cannot be captured well using traditional techniques. It is of the essence to explore the potential of other methods for measuring the utility of medical treatment in a more precise way.

IMPORTANCE OF TREATMENT PREFERENCES TOWARDS CHRONIC DISEASES

Patients are not the only important stakeholders in measuring treatment preferences,. The public contributes to the health care expenditure and their preference should also be valued (Essers et al., 2010). The similarity between Dutch and British health care policy is that the public contributes in providing health care services to patients. A compulsory insurance is involved in the provision of health service in the Netherlands (Essers, Dirsen & Prins, 2010). The public pay premium to insurance companies on regular basis and they can receive reimbursed health care service. On the other hand, health care service is provided for the public which is financed by tax in the United Kingdom (Grosios, Gahan & Burbidge, 2010). Besides, different medical treatments also lead to various health outcomes of patients which affect the functioning of society (Essers et al., 2010).

REGULATORY AND REIMBURSEMENT DECISION MAKING

Within the European Union (EU), European Medicines Agency (EMA) contributes in harmonizing the regulatory medical decision making by assessing the value of medicinal products (Szende, Leidy & Revicki, 2005). As stated by Szende et al. (2005), 38% of the disease specific regulatory document of EMA recommended the use of health related quality of life (HRQoL) and patients' related outcomes (PRO) as the primary or secondary clinical endpoints in investigating the value of the medicinal products of chronic diseases. In spite of the recommendation, there is no standard quantitative method with solid theoretical base for incorporating treatment preferences of stakeholders other than clinicians into medical decision making (Kaplan, 2003; Willke, Burke & Erickson, 2004).

As stated by Bouvy & Vogler (2013), in deciding reimbursement decisions about medical treatments, the eligibility of the treatments should be assessed and the legitimate reimbursable treatments should be examined to decide how much the public payer should support the cost. Based on the EU Transparency Directive (Council Directive 89/105/EEC), the national reimbursement system in EU Member States should follow three conditions related to reasonableness of the decision (European Commission, 1988). The included conditions are: (1) The reimbursement decisions should be publicly accessible; (2) Acceptable and relevant principles should be used to reflect the coverage decision; (3) Transparent procedure should be used to handle the challenging and disputing decisions (European Commission, 1988). In view of the reimbursement policy and the recommendation from EMA, the feasible techniques for assessing treatments for chronic disease should be able to measure and differentiate the holistic impact of the treatments and also with clear principles reflecting the decision making process.

BEST WORST SCALING

One technique that has received much interest in recent literature is Best Worst Scaling (BWS). It is an innovative Discrete Choice Experiment (DCE) based on the inherited comparative judgments when humans face choices (Louviere, Hensher & Swait, 2000). Traditional DCE includes a series of choice questions and respondents have to select their preferred alternative (Arons & Krabbe, 2013). The choice data is then aggregated to estimate the overall choice model (Arons & Krabbe, 2013). The main shortcoming of DCE is the need of large amount of choice questions to establish the choice model (Flynn, Louviere, Peters & Coast, 2007).

BWS overcomes this disadvantage through requiring participants to select the best and the worst option in choice questions (Marley and Louviere, 2005). With this preference eliciting technique, BWS exploits the use of choice questions and possesses higher discriminating power because choice data obtained can be expanded into best-worst pair (Flynn, 2010). This also contributes to more effective sample size since more data are probed from one individual in each question (Lancsar, Louviere & Flynn, 2013).

The advantage of BWS is not limited to the above aspects, the relative importance of attribute levels of medical treatments can be compared based on random utility theory, which assumes there is latent underlying utility scale in human when facing choices (Marley and Louviere, 2005). The utility scale is formulated as the sum of systematic components explained in the description of choice alternative, and a random component, representing all possible unexplained factors leading to variation in utility estimates (Ryan, Amaya-amaya & Gerard, 2008). As stated by Marley & Louviere (2005), it is easier to identify extreme options than to rate them. BWS utilizes this inherited human propensity to uncover the underlying utility scale of respondents. It sufficiently reflects the cognitive process of individual when facing choices, which is the identification of every available pairs of attribute's levels and then selects the pair with maximum difference (Marley and Louviere, 2005). The utility estimates are also free of cognitive bias towards numbers (Lancsar, Louviere, Donaldson, Currie & Burgess, 2013).

TYPES OF BEST WORST SCALING

There are three types of BWS, including object-based, profile-based and multi-profile-based BWS. (Case 1, 2 and 3 BWS respectively). Case 2 and 3 BWS can be used to estimate the overall utility of treatment since the overall impact of key attributes are included in the choice questions, whereas Case 1 can only estimate part of the impacts of the treatment without the complete profile (Flynn, 2010).

1. Profile-based best worst scaling

In the Case 2 BWS choice task, a single and complete profile with varying attribute levels is displayed in each choice question. Respondents are asked to select the best and the worst attribute levels within the profile (Flynn et al., 2007). It is the most common type of BWS used in health economics studies since it possesses two advantages over the other two types (Lancsar et al., 2007). Firstly, the result obtained is more comprehensive than traditional DCE since all the impacts of included attribute levels associated with treatment can then be compared directly on a common scale (McIntosh & Louviere, 2002).The second advantage is that it is less cognitive burdening to select the best and the worst item within a profile than making choices among profiles (Lancsar et al., 2007). One potential limitation of Case 2 BWS is that it does not resemble the actual decision making process because it does not involve a choice between different alternatives (Lancsar et al., 2013).

2. Multi-profile based best worst scaling

On the other hand, Case 3 BWS is more similar to traditional DCE in which the only different point is the requirement of selecting the worst profile in each question among profiles (Flynn, 2010). It is increasingly popular to be used in health economics studies since it mirrors the selection of treatment in the reality and additional information can be yielded (Flynn, 2010). Yet, there are two disadvantages associated with Case 3 BWS, making it less popular than Case 2 BWS. Firstly, unlike Case 2 BWS, there are more than one utility scales in each attribute in the result, thus the impact of attribute levels cannot be compared on the same scale simultaneously (Flynn & Marley, 2012). Secondly, it is more cognitively demanding than Case 2 undoubtedly since respondents need to consider multi-profiles in each choice question (Flynn, 2010). Given the similarities and differences between Case 2 and Case 3 BWS, it is desirable to investigate both methodologies.

STATISITICAL FRAMEWORK OF BEST WORST SCALING

Choice frequency is used as the metric to generate the probability of each items to be selected as the best or the worst (Flynn & Marley, 2012). The underlying statistical theory is that the relative choice probability of a given best-worst pair of options is proportional to the distance between the two options on the latent utility scale (Marley & Louviere, 2005). Based on the random utility theory, the statistical function of BWS is $U_{ij}= V_{ij+} \varepsilon_{ij}$, where U_{ij} represents the utility of a health state of an individual i in choice set j; V_{ij} represents the explainable components included in the choice experiment. ε_{ij} represents the

random components due to variability among individuals and random error within individual (Flynn et al., 2008). It is assumed that the relative choice probability of a given best worst pair is represented by the proportional distance between them on a latent utility scale (Flynn et al., 2007). Another assumption is that individual has an underlying utility function which is linearly additive (Flynn et al., 2007). The value of hypothetical treatment is captured in $V_{ij} = X_{ij} \beta + Z_i \gamma$, where X_{ij} is the vector of attributes of the choice set j as viewed by the individual i, whereas Z_i is the vector of individual characteristics of individual i. Both β and γ represent the coefficients of vectors which need to be estimated.

With the use of systematic variation to include treatments with competing characteristics into choice questions, preference of respondents can be elicited effectively without bias. Each of the best-worst pair selected by respondents is conceptualized as one maximum difference between the two options on an underlying utility scale (Flynn et al., 2007). By aggregating individual best-worst pairs by series of choice questions, the difference in utility between every best-worst pair can be estimated. The choice models of individuals can be estimated and individual choice models can be pooled into an average choice model which represents the population (Flynn et al., 2007).

CHALLENGES IN USING BEST WORST SCALING IN EVALUATING MEDICAL TREATMENTS

The absence of golden standard makes it difficult to select which techniques to measure health utility, external parameters are thus necessary to compare different health-related utility valuation methods (Bijlenga, Birnie & Bonsel, 2009). Based on the theory behind BWS, the utility of treatment process can be captured more systematically and realistically than VAS and TTO (See Appendix I). However, it cannot be used as a standardized instrument for regulatory and reimbursement decision making at this stage, because it only provides cardinal utilities based on the ordinal preferences of treatment outcomes on a latent utility scale, thus a scale-based value ranging from 0 to 1 cannot be derived (Flynn & Marley, 2012). As a consequence, the utility estimates generated through BWS cannot be formulized within the decision framework for constructing QALYs to measure the cost-utility of medical treatments.

On the other hand, VAS and TTO are capable of anchoring the health utilities on the scale of 0 to 1 (Craig, Busschbach & Salomon, 2009; Badia, Monserrat, Roset & Herdman, 1999). Direct comparison between BWS and traditional approaches, as well as between Case 2 and Case 3 BWS are beneficial to the methodological development regarding the health-related valuation technique (See Appendix II).

CASE STUDY

Parkinson's disease (PD) is selected for investigation since it is the second most prevalent neurodegenerative disease after Alzheimer's disease in Europe and it is challenging for clinicians and patients to decide the optimal treatment (Jankovic, 2008). There are more than 1.2 million European having PD (Di Luca et al., 2011). It is also estimated that the incidence of PD will be doubled by 2030 (Dorsey et al., 2007).

PD is caused by the neurodegeneration of dopaminergic pathway in the pars compacta region of the substantia nigra inside the brain, leading to progressive depletion of dopamine in the central nervous system and then failure in nerve impulse transmission (Chaudhuri, Healy & Schapira, 2006; Wirdefeldt, Adami, Cole, Trichopoulos & Mandel, 2011). As a consequence, a wide range of motor and non-motor symptoms are resulted (Wirdefeldt et al., 2011). Common motor symptoms include resting tremor, rigidity and bradykinesia (Wirdefeldt et al., 2011). Besides, mental, cognitive and psychological symptoms are commonly seen; including insomnia, depression and PD initiated dementia (Jankovic, 2008). At later stage, the self-care ability will be seriously compromised (Jankovic, 2008).

Since the root cause of PD is unsolved, current treatment plan aims at providing symptomatic relief against motor symptoms and lengthening the effectiveness period of oral medication (Chaudhuri, Healy & Schapira, 2006). Nevertheless, oral medication is only efficacious at early stage and its effect will be worn off as PD progresses for four to six years (Chaudhuri, Healy & Schapira, 2006). This is due to the extensive loss of dopaminergic neurons (body cell which works by receiving dopamine) which cannot be compensated through oral medication (Ferreira & Rascol, 2000). It is noteworthy that there are alternative treatments for PD. Table 1 provides clear comparison of the alternative approaches.

TABLE 1. ADVANTAGES AND DISADVANGES OF COMMON PD TREATMENTS

	Oral Medication	Duodenum Pump	Deep Brain Stimulation
Main Advantages	• Good motor benefits at early stage • Non-invasive	• Significant motor benefits through continuous and steady release of medication to brain. • No oral medication is required • Less side effects and better symptomatic relief	• Significant motor benefits through sending electrical impulse to the brain • Drug dosage and frequency can be reduced
Main Disadvantages	• Numerous side effects including dyskinesia, postural hypotension and drowsiness. • Patients have to take tablets frequently • Not effective for advanced PD	• Minor surgical procedure is required to place a permanent tube to the duodenum. • Secondary effects include sporadic blockage, displacement of inner tube, leakage and local infections. • The need to change parts of the pump system and cleanse tubing from time to time	• Most invasive since brain surgery is required to place a neurostimulator under the chest skin and insert an electrode inside the brain • Associated with behavioral changes including committing suicide • The need to change battery of machines after 3 to 6 years depending on specific programming settings
	Nilsson, Nyholm & Aquilonius, 2001; De Gaspari et al., 2006		

** Dyskinesia is due to the intermittent fluctuated dopamine level outside the brain and the period of time that patient is suffering from this symptoms is called off-time (Palmer. et al., 2000). It is characterized by rapid uncontrolled movement at limbs, face or trunk (Ferreira & Rascol, 2000; Chaudhuri & Odin, 2006).*

Depending on the view of patients, the treatment burden and side effects may outweigh the clinical effects. The non-compliance rate of PD treatment can be as high as 50% (Leopold, Polansky, Hurka, 2004; Arbouw et al., 2008). Common causes include the associated side effects and treatment burden (Palmer, Schmier, Snyder & Scott, 2000; Sujith & Lane, 2009).

Studies about effectiveness of PD treatment mainly focus on how the treatment can reduce motor fluctuation and extent of off-time (Rascol et al., 2000; Palmer et al., 2000; Siderowf, Holloway & Stern,

2000). These criteria may be more influential for clinicians in deciding PD treatment; however the value of PD treatment goes beyond traditional measures from the view of patients (Dodel, Reese, Balzer & Oertel, 2008). Understanding the heterogeneous treatment preferences of clinicians, patients and the public is of uttermost importance in making unbiased reimbursement decision (Montgomery & Fahey, 2001). Clinicians possess more medical knowledge, whereas patients are the ones who experience the treatment. Besides, the medical expenditure of PD patients is financially supported by the public. Unequal emphasis of the treatment preference among stakeholders will lead to negative outcomes. Solely adopting preference from patients may lead to excessive demand for unnecessary, costly or even harmful treatment because they may not have adequate medical knowledge or their judgement is clouded by emotional, mental and cognitive dysfunction (Lees, Hardy & Revesz, 2009). Solely adopting preference from clinicians is also not justified because it may lead to reduced quality of life of patients and higher non-compliance rate of treatment (Sculpher, Gafni & Watt, 2002). Neglecting the preference of key stakeholders may lead to unwise allocation of health care without societal consensus (Coulter, 1997). Annually, approximately €13.9 billion is spent on PD patients in Europe and this expenditure will continue to rise due to aging population (European Parkinson's Disease Association, 2011). It is important to study how this expenditure can help patients, instead of being an economic burden to the society.

By incorporating key treatment attributes in the choice experiment to assess PD treatments, more comprehensive evaluation can be performed (Arons & Krabbe, 2013). Through combining with the value judgements of stakeholders, social benefits can then be balanced and decision makers can make an informed decision with societal consensus.

RESEARCH GOAL

There is the lack of empirical research related to head-to-head comparison between BWS and traditional valuation methods. Given the need for instrument to capture the small but relevant differences in the utility of treatment based on treatment process and the perceived advantages of BWS, it is uncertain whether BWS is capable of facilitating the reimbursement decision. This study aimed at filling this research gap through assessing the correlation between BWS approaches and traditional approaches in the context of PD. Besides, this study also examines the other aspects of BWS including the theoretical validity and feasibility.

RESEARCH QUESTION

The main research question is stated as follow:

How does BWS correlate with traditional health-related utility measuring approaches?

This question was addressed by assessing the convergent validity of BWS by performing mixed logit regression analysis to generate utilities values of six treatment scenarios estimated by TTO, VAS, Case 2 and Case 3 BWS. They were undergone correlation test and the statistical significance was examined.

Two sub-questions are set which are related to assessment of theoretical validity and feasibility of BWS:

1. With the use of Case 2 and Case 3 BWS, are the impacts of treatment attributes estimated consistent with prior expectation?

2. Is BWS a feasible instrument in measuring treatment preferences from the perspective of the public?

SIGNIFICANCE OF THE STUDY

From a scientific perspective, this study provides evidential support regarding the use of BWS in reimbursement decision making. Examining the utility scale and exploring the use of BWS in assessing medical treatments for chronic diseases is beneficial in making objective decision towards health care resources allocation. As a consequence, the outcome of reimbursement decision can be maximized with the consensus from key stakeholders.

From a societal perspective, the public awareness of social participation in assessing health technology can be increased. More public can realize the importance in balancing the pros and cons of medical treatments and can also recognize their role in improving the outcome of reimbursement decision.

2. METHODOLOGY

In this study, Case 2 and Case 3 BWS are compared with traditional measures, VAS and TTO. An online survey in Dutch and English was prepared for data collection and participants were recruited from the public in the Netherlands and the United Kingdom. The concept of examining the convergent validity, theoretical validity and feasibility are described first, followed by explaining the process of constructing the choice experiment, including the identification of key treatment attributes the construction of choice questions, the requirement of sample size and the components of the survey.

CONVERGENT VALIDTY

Convergent validity is a subtype of construct validity for assessing whether the measurement of an instrument is related to another instrument with the same purpose (Kjær, 2005). As stated by Trochim (2006), the measurements of similar constructs should be related to each other. As BWS, VAS and TTO measure the same construct, it is important to assess whether the utility estimates obtained through Case 2 and Case 3 BWS can be converged to the utility scale of TTO and VAS (Trochim, 2006). In other words, this study aims at showing the convergence among the BWS and traditional valuation approaches. To establish the convergence between similar constructs, correlation coefficient (Also named as Pearson correlation) is the typical approach to assess the intercorrelation among instruments (Trochim, 2006).

THEORETICAL VALIDTY

Theoretical validity focuses on assessing the theoretical foundation of the choice experiment by examining whether the result meets the prior expectation (Kjær, 2005). To the best knowledge of the author, there is no study investigating the treatment preferences of the public, thus no prior knowledge about the treatment preferences. Thus, the prior expectation are made based on the ordinal variables; the coefficient of relatively worse attribute levels (Often suffering from symptoms and side effects) would be more negative than relatively better attribute levels (Seldom suffering from symptoms and side effects).

FEASIBILITY

Feasibility of utility eliciting technique is an important factor to determine the choice of valuation method. It refers to the capability of an instrument to achieve the aim and the acceptability of respondents (Dolan, Gudex, Kind & Williams, 1996). This can be assessed by completion rates, time needed and quality of response (Brazier & Deverill, 1999; Canaway & Frew, 2013).

IDENTIFICATION OF KEY TREATMENT ATTRIBUTES

Extensive literature review was conducted to select key treatment attributes into the choice experiment. Additional information was also obtained through patients' interviews and consulting neurologists conducted by research group to include clinically relevant, operational, mutually independent attributes. Seven treatment attributes with three levels were identified in regard to treatment methods, symptomatic control and side effects (Table 2).

TABLE 2. TREATMENT ATTRIBUTE AND ASSOCIATED LEVELS INCLUDED IN THE STUDY

	Attributes and definition	Associated types or levels
Treatment methods	• Treatment methods: Ways to receive treatments.	• Oral tablets • Pump • Brain surgery
Symptomatic control	• Resting tremor: Involuntary oscillations in the limbs during rest, increased stress, emotion or fatigue. • Posture and balance problems: Difficulty in maintaining body balance and posture due to lack of control, pain and tightness in muscles, leading to high risk of fall. • Slowness in motion: Delay in initiating movements.	• Seldom to never • Sometimes • Often
Side effects	• Dizziness: Having a sensation of whirling upon postural change which is due to sudden decrease in blood pressure. • Drowsiness: Excessive sleepiness in daytime. • Rapid uncontrolled movement: Shivering of limbs, face or trunk which is resulted from long term medication.	
	Senard et al., 1997; Jankovic, 2008; Sujith & Lane, 2009; Wirdefeldt et al., 2011	

EXPERIMENTAL DESIGN

In order to measure the treatment preference quantitatively, a precise choice experiment was designed with the use of computer software. As mentioned previously, choice modeling in BWS is achieved through aggregating the selected best-worst pairs in the choice experiment. It is important to display attributes with the same frequency, so that equal focus is placed to all attribute levels. As a result, each one of them has equal probability to be chosen in the choice questions (Flynn, 2010).

The number of unique best-worst pair to be chosen in a choice experiment can be calculated by 2[(K-1) + (K-2) + (K-3) +...+2 + 1], where K refers to the number of attributes in the profile (Flynn, 2010). It is necessary to include sufficient number of scenarios for choice modeling, whilst not overburdening the participants. The number of possible scenarios was levels ^{attributes}, which is 2187 (3^7) in this study (Flynn, 2010). Given this large amount of hypothetical scenarios which may overburden respondents, orthogonal main effects plan was used to include a subset of treatments from the original full factorial design, yet, the statistical properties of the design are maintained (Louviere, Hensher & Swait, 2000; Marshall, et al., 2010).

D-efficiency, the most common optimality criterion for DCE, is used as the principle in designing the choice experiment through an optimization of the level balance and orthogonality, thus maximizing the statistical efficiency (Huber & Zwerina, 1996; Carlsson & Martinsson, 2003; Johnson et al., 2013). Level balance refers to equal frequency of each level of attributes in the design (Huber & Zwerina, 1996). Orthogonality refers to equalizing the frequency of co-existence of any two different attribute levels in choice sets to the product of their individual marginal frequency. In such way, the different treatment scenarios with different levels vary independently with the design in a criss-cross manner and the main effects in the model can be estimated independently (Kuhfeld, 2003). Statistical software was used to maximize the design of choice experiment in this study. By using statistical software, the occurrence and co-occurrences of attribute levels can be well balanced (Flynn & Marley, 2012). The design is optimal in a region around the perfect solution which can estimate all the main effects within the β space and can extract maximum amount of choice data with the use of statistical software (Street & Burgess, 2007).

The Case 2 and Case 3 BWS choice questions contained the same attribute levels. In Case 2 choice tasks, all seven attributes were presented in a single treatment where as that of Case 3 were presented in three treatments. Four versions of choice sets in both Case 2 and Case 3 BWS tasks were designed to avoid version effect (Louviere et al., 2008). Participants were randomly assigned to perform one of the four versions in both Case 2 and Case 3 BWS choice tasks. Furthermore, the sequence of Case 2 and Case 3 BWS choice tasks was randomized to prevent potential ordering effects (Farrar & Ryan, 1999; Scott & Vick, 1999).

SAMPLE SIZE

There is no guideline concerning the size of optimal sample for BWS study. To obtain a statistically significant result, it is recommended to include a sample size of at least 300 participants (Orme, 2006). Given the possibilities of receiving incomplete response, a target of 500 respondents was set. There is no exclusion criterion for recruitment.

COMPONENTS OF ONLINE SURVEY

The online survey was composed of different sections which began with the introduction of the importance of their view towards PD treatment and questions related to their sociodemographic background. It was followed by a short and precise explanation of the main characteristics of PD treatments and an instruction of how to perform choice tasks with examples. After the choice tasks, questions about their preference and feedback were included. Main parts are highlighted.

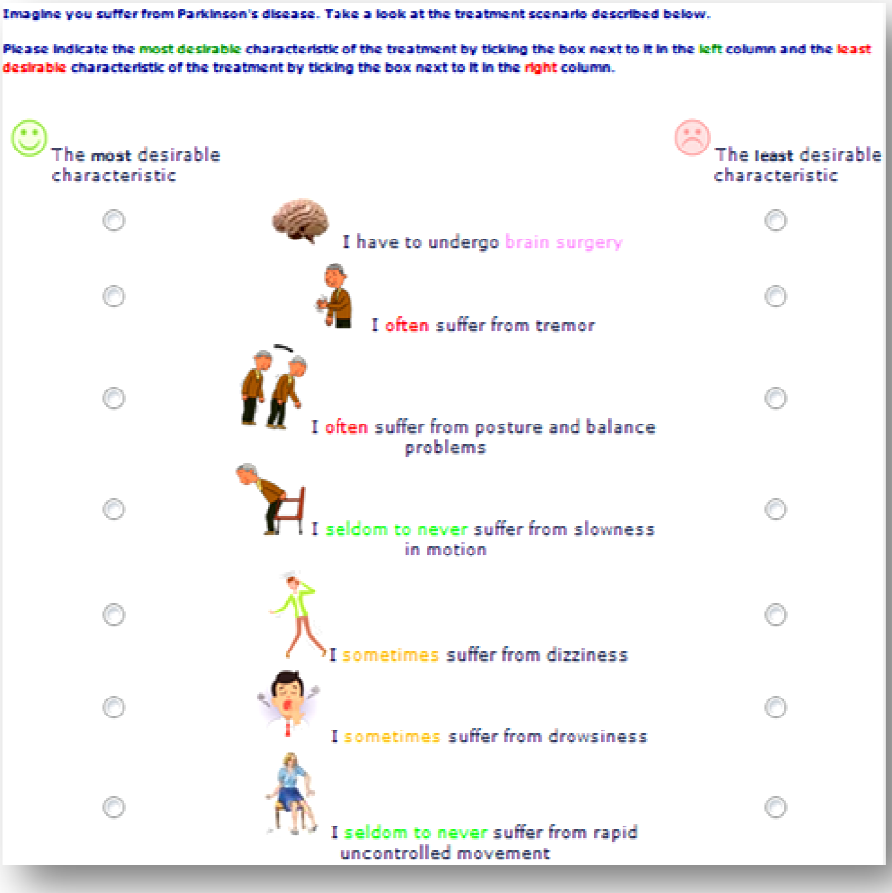
QUESTIONS RELATED TO SOCIODEMOGRAPHIC DATA

Respondents were asked to give information about their age, education level, nationality and any experience related to PD. Their health states were also measured by using European Quality of Life-5 Dimensions (EQ-5D) to have a simple and thorough description about the health-related quality of life of participants. Five dimensions of health were assessed, including mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension consisted of five levels ('No problems', 'Slight problems', 'Moderate problems', 'Severe problems' and 'Unable to') (EuroQol Group, 2013). An index value reflecting HRQoL can be calculated based on the result. It is likely that PD patients may be recruited in the study, thus Quality of Life Questionnaire-8 (PDQ-8), a specific instrument for measuring HRQoL of PD patients, was also included. Five levels (‘Never’, ‘Occasionally’, ‘Sometimes’, ‘Often’ and ‘Cannot do at all’) are included to assess eight dimensions of health for patients, including ‘Difficulty in getting around in public’, ‘Difficulty in dressing’, ‘Depressed feeling’, ‘Problems with close personal relationships’, ‘Problems with concentration’, ‘Feeling of unable to communicate with people properly’ and ‘Painful muscle cramps or spasms’. Similarly, an index value reflecting the HRQoL of PD patients can be calculated (Jenkinson et al., 1997).

CHOICE TASK USING PROFILE-BASED BEST WORST SCALING

Respondents were asked to evaluate ten profiles of treatments by selecting the best and the worst attribute levels. Each choice represented a pair of best-worst attribute levels. The sample choice task is shown in Figure 1.

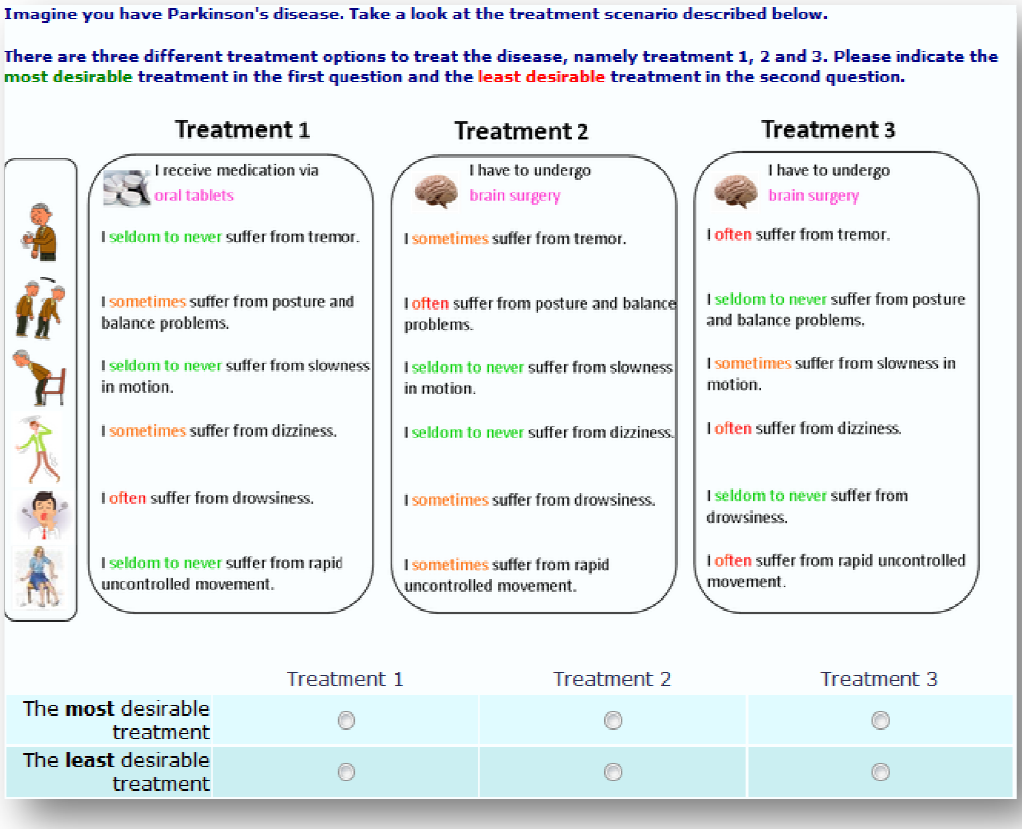
FIGURE 1. SAMPLE CHOICE TASK USING CASE 2 BWS



CHOICE TASK USING MULTI-PROFILE-BASED BEST WORST SCALING

Respondents were asked to evaluate three treatment options in each of the eleven choice sets by selecting the best and the worst treatment options. Each choice represented a pair of best-worst treatments. The sample choice task is shown in Figure 2.

FIGURE 2. SAMPLE CHOICE TASK USING CASE 2 BWS



CHOICE TASKS USING TIME TRADE OFF AND VISUAL ANALOGUE SCALE

After using BWS to value treatments, participants were asked to put themselves into imaginary position of having PD and value four treatments using VAS and TTO. In order to enable respondent to familiarize with these techniques to obtain more reliable utilities estimates, they had to use VAS and TTO to evaluate their current health first. Among the included treatments, one of them consisted of all the lowest level of ordinal variables of treatment attributes and another one consisted of all the highest level of ordinal variables of treatment attributes. The other two treatment options were composed of

varying combinations of attribute levels in which it is difficult to prioritize their values. To avoid version effects, participants were randomized to perform two out of the four intermediate treatments (Louviere et al., 2008).

PREFERENCE BETWEEN PROFILE-BASED AND MULTI-PROFILE-BASED BEST WORST SCALING

The last part included question asking participants' preference between Case 2 and Case 3 BWS in terms of the ease of use. An optional open-ended question was also included to collect feedback regarding the whole survey.

4. DATA ANALYSIS

The primary objective of data analysis is to examine the convergent validity of BWS and the secondary objectives include examining the theoretical validity and the feasibility of BWS.

In order to improve the quality of choice data for data analysis, returned questionnaires were screened to exclude data from participants who did not complete the survey and/or consistently selected two options solely in Case 3 BWS. Given the length of the survey, it is also estimated that it is impossible to be completed within less than five minutes unless full attention was not given by respondents. Thus the returned survey from participants which were completed within five minutes would be eliminated from data analysis. Further screening was not conducted to preserve potentially valuable data.

ASSESSING THE CONVERGENT VALIDITY

Assessing the convergent validity is the main objective of this study. The null hypothesis was tested against the alternative hypothesis as follow:

***Null Hypothesis**: The utility estimates of treatments measured by Case 2 and Case 3 BWS are not correlated with those measured by VAS and TTO.*

***Alternative Hypothesis**: The utility estimates of treatments measured by Case 2 and Case 3 BWS are correlated with those measured by VAS and TTO.*

Convergent validity among VAS, TTO, Case 2 and Case 3 BWS was determined by using linear correlation with 95% confidence interval. Mixed logit regression model was applied to the choice data obtained through Case 2 and Case 3 BWS to examine the individual preference model of each respondent. This is

a mixed model which combines the features of standard conditional logit regression model, as well as incorporating individual preferences in the logit model (Agresti, 2002). Through applying this model, two sets of coefficients of attribute levels were computed. They represented the impact of each attribute level to the overall utility of treatment measured by Case 2 and Case 3 BWS respectively. Based on these two sets of co-efficient associated with attribute levels, the utilities of the six treatments at mean level and at individual level (which were also measured through VAS and TTO) were computed.

Afterwards, the utilities of the six treatments measured using VAS, TTO, Case 2 and Case 3 BWS were compared to examine the strength of linear dependence. Pearson correlation coefficient is used since the utility scales of TTO and VAS is from 0 to 1, whereas that of Case 2 and Case 3 BWS is an underlying latent utility scale without fixed boundary. As stated by Portney and Watkins (2000), correlation coefficient between 0.25 to 0.50 is considered as having fair relationship; those with 0.5 to 0.75 indicates moderate to good relationship; those above 0.75 is deemed as having good to excellent relationship.

Correlation analysis was also performed through constructing bivariate plot of the average utilities of the six treatments between every pairs of valuation methods. The change in utility of the six treatments between each pairs of valuation method s can be examined. Scatterplot of the individual utility estimates of the six treatments between every pairs of valuation methods was also constructed since the utility estimates through different valuation methods can be examined at individual level.

ASSESSING THE THEORETICAL VALIDITY

To answer the sub question (1) concerning the theoretical validity of BWS, Conditional logit regression model is used to estimate the main effects of treatment preference of participants for assessing the theoretical validity. Paired model conditional logit regression was used to estimate the choice data obtained through Case 2 and Case 3 BWS. Each unique best-worst pair was treated as the possible choice outcome and selected best-worst pair was modeled as the outcome with maximum difference within each choice question (Flynn et al., 2008). It is also important to use alternative approaches to compare whether the results obtained are consistent. Frequency count of attribute levels was calculated to obtain the best-minus-worst scores for each attribute levels. This approach has linear relationship with conditional logit regression model and thus a suitable approach to estimate the choice model (Flynn & Marley, 2012).

The dependent variables are the best worst pairs and the independent variables are the 21 attribute levels. Dummy coding is used to translate the categorical choice data into quantitative variables for conducting regression analysis in both Case 2 and Case 3 BWS (Alkaharusi, 2012). With the use of conditional logit regression model and dummy coding, the utility differences within every pairs can be modeled and the impact of attribute levels can be presented as utility coefficients (Flynn & Marley, 2012). In Case 2 BWS, the selected best worst pair of attribute levels were coded as 1 while the remaining possible pairs in each choice question were coded as 0. In Case 3 BWS, the attribute levels of the selected best worst pair of treatment were coded as 1 whereas the remaining attribute levels of the remaining treatments were coded as 0. By setting one attribute level as the reference variable to estimate the conditional regression model, the overall result indicates whether there is a relationship between dummy variables and the dependent variables. The reference variable is preferably the least important attribute level (Flynn et al., 2008). Since it is uncertain which attribute level is the least important before study, preliminary regression is performed first to identify the least important attribute. This attribute level is then set to be the reference in the final regression model.

The coefficients of attribute levels reflect their relative impacts (Flynn & Marley, 2012). There are slight differences between the utility coefficients of Case 2 and Case 3 BWS. In Case 2 BWS, all the utility coefficients of attribute levels lie on the same difference scale based on reference attribute level. In contrast, the utility coefficients of attribute levels of Case 3 BWS are anchored within the scale of each attribute with one level as the reference point (Flynn & Marley, 2012).

ASSESSING THE FEASIBILITY

To answer the sub question (2) related to feasibility, the completion rates, the proportion of participants who passed the data screening test and the length of time needed for completing the Case 2 and Case 3 BWS choice tasks were analyzed using descriptive statistics. Besides, participants were asked to express their preference between Case 2 and Case 3 BWS based on ease of use. The proportion of participants choosing Case 2 and Case 3 BWS was calculated. The optional feedbacks of participants towards the whole survey were also recorded.

Descriptive and correlation analysis was conducted using SPSS 20 and regression analysis was conducted using Stata 12.0. A p-value of ≤ 0.05 was considered as statistically significant.

4. RESULTS

With the use of online survey software platform, 851 individuals accessed to the online questionnaire. Completion rate is 71.7% (n=610). Choice data from 15 individuals were eliminated since they spent less than five minutes to finish the questionnaire. There were four individuals consistently selected the same treatment options in Case 3 BWS choice task. One of them had already been eliminated due to the time spent on finishing the survey. Thus 18 individuals were eliminated from the data analysis in total. Data analysis is pertained within the 592 individuals who passed all the three screening tests.

SOCIODEMOGRAPHIC CHARACTERISTICS AND HEALTH STATE OF PARTICIPANTS

The sociodemographic background of participants was measured. The sample population consists of 277 individuals from the Netherlands and 315 from the United Kingdom. There were 267 male and 325 female respondents. The average age of the respondents was 42 years old (SD=14). The self-reported sociodemographic information of Dutch and British population and the differences between them are shown in Table 3. There are significant differences between the Dutch and British population in terms of their education level and their gender.

TABLE 3. SOCIODEMGRAPHIC CHARACTERISTICS OF PARTICIPANTS

Count (Percentage)		NL	UK	Total	χ^2 (p value)
Number of participants		476	375	851	
Number of participants completed the survey		284 (47%)	326 (53%)	610 (97.1%)	
Number of participants fulfilled the screening		277 (74%)	315 (66%)	592 (70%)	
Gender	Male	100 (36%)	167 (53%)	267 (45%)	17.80*
	Female	177 (64%)	148 (47%)	325 (55%)	(0.000)
Age	≤30	77 (28%)	82 (26%)	159 (27%)	3.048
	30-60	152 (55%)	192 (61%)	343 (58%)	(0.225)
	≥60	48 (17%)	41 (13%)	89 (15%)	
Education level	Low	49 (18%)	2(1%)	51 (9%)	55.450*
	Medium	121 (44%)	181 (57%)	302 (51%)	(0.000)
	High	104 (37%)	124 (39%)	227 (38%)	
	Other	3 (1%)	8 (3%)	11 (2%)	

HEALTH CONDITION AND PD RELATED EXPERIENCE OF PARTICIPANTS

The health state and PD related experiences of participants were measured (Table 4). The mean score of EQ5D of participants is 0.814. Their health states did not vary significantly between the Dutch and British population. There were only 9 participants who had PD and 135 participants who had friends or relatives with PD. Majority of the participants did not have PD related experience.

TABLE 4. HEALTH STATE AND PD RELATED EXPERIENCE OF PARTICIPANTS

Count (Percentage)		NL	UK	Total	χ^2 (p value)
Health state	EQ5D <0.5	18 (6%)	32 (10%)	50 (8%)	2.476
	EQ5D index value ≥0.5	259 (93%)	283 (90%)	541 (92%)	(0.116)
	EQ5D Score	0.848 (SD=0.189)	0.786 (SD=0.234)	0.814 (SD=0.217)	
PD	Having PD	4 (1%)	5 (2%)	9 (2%)	0.021
	Not having PD	273 (99%)	310 (98%)	582 (98%)	(0.887)
Experience about PD	Having relatives/ friends with PD	57 (21%)	78 (25%)	135 (23%)	1.512
	No relatives/ friends with PD	220 (79%)	235 (75%)	454 (77%)	(0.195)

CORRELATION AMONG VAS, TTO, CASE 2 AND CASE 3 BWS

The convergent validity of Case 2 and Case 3 BWS were measured. The average utilities of the six different treatment scenarios measured by the four approaches were computed (Table 7).

TABLE 7. UTILITIES ESTIMATES OF THE SIX TREATMENTS THROUGH VAS, TTO, CASE 2 AND CASE 3 BWS

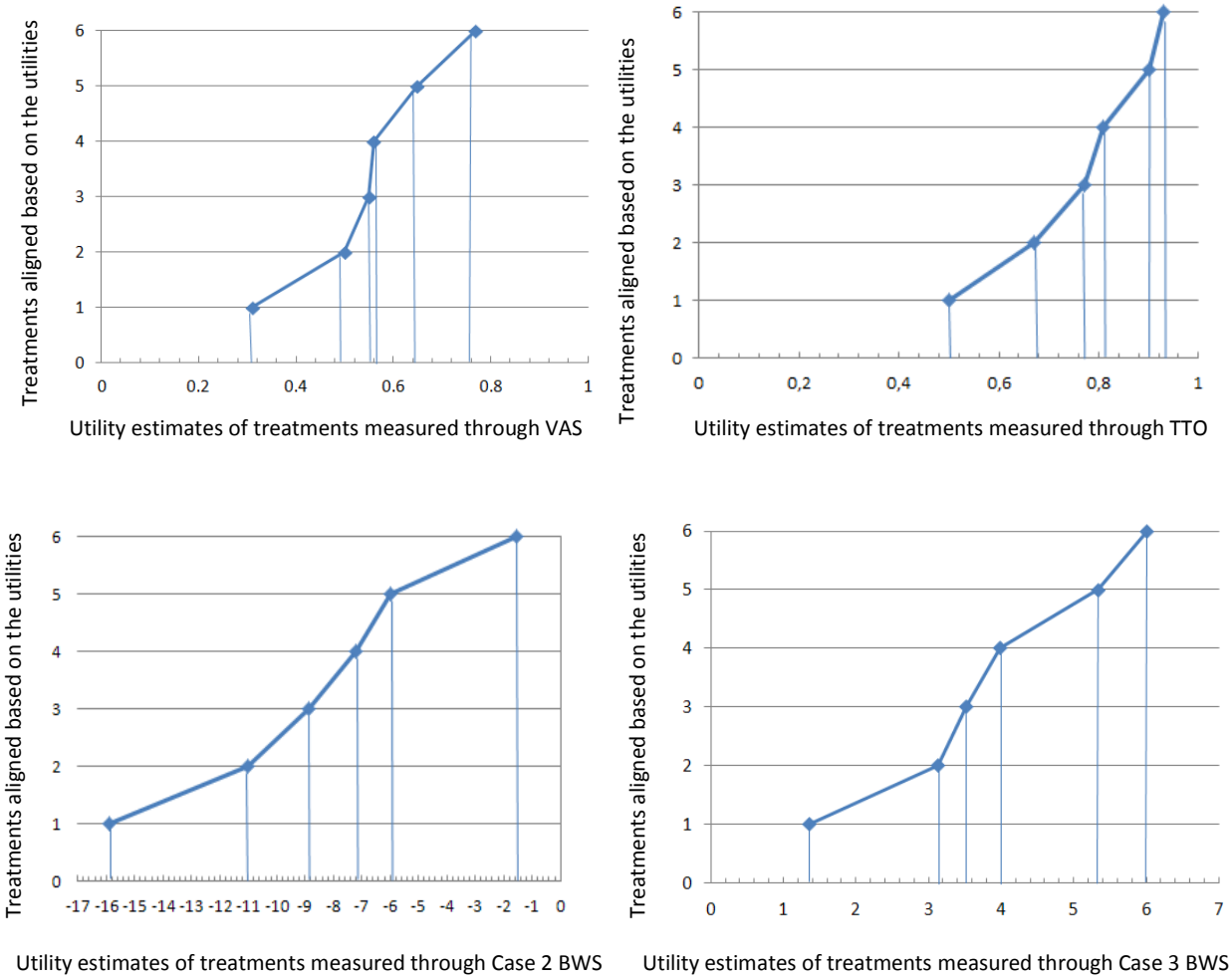
Treatments	Utility estimated by VAS (S.D)	Utility estimated by TTO (S.D)	Utility estimated by Case2 BWS (S.D)	Utility estimated by Case3 BWS (S.D)
Treatment with all 'often' levels	0.314 (0.265)	0.504 (0.364)	-15.878 (1.904)	1.362 (0.949)
Intermediate Treatment 1	0.551 (0.180)	0.767 (0.313)	-8.867 (2.020)	3.529(0.941)
Intermediate Treatment 2	0.652 (0.153)	0.907 (0.220)	-6.026 (1.765)	5.345 (0.909)
Intermediate Treatment 3	0.508 (0.193)	0.671 (0.351)	-11.03 (1.909)	3.138 (0.923)
Intermediate Treatment 4	0.560 (0.178)	0.808 (0.300)	-7.241 (1.881)	3.986 (0.959)
Treatment with all 'seldom' levels	0.768 (0.184)	0.929 (0.194)	-1.523 (1.934)	6.008 (0.996)

On average, the sequences of the six treatments based on their utility values are the same among the four approaches (Treatment with all 'often' levels <Intermediate 3 < Intermediate 1 < Intermediate 4 < Intermediate 2 < Treatment with all 'seldom' levels). The worst treatment scenario with the highest occurrence of symptoms and side effects yielded the lowest utility in all valuation methods as expected. The best treatment scenario with the lowest occurrence of symptoms and side effects also yielded the highest utility in all valuation methods as expected. Figure 3 visualizes the contents of the four intermediate treatment scenarios. The utility of intermediate treatment 2 is the highest utility among all the intermediate treatments, which is also same as expected, since it is the only intermediate treatment without 'often' level in symptoms and side effects. It is concluded that the order of treatment scenarios estimated through all valuation methods are in right position.

FIGURE 3. THE FOUR INTERMEDIATE TREATMENTS EVALUATED THROUGH VAS, TTO AND BWS APPROACHES

Secondly, the graphical illustration of the average utility estimates of the six treatments with the scales of the four valuation methods was conducted and shown in Figure 4.

FIGURE 4. THE SCALES OF UTILITY ESTIMATES OF TREATMENTS THROUGH DIFFERENT METHODS



Even though the orders of treatments' value of the four approaches are the same, the scales of both Case 2 and Case 3 BWS are much larger than that of VAS and TTO. This shows that all valuation approaches are valid in ranking the overall value of various treatment scenarios but only differs in scales.

After examining the mean values of treatment scenarios using the four approaches. Convergent validity of Case 2 and Case 3 BWS were measured. Figure 5 to Figure 9 visualize the result of correlation analysis of each pair of valuation methods, which reflect the strength of linear dependence.

FIGURE 5. BIVARIATE PLOT AND SCATTERPLOT OF CASE 2 AND CASE 3 BWS

Pearson correlation coefficient: 0.830 (p-value: 0.000)

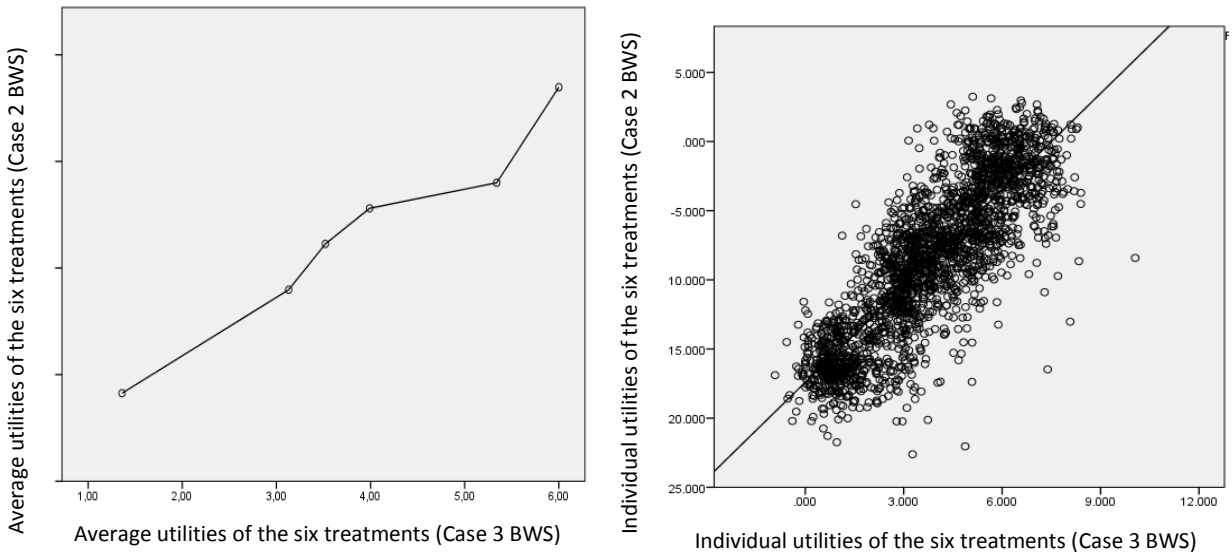


FIGURE 6. BIVARIATE PLOT AND SCATTERPLOT OF VAS AND CASE 2 BWS

Pearson correlation coefficient: 0.593 (p-value: 0.000)

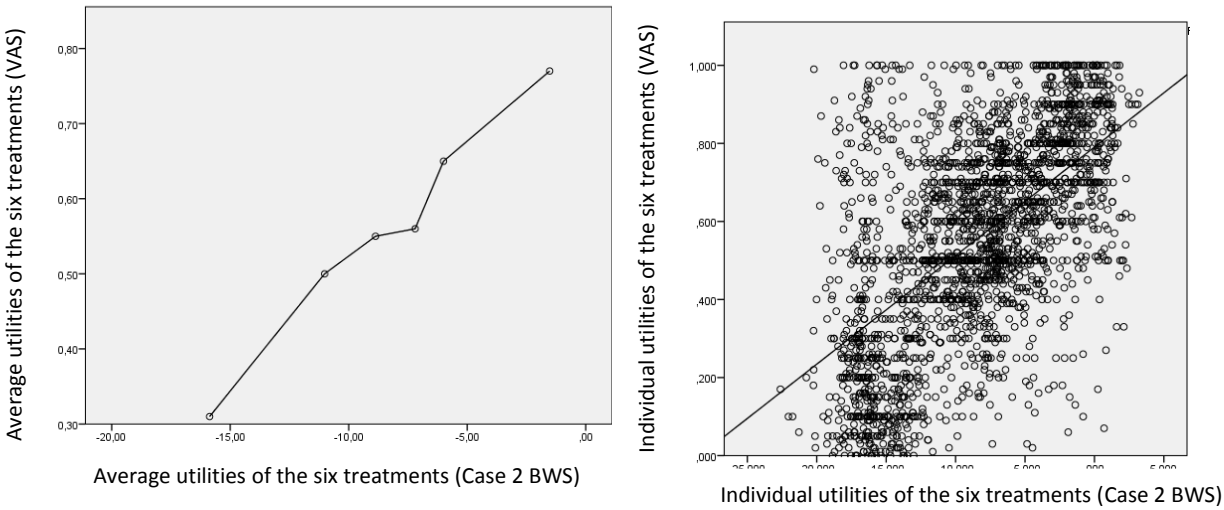


FIGURE 7. BIVARIATE PLOT AND SCATTERPLOT OF VAS AND CASE 3 BWS

Pearson correlation coefficient: 0.563 (p-value: 0.000)

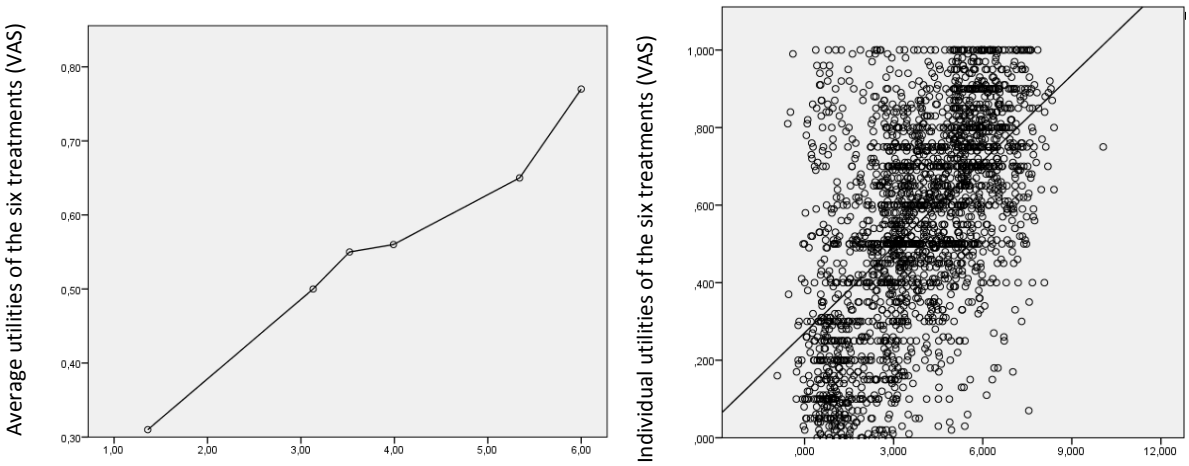


FIGURE 8. BIVARIATE PLOT AND SCATTERPLOT OF TTO AND CASE 2 BWS

Pearson correlation: 0.447 (p-value: 0.000)

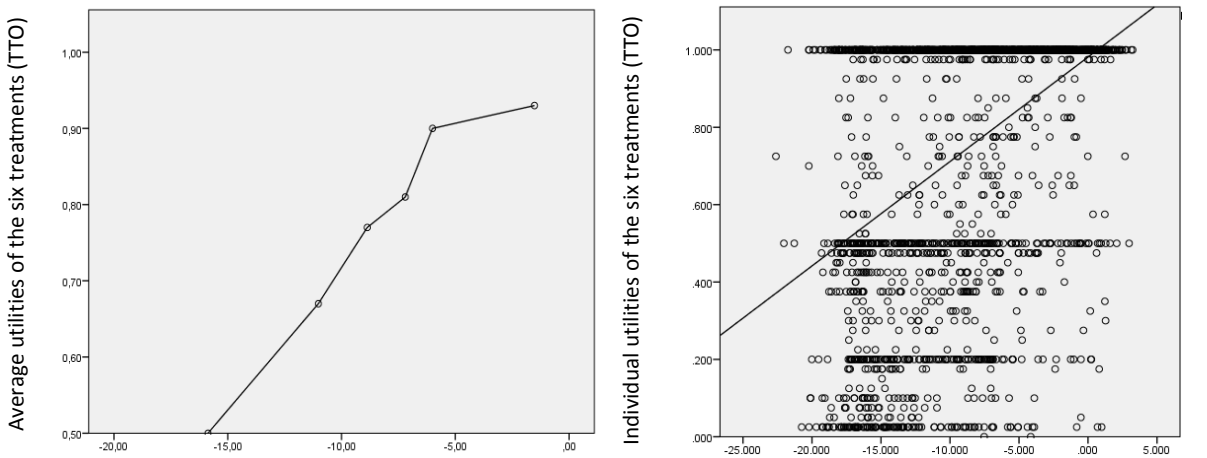
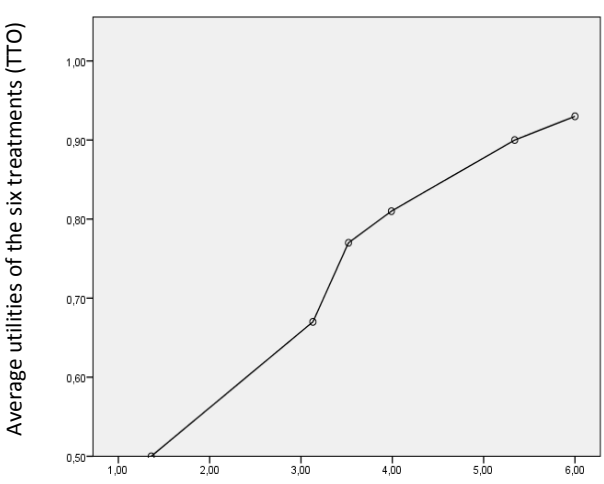
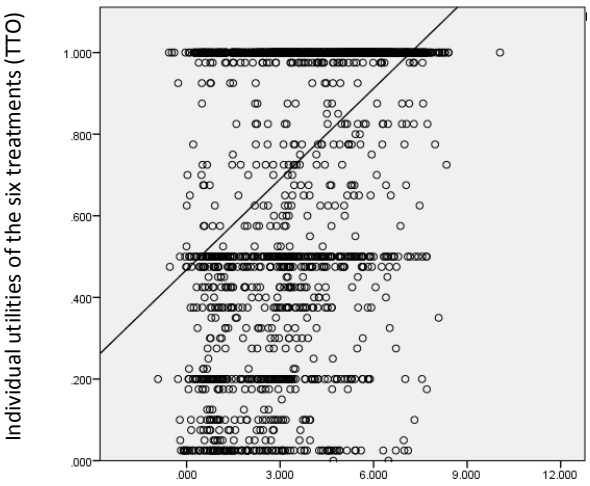


FIGURE 9. BIVARIATE PLOT AND SCATTERPLOT OF TTO AND CASE 3 BWS

Pearson correlation coefficient: 0.438 (p-value: 0.000)



Average utilities of the six treatments (Case 3 BWS)



Individual utilities of the six treatments (Case 3 BWS)

The result of correlation analysis at individual level is coherent with that examined through Pearson correlation test at mean level. The correlation between Case 2 and Case 3 BWS is the highest among the five pairs of valuation methods, as seen from the densest clustering of utility estimates which form a clear positive slope. The utility estimates also clustered densely in the two scatterplots of VAS versus Case 2 and Case 3 BWS, forming a positive slope, indicating linear convergence. On the other hand, the scatterplots related to BWS and TTO are much more dispersed when compared to that between BWS and VAS. . This indicates that the correlation between TTO and BWS approaches is lower than that between VAS and BWS approaches. Also, unlike the other scatterplots, there is dense clustering of utility estimates when utility value equals to 1 and 0.5 in the scatterplots of TTO with BWS approaches. Many participants chose these two numbers when expressing treatment preferences using TTO, which indicates possible numerical error or different risk attitudes which affect the result of utility estimates.

To sum up the result of convergent validity of BWS, there are significant convergence between BWS and traditional valuation methods, VAS and TTO. There is relatively highest convergence between Case 2 and Case 3 BWS, further demonstrated the convergent validity of BWS. Besides, Case 2 and Case 3 BWS correlate with VAS more than with TTO.

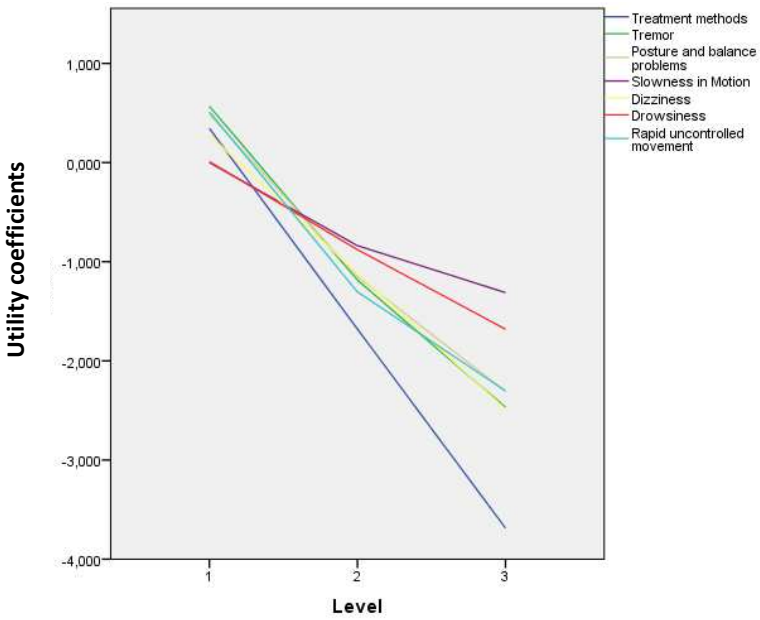
THEORETICAL VALIDITY OF CASE 2 BWS

Theoretical validity of Case 2 BWS was examined. After preliminary regression analysis, it was shown that 'Seldom suffer from slowness in motion' was the least influential attribute level from the view of participants. Thus it was set to be the reference level in the main regression model (Flynn et al., 2008). Table 5 lists the result of analysis which is the utility coefficients of attribute levels. Figure 1 illustrates the trend of each attribute level from 'seldom to never' to 'often' suffer from symptoms and side effect and from non-invasive to invasive approach.

TABLE 5 (LEFT). UTILITY COEFFICIENT ESTIMATES OF CASE 2 BWS (Coefficient with* indicates p <0.00)

FIGURE 1 (RIGHT). THE IMPACTS OF TREATMENT ATTRIBUTES BASED ON CASE 2 BWS

Attributes		Coefficient
Treatment methods	Oral tablets	0.345*
	Pump	-1.679*
	Brain Surgery	-3.687*
Tremor	Seldom to never	0.569*
	Sometimes	-1.186*
	Often	-2.469*
Posture & Balance problem	Seldom to never	0.488*
	Sometimes	-1.151*
	Often	-2.309*
Slowness in motion	Seldom to never	Reference
	Sometimes	-0.838*
	Often	-1.312*
Dizziness	Seldom to never	0.278*
	Sometimes	-1.012*
	Often	-2.487*
Drowsiness	Seldom to never	0.009
	Sometimes	-0.848*
	Often	-1.681*
Rapid uncontrolled movement	Seldom to never	0.511*
	Sometimes	-1.304*
	Often	-2.302*



As observed from the graph above, the result is consistent with the prior expectation, the more frequent the symptoms and side effects are the more negative the utility coefficients are. The treatment method is the most influential attribute impacting the utility of PD treatment. The utility coefficients of treatment method, from the non-invasive approach to more invasive approach decreased dramatically

from 0,345 to –3,687, which is the highest reduction in overall treatment utility across attribute levels. The impacts of having tremor, dizziness, rapid uncontrolled movement and posture and balance problems on treatment utility are also high next to the impact of treatment methods. Slowness in motion is relatively less significant in determining the utility of treatment as the reduction in treatment value is the least when this symptom occurs more frequently.

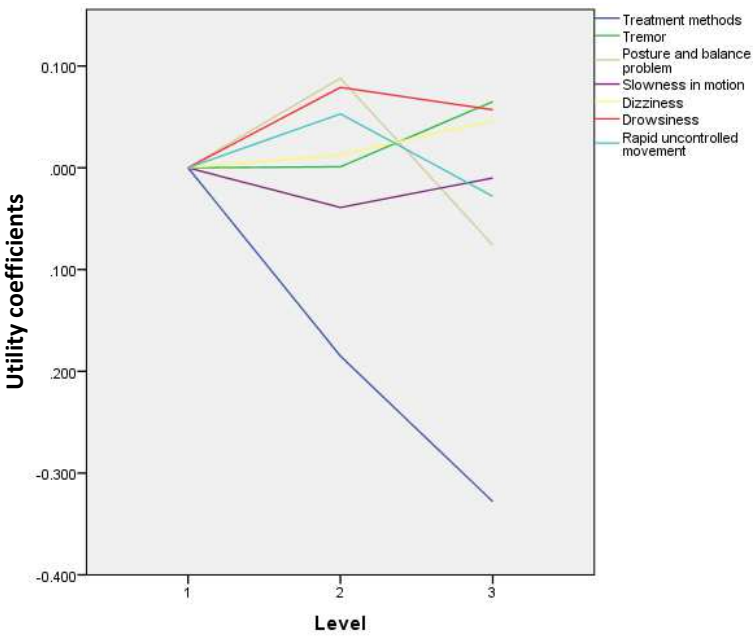
THEORETICAL VALIDITY OF CASE 3 BWS

The theoretical validity of Case 3 BWS was examined. The 'often' levels of attributes were set to be the reference levels in Case 3 BWS. Table 6 lists the outcome of analysis utility coefficient of attribute levels. Figure 2 illustrates the trend of each attribute level from 'seldom to never' to 'sometimes' suffer from symptoms and side effect, as well as from non-invasive to invasive approach.

TABLE 6 (LEFT). UTILITY COEFFICIENT ESTIMATES OF CASE 3 BWS (Coefficient with* indicates p <0.00)

FIGURE 2 (RIGHT). THE IMPACTS OF TREATMENT ATTRIBUTES BASED ON CASE 3 BWS

Attributes		Coefficient
Treatment methods	Oral tablets	-0.185*
	Pump	-0.328*
	Brain Surgery	Reference
Tremor	Seldom to never	0.001
	Sometimes	-0.065
	Often	Reference
Posture & Balance problem	Seldom to never	0.088*
	Sometimes	-0.076*
	Often	Reference
Slowness in motion	Seldom to never	-0.039
	Sometimes	-0.010
	Often	Reference
Dizziness	Seldom to never	0.013
	Sometimes	0.046
	Often	Reference
Drowsiness	Seldom to never	0.079*
	Sometimes	0.057*
	Often	Reference
Rapid uncontrolled movement	Seldom to never	0.053*
	Sometimes	-0.028
	Often	Reference



As seen from the graph of Case 3 BWS, the relatively steeper slopes compared to other attributes indicates their higher influence to the overall utility. Treatment method, as well as posture and balance problem are the most important attributes. The utility coefficient dropped from -0,185 to -0,328 when the treatment is more invasive, whereas the utility coefficient of posture and balance problem dropped from 0,088 to -0,076 when the symptom is more frequent. Rapid uncontrolled movement is the third influential attributes.

One more observation is that the slopes of slowness in motion, tremor, and dizziness are positive, which is not consistent with the prior expectation. This indicates that the utility of treatment increases when the frequency of having slowness in motion, tremor, and dizziness increase. Participants tend to trade off these two attributes to increase the overall utility of treatments.

COMPARISON BETWEEN THE RESULTS OF CASE 2 AND CASE 3 BWS

The result of Case 2 and Case 3 BWS was compared based on the measurements of utility coefficients. The similarity between them is that treatment method, as well as posture and balance problems are shown as the highly influential attribute to the utility of PD treatment in both approaches. Another similarity is that rapid uncontrolled movement are also another significant attribute. Besides, Slowness in motion is considered as the least influential attribute in Case 2 BWS and its utility coefficient also increases as its occurrence increases in Case 3 BWS, indicating that 'Slowness in motion' is relatively less important in determining the value of treatment.

There are differences in the estimation of tremor and dizziness in Case 2 and Case 3 BWS. As the frequency of having tremor and dizziness increases, their utility coefficients decrease in Case 2 BWS but they increase in Case 3 BWS.

To summarize the examination of theoretical validity of Case 2 and Case 3 BWS, the result of Case 2 is same as prior expectation whereas that of Case 3 BWS is not. One possible reason is that Case 3 BWS elicited the trade off behavior of participants. Besides, treatment method, posture and balance problem and rapid uncontrolled movements are the relatively important attributes of PD treatment.

For further details about the numerical result of conditional logit regression analysis based on Case 2 and Case 3 BWS (coefficients, standard error, associated p-value and 95% confidence interval) are listed in Appendix III.

FEASIBILITY OF USING BWS

Alongside the convergent and theoretical validity, feasibility of Case 2 and Case 3 BWS from the participants' perspective was measured based on the participants' preferences, quality of response, the time taken by respondents and the optional feedback from participants.

TABLE 8. PREFERENCE BETWEEN CASE 2 AND CASE 3 BWS BASED ON THE TASK SEQUENCE OF PARTICIPANTS

Sequence of performing BWS task	Number of participants who prefer Case 2 BWS (%)	Number of participants who prefer Case 3 BWS	χ^2 (p-value)
Case 2 BWS and then Case 3 BWS	128 (43%)	168 (57%)	11.92*
Case 3 BWS and then Case 2 BWS	170 (57%)	126 (43%)	(0.001)
Preference between Case 2 and Case 3 BWS	298 (50%)	294 (50%)	

As seen from the table, there is no particular preference for either Case 2 or Case 3 BWS. Differences by age and gender were also absent. 298 participants preferred Case 2 BWS, and 294 participants preferred Case 3 BWS. It is remarkable that relatively more participants preferred the BWS approach which was performed as the second BWS task. It is uncertain whether it is solely a statistical observation or the fact that the order of choice task influences the preference of participants.

In terms of the quality of response, 851 individuals participated in the survey. The completion rate is 71.7% and 97.0% of this population passed the data screening tests. The mean and median time for completing the Case 2 BWS are 5.0 minutes (S.D =5.6) and 3.8 minutes respectively, whereas those for completing the Case 3 BWS are 5.9 minutes (S.D =5.2) and 4.6 minutes respectively. The difference in time spent is approximately 1 minute longer in Case 3 BWS.

Speaking of the feedback from participants, 71 feedbacks were collected. Most participants commented that the survey was interesting and increased their awareness of overall treatment outcome. Nevertheless, 11 respondents expressed that questions were difficult to answer. In particular, one participant responded '*How can you often dizzy and yet seldom to never having balance problems? '*. This may be due to the fact that not every respondent read the explanation of PD treatments provided in the survey with full concentration. Nevertheless, the comment reveals the challenge to include mutually independent treatment attributes. Many symptoms of PD are interconnected with each other even though their pathological origins are different. More attention is required to select completely independent treatment attributes and yet capture the key characteristics of treatment outcome.

Another important issue is that two participants expressed their disagreement towards TTO task. One mentioned that '*If patients are happy, then regardless of their health state they will not choose to shorten their life span for perfect health*.', whereas the other one stated that 'Making *choice to shorten lifespan is not ethical*'. This reflects that TTO is not an appropriate utility eliciting approach for those participants who are against trading off life span for better health, even though it is considered as a better tool for valuing different health outcomes than VAS since it addresses the time duration spent on the health state (Tolley, 2009).

5. DISCUSSION

INTERPRETATIONS OF FINDINGS

CONVERGENT VALIDITY OF CASE 2 AND CASE 3 BWS

This study visualized the convergent validity of Case 2 and Case 3 BWS. The main research question is answered and the null hypothesis has been rejected. The alternative hypothesis is accepted, indicating that the utility estimates of treatments measured by Case 2 and Case 3 are both correlated with those measured by VAS and TTO. The convergence between both BWS approaches and methods are all significant with satisfactory correlation coefficient, indicating adequate convergent validity. Two literatures investigating on the convergent validity of BWS has been found. Both yielded significant convergent validity. However, direct comparison of findings is difficult since different methods were used to assess convergent validity. One of the studies investigated the correlation between BWS and two traditional questionnaires in measuring the preferences in handling conflicts, which demonstrated the convergent validity of BWS with significant Pearson correlation coefficient (Daly, Lee, Soutar & Rasmi, 2010). However, four correlation coefficients, ranging from 0.22 to 0.68 were calculated in each pair of valuation methods instead of one correlation coefficient. The values of four different conflict handling styles were correlated instead of an overall value, resulting in four correlation coefficients (Daly et al., 2010). On the other hand, the other study focused in measuring the convergent validity of BWS through correlating the values of food product by correlating the estimates measured through BWS and willingness-to-pay (Lusk & Briggeman, 2009). Spearman-ranking correlation coefficient was used in that study to account for the fact that four ordinal price options were provided for respondent to choose in willingness-to-pay to be correlated with BWS (Lusk & Briggeman, 2009). Pearson correlation was used in this study to account for the linear relationship among valuation methods. In view of the different concept and methodologies, it is difficult to compare the findings of these two studies with this study.

Comparatively, the convergence between BWS and VAS is higher than that between BWS and TTO. This may be due to the different risk attitudes to trade off life span for better health among participants, which resulted in weaker linear relationship between the utility estimates of TTO and BWS. Many participants express their utility judgment towards treatments as 1 and 0.5. By valuing treatment scenarios as 1, it indicates that participants refused to trade off life span for better health. By valuing most treatment scenario as 0.5, it indicates that numerical error is present as 0.5 may be the relatively easier option for making trade off, but not representing the actual utility of the treatment. This phenomenon indicates that some participants were insensitive in using number of years to trade off to express their treatment preferences. The correlation coefficient between Case 2 and Case 3 indicates good or even excellent relationship which illustrates the convergent validity between them. Given the completely different psychometric profiles and the different utility scales between traditional valuation methods and BWS, together with the satisfactory convergence between them, it can be concluded that convergent validity of BWS is adequate.

This study is the first one to illustrate the convergent validity of BWS under the context of medical decision making. Even though traditional methods can also prioritize the four intermediate treatments in right order as what BWS does, there are superiorities of BWS over traditional techniques which make it a more desirable tool for making the reimbursement decisions. Firstly, as aforementioned, the estimation error due to numbers and different attitudes towards trading off life span can be avoided in using BWS. Secondly, in order to make reimbursement decision towards different treatment options based on the preferences of stakeholders, each treatment options have to be rated by individuals through VAS and TTO to obtain their utility estimates. In contrast, with the use of BWS, once the treatment preferences of stakeholders has been measured using a set of approximately ten choice questions, the utility estimates of all possible treatment options with competing characteristics can be calculated without repeatedly asking individuals to rate any newly invented medicinal products. This can facilitate the assessment of the value of ever growing medical treatments, thus enhance the progress of making reimbursement decisions.

THEORETICAL VALIDITY OF CASE 2 AND CASE 3 BWS

Corresponding to the sub-question (1), the theoretical validity of Case 2 and Case 3 BWS were both examined. Unlike the result of Case 2 BWS, the result of three attributes in Case 3 BWS is opposite to the prior expectation. Possible interpretation is that even though having slowness in motion, tremor and

dizziness are significant factors, respondents prefer trading off these three attributes to have less invasive treatment method and higher performance in reducing posture and balance problems, rapid controlled movement, as well as drowsiness when they face treatment options with competing multidimensional outcomes.

The utility coefficient estimates of Case 2 BWS yields higher statistical significance. Almost all attribute levels are statistically significant except 'Seldom having drowsiness'. On the other hand, Case 3 BWS yields less statistically significant result, in which only seven attributes yield sufficient statistical result. One possible explanation is that the public is less experienced in PD and it is more cognitive demanding and challenging for them to make complicated trade off among treatments based on their true underlying utility scale, leading to more variations in their decisions. In light of the result of assessing theoretical validity, Case 2 BWS might be a better technique for measuring the treatment preference of the public in reimbursement decision since it is cognitively easier.

Also, treatment method, posture and balance problems, as well as rapid uncontrolled movements are the most influential attributes affecting the value of PD treatments. From the clinical point of view, the result is consistent with existing literature related to patient preferences. The invasiveness of drug delivery, gait balance and rapid uncontrolled movement are the major concerns when deciding treatments (Machado et al., 2006; Sujith & Lane, 2009).

FEASIBILITY OF BWS

In regard to the sub-question (2), it is concluded that BWS is a feasible instrument in measuring treatment preferences based on the response from the public. The quality of response is deemed satisfactory, with 71.7% completion rate and 97.0% of valid choice data. It is unexpected that participants has no particular preference between Case 2 and Case 3 BWS, however, it is worth for further examination since it is uncertain whether their preference is related to their sequence of performing choice task. Additionally, the time spent on Case 2 and Case 3 BWS are acceptable. The large value of standard deviation in mean time is possibly resulted from participants who took pause when performing the choice tasks. On average, only one more minute is needed to complete the Case 3 BWS choice task when compared to Case 2 BWS. It is important to note that this study only addressed the feasibility of BWS based on the public. Further assessment of feasibility of BWS is needed to explore this issue from the view of other parties, including the pharmaceutical companies.

This study demonstrated the convergent validity of BWS and highlighted the key advantages of BWS over traditional approaches. Compared between the performance of Case 2 and Case 3 BWS, it is difficult to judge which valuation method is better. Case 3 BWS captures the actual utility scale of respondents by mirroring the reality of decision making settings. It was also equally acceptable by participants as Case 2 BWS. The extra time required to perform Case 3 BWS choice task is minimal. On the other hand, the Case 2 BWS is cognitively easier for participants and the result is statistically more appealing. This study can only conclude that both Case 2 and Case 3 BWS correlate with VAS and TTO significantly and they are both suitable and feasible in measuring treatment preference towards chronic diseases, which is beneficial in reimbursement decision making.

LIMITATION

The main limitation is the insufficient number of subjects with prior knowledge about PD and those with relatively bad health state. These are the two main factors leading to differences in treatment preference (Montgomery & Fahey, 2001). Thus, the preference heterogeneity was not investigated.

FUTURE RESEARCH DIRECTION

Through this study, the convergent validity, theoretical validity and feasibility of BWS have been examined with positive results. BWS is worth for further investigation, so that this technique can be used in reimbursement and regulatory decision making. Based on the adequate convergent validity, the next step to validate BWS should be to investigate on how BWS estimated can be converted and be anchored onto the full QALY scale. There are ongoing researches investigating how to transform DCE utility estimates (Flynn et al., 2008; Rowen, Brazier and Van Hout, 2011). Three main approaches for conversion of utility estimates include anchoring one latent estimate obtained through DCE onto TTO estimates, mapping the DCE utility estimates in trial study to establish a statistical relationship between TTO and DCE, as well as using hybrid models (i.e. likelihood approach and Bayesian approach) to combine the DCE choice data and TTO choice data (Rowen, Brazier and Van Hout, 2011). The former approach has been criticized because of the lack of empirical basis and inconsistent result, whereas the two latter approaches are increasingly gaining recognition since both of them achieved consistent transformation of DCE values onto QALY scale (Flynn et al., 2008). Mapping approach uses the mean level data whereas hybrid model approach uses individual level data (Rowen, Brazier and Van Hout, 2011). Hybrid model may be more statistically appealing as it utilizes the choice data more effectively and worth further investigation on how it can be used in BWS.

Moreover, conditional logit regression model is used in this study to measure the treatment preference which functioned adequately in pooling the choice data of Case 2 BWS, yielding utility coefficients with statistically significant figures. The heterogeneity of treatment preferences was also captured using this regression model. However, the choice data of Case 3 BWS is less fit in this regression model compared to the choice data of Case 2 BWS. It is uncertain whether it is due to the quality of choice data or the use of inadequate regression model. It is of interest to explore better regression models. Nested logit model may be a plausible alternative in which the assumption of independence of irrelevant alternatives (IIA) associated with conditional logit regression model can be loosened and the error terms of the pairs treatment profiles can be correlated (Wen & Koppelman, 2001). Even though there may be other practical difficulties and theoretical limitations, ongoing investigations on BWS are crucial steps in overcoming the limitation of existing methods and improving the valuation methods for treatment outcomes of chronic diseases.

Besides, future research should also be more specifically targeting at further comparison between Case 2 and Case 3 BWS, in terms of their usage in various medical decision problems, their application on different types of population, as well as the opinions from various parties using more in-depth approach. Only by accumulating experiences, the methodology of BWS can be fine-tuned and become standardized measurement instrument to be used in reimbursement and regulatory medical decision making process.

7. CONCLUSION

To summarize, this study showed that Case 2 and Case 3 BWS have significant convergence to traditional measurement in the context of PD. From the view of respondents, Case 2 and Case 3 BWS are equally attractive; from theoretical point of view, Case 3 BWS captures the choice process better than Case 2 BWS does; however, from the point of view of the author, Case 2 BWS might be more feasible for public since it is easier to capture the treatment preference with significant values of the attributes' impact.

Last but not least, this study demonstrated how to apply BWS to measure treatment preferences in the context of regulatory medical decision making which involves multidimensional treatment with complicated trade-offs. BWS is recommended over traditional approaches to measure the preferences of various stakeholders. As a consequence, comparison among competing treatments can be conducted systematically, making the reimbursement decision with higher consensus. Ultimately, the societal resources allocation decisions can be wiser and the social benefits can be maximized.

Acknowledgment

First and foremost I offer my sincerest gratitude to my supervisor, Janine van Til, who has supported me throughout my research with her knowledge and guidance. Without her patience and tireless effort in guiding me to the right track, I could not perform the research at my best level. A hearty thanks to my co-supervisor, Karin Groothuis-Oudshoorn, who has provided valuable advice in statistics and support to me. I could not be able to proceed in data analysis without her help. I am also grateful to the PhD student in this research group, Marieke Weernink, who always supported me when I faced difficulties, especially in translating the questionnaire into Dutch. A special thanks also to the research group for their contributions to the research participants. I am also indebted to all the many research participants, I applaud them for their dedication and readiness to participate and I received valuable feedback from them.

Ultimately, the greatest motivation sustaining me through the research came from my family. I would like to thank my parents for their unconditional love and encouragement. My gratitude also goes to my beloved sister, Jessica, who has been encouraging me all the way to the terminus of this Master thesis.

REFERENCES

Agresti, A. (2002). *An Introduction to Categorical Analysis.* Hoboken: John Wiley & Sons, Inc.

Alkaharusi, H. (2012). Categorical Variables in Regression Analysis: A comparison of Dummy and Effect Coding, 4(2), 202-209.

Arbouw, M. E. L, Movig, K. L.L, Guchelaar, H. J., Poels, P. J. E., Van Vugt, J. P. P.& Neef, C. et al. (2008). Discontinuation of ropinirole and pramipexole in patients with Parkinson's disease: clinical practice versus clinical trials. . *European Journal of Clinical Pharmacology, 64,* 1021-1026.

Apolone, G., De Carli, G., Brunetti, M. & Garattini, S. (2001). Health-related quality of life (HR-QOL) and regulatory issues. An assessment of the European Agency for the Evaluation of Medicinal Products (EMEA) recommendations on the use of the HRQOL measures in drug approval. *Pharmacoeconomics, 19*, 187–195.

Asadi-lari, M., Tamburini, M. & Gray, D. (2004). Patients' needs, satisfaction, and health related quality of life: towards a comprehensive model. *Health and quality of life outcomes, 2* (32), 1-15.

Atthobari, J., Bos, J. M., Boersma, C., de Jong-van dan Berg, L. T. & Postma, M. J. (2005). Adherence of pharmacoeconomic studies to national guidelines in the Netherlands. *Pharmacy World & Science, 27* (5), 364-370.

Badia, X., Monserrat, S., Roset, M. & Herdman, M. (1999). Feasibility, validity and test±retest reliability of scaling methods for health states: The visual analogue scale and the time trade-off. *Quality of Life Research, 8*, 303-310.

Bennett, J. & Blamey, R. K. (2001). *The choice modelling approach to Environmental Valuation.* England: Edward Elgar Publishing Limited.

Bijlenga, D., Birnie, E. & Bonsel, E. (2009). Feasibility, Reliability, and Validity of Three Health-State Valuation Methods Using Multiple-Outcome Vignettes on Moderate-Risk Pregnancy at Term. *Value in Health, 12* (5), 821-827.

Bridges, J. F. P., Hauber, B., Marshall, D., Lloyd, A., Prosser, L. A., Regier, D. A. et al., (2011). Conjoint Analysis Applications in Health - A checklist: A report of the ISPOR Good Research Practises for conjoint analysis task force. *Value in Health, 14*(4), 403-413.

Carlsson, F. & Martinsson, P. (2003). Design techniques for stated-preference methods in health economics. *Health Economics, 12*(4), 281-294.

Chaudhuri, K. R., Healy, D. G. & Schapira, A. H. (2006). Non-motor symptoms of Parkinson's disease: diagnosis and management . *The Lancet Neurology, 5,* 235-245.

REFERENCES - CONTINUED

Chaudhuri, K. R. & Odin, P. (2010). The challenge of non-motor symptoms in Parkinson's disease . *Progress in Brain Research, 184,* 325-341.

Coulter, A. (1997). Partnerships with patients: the pros and cons of shared clinical decision-making. *Journal of Health Services Research & Policy, 2* (2), 112-121.

Craig, B. M., Busschbach, J. J.V. & Salomon, J. A. (2009). Modeling Ranking, Time Trade-Off and Visual Analogue Scale Values for EQ-5D Health States. *Medical Care, 47* (6), 634–641.

Cubo, E. (2010). Pharmacotherapy in the management of early Parkinson's disease: Cost-effectiveness and patient acceptability. *Clinic Economic Outcomes Research, 2* (1), 127-134.

Daly, T. M., Lee, J. A., Soutar, G. N. & Rasmi, S. (2010). Conflict-handling style measurement: a best-worst scaling application. *International Journal of Conflict Management, 21*(3), 281-308.

De Gaspari, D., Siri, C., Landi, A., Cilia, R., Bonetti, A., Natuzzi, F. et al. (2006) Clinical and neuropsychological follow up at 12 months in patients with complicated Parkinson's disease treated with subcutaneous apomorphine infusion or deep brain stimulation of the subthalamic nucleus, *Journal of Neurology Neurosurgery & Psychiatry, 77*, 450–453.

Di Luca, M., Baker, M. & Corradetti, R. et al. (2011). Consensus document on European brain research. *European Journal of Neuroscience, 33* (5), 768-818.

Dodel, R., Reese, J. P., Balzer, M. & Oertel, W. H. (2008). The economic burden of Parkinson's disease. . *European Neurological Review, 3* (2), 11-14.

Dolan, P., Gudex, C., Kind, P. & V. Williams. (1996). Valuing health states: A comparison of methods. *Journal of Health Economics, 15,* 209-231.

Dorsey, E. R., Constantinescu, R., Thompson, J. P., Biglan, K. M., Holloway, R. G., Kieburtz, K. et al. (2007). Projected number of people with Parkinson's disease in the most populous nations, 2005 through 2030. *Neurology, 68* (5), 384-386.

Essers, B. A, Dirksen, C. D. & Prins, M. H. (2010). Assessing the Public's Preference for Surgical Treatment of Primary Basal Cell Carcinoma: A Discrete-Choice Experiment in the South of The Netherlands. *Dermatologic Surgery, 36,* 1950–1955.

European Commission (1988). *Council Directive 89/105/EEC of 21 December 1988 relating to the transparency of measures regulating the prices of medicinal products for human use and their inclusion in the scope of national health insurance systems*. Retrieved 31 July, 2013 from http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31989L0105:en:HTML.

Euro Group. (1990). Euroqol: A new facility for the measurement of health related quality of life. *Health Policy, 16,* 199-208.

REFERENCES - CONTINUED

Euroqol Group (2013). *Valuation of EQ-5D*. Retrieved February 5, 2013 from http://www.euroqol.org/about-eq-5d/valuation-of-eq-5d.html

European Parkinson's Disease Association (2011). *European Parkinson's Disease Standards of Care Consensus Statement.* Retrieved July 10, 2013 from http://www.epda.eu.com/en/resources/parkinsons-consensus-statement/

Farrar, S., Ryan, M., Ross, D., & Ludbrook, A. (2000). Using discrete choice modelling in priority setting: an application to clinical service developments. *Social Science &* Medicine, 50(1), 63-75.

Feeny, D. (2000). A Utility Approach to the Assessment of Health-Related Quality of Life. *Medical Care, 38* (9), 151-154.

Ferreira, J. J. & Rascol, O. (2000) Prevention and therapeutic strategies for levodopa-induced dyskinesias in Parkinson's disease. *Current Opinion in Neurology, 13,* 431–436.

Flynn, T., Louviere, J., Peters, T.,& Coast, J. (2007). Best-worst scaling: what it can do for health care and how to do it. *Journal of Health Economics, 26,* 171-189.

Flynn, T. N., Louviere, J. J., Marley, A. A. J., Coast, J. & Peters, T. J. (2008). Rescaling quality of life values from discrete choice experiments for use as QALYs: a cautionary tale. *Population Health Metrics, 6* (6), 1-6.

Flynn, T. N., Louviere, J. J., Peters, T. J. & Coast, J. (2008). Estimating preferences for a dermatology consultation using Best-Worst Scaling: Comparison of various methods of analysis. *BioMed Central Medical Research Methodology, 8*, 76.

Flynn, T. (2010). Valuing citizen and patient preference in health: recent developments in three types of best-worst scaling. *Expert Review of Pharmacoeconomics & Outcomes Research, 10* (3), 259-267.

Flynn, T. & Marley, A. J. (2012) *Best Worst Scaling: Theory and Methods.* Retrieved July 10, 2013 from University of Technology Sydney, Centre for the Study of Choice Website: http://www.censoc.uts.edu.au/researchoutput/wp12-01.pdf

Grosios, K., Gahan, P. B. & Burbidge, J. (2010). Overview of health care in the UK. *The EPMA Journal, 1* (4), 529-534.

Bouvy, J. & Vogler, S. (2013). *Pricing and reimbursement policy: Impacts on Innovation.* Retrieved 31 July, 2013 from World Health Organization Website: http://www.who.int/medicines/areas/priority_medicines/BP8_3_pricing.pdf

Hensher, D. A., Rose, J. M. & Greene, W. H. (2005). Applied choice analysis: a primer. Cambridge: Cambridge University Press.

REFERENCES - CONTINUED

Huber, J. & Zwerina, K. (1996). The importance of utility balance in efficient choice designs. *Journal of Marketing Research, 33*(3), 307-317.

Jankovic, J. (2008). Parkinson's disease: clinical features and diagnosis . *Journal of Neurology, Neurosurgery & Psychiatry, 79* (4), 368-376.

Jenkinson, C., Fitzpatrick, R., Peto V., Greenhall, R. & Hyman N. (1997). The PDQ-8: Development and validation of a short-form Parkinson's disease questionnaire. *Psychology and Health*, *12*, 805-814.

Johnson, F. R., Lancsar, E., Marshall, D., Kilambi, V., Mühlbacher, A. & Regier, D. A. et al. (2013). Constructing Experimental Designs for Discrete-Choice Experiments: Report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force . *Value in Health, 16* , 3-13.

Kaplan, R. M. (2003). The significance. *Quality of Life Research, 12* (Supp 1), 74-82.

Keeney, R. L. & Raiffa, H. (1976). *Decisions with Multiple Objectives - Preferences and Value Tradeoffs*, New York: John Wiley & Sons.

Kind, P., Lafata, J. E., Matuszewski, K. & Raisch, D. (2009). The Use of QALYs in Clinical and Patient Decision-Making: Issues and Prospects. *Value in Health, 12,* Supplement 1, s227-s230.

Kjær, T. (2011). A review of the discrete choice experiment - with emphasis on its application in health care. Retrieved May 6, 2013 from University Of Southern Denmark, Web site: http://static.sdu.dk/mediafiles/Files/Om_SDU/Centre/c_ist_sundoke/Forskningsdokumenter/publications/Working%20papers/20051pdf.pdf

Kuhfeld, W. F. (2003). *Marketing Research Methods in SAS Experimental Design, Choice, Conjoint, and Graphical Techniques*, Cary: SAS Institute Inc.

Lancsar, E., Louviere, J., & Flynn, T. (2007). Several methods to investigate relative attribute impact in stated preference experiments. *Social Science & Medicine*, 64(8), 1738-1753.

Lancsar, E., Louviere, J., Donaldson, C., Currie, G. & Burgess, L. (2013). Best worst discrete choice experiments in health: Methods and an application. *Social Science & Medicine, 76,* 74-82.

Lees, A. J., Hardy, J., Revesz, T. (2009). Parkinson's disease. *Lancet, 373* (9680), 2055-2066.

Lee, S., Tsang, A. & Kwok, K. (2007). Twelve-month prevalence, correlates, and treatment preference of adults with DSM-IV major depressive episode in Hong Kong. *Journal of Affective Disorders, 98*, 129-136.

Leopold, N. A., Polansky, M. & Hurka, M. R. (2004). Drug Adherence in Parkinson's Diseases. *Movement Disorders, 19* (5), 513-517.

REFERENCES - CONTINUED

Louviere, J. J., Street, D., Burgess, L., Wasi, N., Islam, T. & Marley, A. A. J. (2008). Modeling the choices of individual decision-makers by combining efficient choice experiment designs with extra preference information. *Journal of Choice Modeling, 1* (1), 128-163.

Louviere, J., Hensher, D. A. & Swait, J. (2000). *Stated Choice Methods, analysis and application.* England: Cambrige University Press.

Lusk, J. L. & Briggeman, B. C. (2009). Food Values. *American Journal of Agricultural Economics, 91*(1), 184-196.

Machado, A., Rezai, A. R., Kopell, B. H., Gross, R. E., Sharan, A. D. & Benabid, A. L. (2006). Deep brain stimulation for Parkinson's disease: Surgical technique and perioperative management. *Movement Disorders, 21* (Supp 14), S247-258.

Marinus, J., Ramaker, C., van-Hilten, J.J. & Stiggelbout, A.M. (2002). Health related quality of life in Parkinson's disease: a systematic review of disease specific instruments. *Journal of Neurology Neurosurgery & Psychiatry, 72,* 241-248.

Marshall D., Bridges, J. F. P., Hauber, B., Cameron, R., Donnalley, L. & Fyie, K. et al. (2010). Conjoint analysis applications in health, how are studies being designed and reported. *The Patient, 3*(4), 249-256.

Marley, A., & Louviere, J. (2005). Some probabilistic models of best, worst, and beste worst choices. *Journal of Mathematical Psychology*, 49(6), 464-480.

McIntosh, E., & Louviere, J. (2002). *Separating weight and scale value: An exploration of best-attribute scaling in health economics*. Health Economics Study Group Meeting. United Kingdom: Brunel University.

McIntosh, E. & Ryan, M. (2002). Using discrete choice experiments to derive welfare estimates for the provision of elective surgery: Implications of discontinuous preferences. *Journal of Economic Psychology*, *23* (3), 367-382.

Montgomery, A. A. & Fahey, T. (2001). How do patients' treatment preferences compare with those of clinicians? *Quality Health Care, 10* (Supp I), 39-43.

National Public Health Compass (2011). Retrieved February 4, 2013 from National Institute for Public Health and the Environment, Ministry of Health Welfare and Sport Web site: http://www.nationaalkompas.nl/gezondheid-en-ziekte/ziekten-en-aandoeningen/zenuwstelsel-en-zintuigen/ziekte-van-parkinson/cijfers-ziekte-van-parkinson-prevalentie-incidentie-en-sterfte-uit-de-vtv-2010/

Nilsson, D., Nyholm, D. and Aquilonius, S. M. (2001). Duodenal levodopa infusion in Parkinson's disease–long-term experience, *Acta Neurologica Scandinavica*, *104*, 343–348.

REFERENCES - CONTINUED

Nuijten, M. C., Van Iperen, P., Palmer, C., Van Hilten, B. J. & Snyder, E. (2001). Cost-Effectiveness Analysis of Entacapone in Parkinson's Disease: A Markov Process Analysis. *Value in Health*, *4*(4), 316-328.

Obering, C.D., Chen, J.J. and Swope, D.M. (2006) Update on apomorphine for the rapid treatment of hypomobility ('off ') episodes in Parkinson's disease, *Pharmacotherapy, 26*, 840–852.

Orme, B. K. (2006). *Getting started with conjoint analysis: Strategies for product design and pricing research.* Madison Wisconsin: Research Publishers LLC.

Palmer, C. S., Schmier, J. K., Snyder, E. & Scott, B. (2000). Patient preferences and utilities for "off-time" outcomes in the treatment of Parkinson's disease . *Quality of life research, 9,* 819-827.

Pinto-Prades, J. L., & Rodríguez-Míguez, E. (2011). *The Lead Time Trade-Off: The case of health states better than death* (No. 1101). Universidade de Vigo, Departamento de Economía Aplicada.

Portney, L.G. & Watkins, M.P. (2000). *Foundations of Clinical Research: Applications to Practice.* New Jersey: Prentice Hall Upper Saddle River.

Prades, J. L. P. & Miguez, E. R. (2011). The Lead Time Trade-Off: The Case of Health States Better than Death. Working Paper 11.10. Universidad Pablo de Olacide, Department of Economics.

Rascol, O., Brooks, D. J., Korczyn, A. D.,De Deyn, P. P., Clarke, C. E. & Lang, A. E. (2000). A five-year study of the incidence of dyskinesia in patients with early Parkinson's Disease who were treated with Ropinirole or Levodopa. *The New England Journal of Medicine*, *342* (20), 1484-1491.

Richardson, J. (1994). Cost utility analysis: what should be measured? *Social Science & Medicine, 1* (7), 21.

Robinson, A. and Spencer, A., 2006. Exploring challenges to TTO utilities: Valuing states worse than dead. *Health Economics, 15*(4), 393-402.

Rowan, D.,Brazier, J. & Van Hout, B. (2011). A comparison of methods for converting DCE values onto full health-dead QALY scale. Retrieved July 24, 2013 from The University of Sheffield, Faculty of Medicine, Dentistry and Health Website: http://eprints.whiterose.ac.uk/74892/1/HEDSDP11.15.1.pdf

Ryan, M., Amaya-amaya, M. & Gerard, K (2008). Using discrete choice experiments to value health and health care. Dordrecht: Elsevier.

Senard, J. M., Raï, S., Lapeyre-mestre, M., Brefel, C., Rascol, O. & Montastruc, J. L. (1997). Prevalence of orthostatic hypotension in Parkinson's disease*. Journal of Neurology, Neurosurgery and Psychiatry,63*, 584-589.

REFERENCES - CONTINUED

Scott, A. & Vick, S. (1999). Patients, doctors and contracts: An application of principal-agent theory to the doctor-patient relationship. *Scottish Journal of Political Economy*, 46(2), 111-134.

Siderowf, A. D., Holloway, R. G. & Stern, M. B. (2000). Cost-Effectiveness Analysis in Parkinson's Disease: Determining the Value of Interventions. *Movement Disorder Society*, *15*(3), 439-445.

Street, D.J. & Burgess, L. (2007). *The Construction of Optimal Stated Choice Experiments: Theory and Methods.* Hoboken: Wiley.

Sujith, O. K. & Lane, C. (2009). Therapeutic options for continuous dopaminergic stimulation in Parkinson's disease*. Therapeutic Advances in Neurological Disorders, 2* (2), 105-113.

Szende, A. M., Leidy, N. K. & Revicki, D. (2005). Health-Related Quality of Life and Other Patient-Reported Outcomes in the European Centralized Drug Regulatory Process: A Review of Guidance Documents and Performed Authorizations of Medicinal Products 1995 to 2003. *Quality of Life Research, 8* (5), 534-548.

Tolley (2009). What is Health Utilities? Retrieved July 24, 2013 from Hayward Group Ltd, Hayward Medical Communications, Web site: http://www.whatisseries.co.uk/whatis/pdfs/What_are_health_util.pdf

Trochim, M. K. (2006). Retrieved July 10, 2013 from Research Methods Knowledge Base Website: http://www.socialresearchmethods.net/kb/convdisc.php

Wen, C. H. & Koppelman, F. S. (2001). The generalized nested logit model. *Transportation Research, 35,* 627-641.

Willke, R. J., Burke, L. B. & Erickson, P. (2004). Measuring treatment impact. a review of patient-reported outcomes and other efficacy endpoints in approved product labels. *Controlled Clinical Trials, 25,* 535– 552.

Wirdefeldt, K., Adami, H. O., Cole, P., Trichopoulos, D. & Mandel, J. (2011). Epidemiology and etiology of Parkinson's disease: a review of the evidence*. European Journal of Epidemiology, 26,* S1-S58.

World Health Organization (2013). Retrieved July 8, 2013 from World Health Organization Website: http://www.who.int/chp/en/

APPENDIX I

ADVANTAGES AND DISADVANTAGES OF TTO, VAS, CASE 2 AND CASE 3 BWS

	Time trade off	Visual Analogue Scale	Best-Worst Scaling
Advantage	• Convenient approach to prepare questions. • Individual utility estimates can be made easily. • The time duration being in the certain health situation is addressed.	• Convenient approach to prepare questions and answer questions. • Time requirement is short. • Individual utility estimates can be made easily.	• Cognitively easier since the ability of identifying extreme items in human is used. • Higher discriminative power • Free of numerical bias associated with rating. • The impact of attributes on health state can be analyzed. • Case 3 BWS can capture the reality of decision king process, thus giving reliable measures. • Obtain more data about the underlying utility scales of respondents based on the random utility framework.
Disadvantage	• Not reflecting the actual decision making process. • Lower discriminative power • Cognitive demanding to answer series of questions about how to make trade-offs between health state and full health. • Different trade off behavior among individuals, leading to error in estimates. • Some people may disagree to trade off time for better health state. • Inability to measure the impact of attributes.	• Not reflecting the actual decision making process. • Lower discriminating power • It can be abstract for respondent to use number to value treatment, which induces numerical errors. • Duration of health state cannot be addressed. • Inability to measure the impact of attributes.	• Individual choice model is more difficult to obtain. • The choice experiment and data analysis are more sophisticated.

(Badia, Monserrat, Roset & Herdman, 1999 ; Flynn & Marley, 2012)

APPENDIX II

COMPARISON BETWEEN CASE 2 AND CASE 3 BWS

	Case 2 BWS	Case 3 BWS
Similarities	• Both obtain more information by exploiting the choice questions asking respondents to state the best and the worst option. • Both utilize the human's ability of identifying extremes, thus with higher discriminative power. • Both include attributes associated levels. • Utility estimates are measured based on aggregating choice data. • Both utilize the frequency of best-worst pair within a choice to generate the impact of attribute levels and the associate utility of profiles.	
Differences	• More popular in medical decision making since the decision are relatively complicated and cognitive demanding, especially more difficult for patients.	• More popular in marketing industry, in which the decision are relatively more understandable and easy by respondents.
	• Choice set is presented as single profile, which is relatively easier.	• Choice set is presented as multiple profiles which are more cognitive demanding.
	• Respondent make choices within the attribute levels included in a single profile, which is not the reality in decision making.	• Respondent make choices among profiles with different combinations of attribute levels, which is an extended task of traditional discrete choice experiment. It reflects the real decision making process.
	• Direct comparison of the attribute levels' impact jointly is possible. The utility co-efficient of attribute levels are anchored on the same scale with one attribute level as the reference point.	• Direct comparison of attribute levels' impact is not possible. There are separate scales for each attribute. There is one attribute level being the reference point in the scale.
		(Hensher, Rose & Greene, 2005; Louviere et al., 2008; Flynn & Marley, 2012)

APPENDIX III

CONDITIONAL LOGIT REGRESSION ESTIMATES OF CASE 2 AND CASE 3 BWS

Attributes		Utility values estimates using Case 2 BWS					Utility values estimates using Case 3 BWS				
		Coefficient	Std. Error	P value	95% Conf. Interval		Coefficient	Std. Error	P value	95% Conf. Interval	
Treatment methods	Oral tablets	0.345*	0.681	0.000*	0.211	0.478	-0.185*	0.024	0.000*	-0.232	-0.137
	Pump	-1.679*	0.074	0.000*	-1.823	-1.534	-0.328*	0.035	0.000*	-0.378	-0.278
	Brain Surgery	-3.687*	0.073	0.000*	-3.83	-3.543	Reference				
Tremor	Seldom to never	0.569*	0.067	0.000*	0.437	0.670	0.001	0.024	0.972	-0.047	0.048
	Sometimes	-1.186*	0.075	0.000*	-1.335	-1.0.8	-0.065	0.025	0.008	-0.113	-0.165
	Often	-2.469*	0.075	0.000*	-2.616	-2.322	Reference				
Posture & Balance problem	Seldom to never	0.488*	0.069	0.000*	0.353	0.624	0.088*	0.024	0.000*	0.040	0.136
	Sometimes	-1.151*	0.075	0.000*	-1.298	-1.004	-0.076*	0.025	0.003*	-0.125	-0.026
	Often	-2.309*	0.075	0.000*	-2.456	-2.162	Reference				
Slowness in motion	Seldom to never	Reference					-0.039	0.026	0.135	-0.090	0.012
	Sometimes	-0.838*	0.078	0.000*	-0.991	-0.686	-0.010	0.025	0.698	-0.059	0.040
	Often	-1.312*	0.079	0.000*	-1.466	-1.157	Reference				
Dizziness	Seldom to never	0.278*	0.069	0.000*	-0.142	-0.414	0.013	0.025	0.592	-0.035	0.620
	Sometimes	-1.012*	0.076	0.000*	-1.161	-0.863	0.046	0.025	0.068	-0.003	0.951
	Often	-2.487*	0.074	0.000*	-2.633	-2.342	Reference				
Drowsiness	Seldom to never	0.009	0.071	0.900	-0.130	0.148	0.079*	0.025	0.002*	0.029	0.129
	Sometimes	-0.848*	0.075	0.000*	-0.996	-0.699	0.057*	0.026	0.027*	0.007	0.107
	Often	-1.681*	0.077	0.000*	-1.832	-1.530	Reference				
Rapid uncontrolled movement	Seldom to never	0.511*	0.068	0.000*	0.377	0.645	0.053*	0.025	0.032*	0.004	0.102
	Sometimes	-1.304*	0.077	0.000*	-1.454	-1.153	-0.028	0.024	0.259	0.075	0.020
	Often	-2.302*	0.074	0.000*	-2.447	-2.157	Reference				

APPENDIX IV

BEST-MINUS-WORST COUNT OF TREATMENT ATTRIBUTE LEVELS

		Best	Worst	Best-minus-worst
Treatment Methods	**Oral**	628	174	454
	Pump	237	510	-273
	Brain Surgery	53	1132	-1079
Tremor	**Seldom to never**	661	89	572
	Sometimes	101	161	-60
	Often	43	468	-425
Posture and Balance Problems	**Seldom to never**	637	45	592
	Sometimes	88	120	-32
	Often	63	473	-410
Slowness in Motion	**Seldom to never**	391	49	342
	Sometimes	120	34	86
	Often	146	205	-59
Dizziness	**Seldom to never**	540	50	490
	Sometimes	92	110	-18
	Often	41	499	-458
Drowsiness	**Seldom to never**	391	47	344
	Sometimes	133	65	68
	Often	64	243	-179
Rapid Uncontrolled Movement	**Seldom to never**	663	115	548
	Sometimes	131	185	-54
	Often	104	554	-450

APPENDIX IV (CONTINUED)

GRAPH OF BEST-MINUS-WORST COUNT OF TREATMENT ATTRIBUTE LEVELS