Treatment Preferences in Parkinson's Disease

Marieke G. M. Weernink

Treatment Preferences in Parkinson's Disease

Marieke G. M. Weernink

Treatment Preferences in Parkinson's Disease

ISBN: 978-90-365-4331-6

DOI: 10.3990/1.9789036543316

This thesis is part of the Health Sciences Series, HS 17-14, department Health Technology and Services Research, University of Twente, Enschede, the Netherlands. ISSN 1878-4968.

Printed by: Ipskamp Printing, Enschede

Cover design by: Kevin van Dijk

© Copyright 2017: Marieke G.M. Weernink, Enschede, the Netherlands.

All rights reserved. No part of this publication may be reproduced without permission of the copyright holder.

TREATMENT PREFERENCES IN PARKINSON'S DISEASE

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Twente,

op gezag van de rector magnificus,

prof. dr. T.T.M. Palstra,

volgens besluit van het College voor Promoties

in het openbaar te verdedigen

op donderdag 1 juni 2017 om 14.45

door

Marieke Geertruida Maria Weernink

geboren op 16 September 1988

te Beckum

Dit proefschrift is goedgekeurd door:

prof. dr. M.J. IJzerman (promotor)

dr. J.A. van Til (co-promotor)

Samenstelling promotiecommissie

Voorzitter

Prof. dr. Th.A.J. Toonen

Universiteit Twente

Promotor

Prof. dr. M.J. IJzerman

Universiteit Twente

Co-promotor

Dr. J.A. van Til

Universiteit Twente

Referent

Dr. P.F.M. Krabbe

Rijksuniversiteit Groningen

Leden

Prof. dr. J.H. Kerstholt

Universiteit Twente

Prof. dr. M.J.A.M. van Putten

Universiteit Twente

Prof. dr. W.B.F. Brouwer

Erasmus Universiteit Rotterdam

Prof. dr. A.M. Stiggelbout

Univeriteit Leiden

Paranimfen

Michelle Kip

Marloes Uitdewilligen

Contents

Chapter 1	General introduction	9
<i>Part I Preference elicitation methods in health care</i>		
Chapter 2	A systematic review to identify the use of preference elicitation methods in healthcare decision making	19
<i>Part II Estimating process-utility for treatments in Parkinson's Disease</i>		
Chapter 3	Involving patients in weighting benefits and harms of treatment in Parkinson's disease	45
Chapter 4	Comparison of patient and public preferences for treatment attributes in Parkinson's Disease	73
Chapter 5	Valuing treatments for Parkinson disease incorporating process utility: performance of best-worst scaling, time trade off, and visual analogue scales	91
<i>Part III Value clarification in clinical practice</i>		
Chapter 6	Exploration of values clarification methods based on conjoint analysis techniques: a systematic review	109
Chapter 7	Best-worst scaling case 2 for value clarification and decision support: Structured approach	137
Chapter 8	Best-Worst Scaling case 2 for value clarification and decision support: Patient's experiences	153
Chapter 9	General Discussion	169
Summary		179
Samenvatting		183
Dankwoord		189
Curriculum Vitae and Publications		193

1

General Introduction

Parkinson's disease (PD) is one of the most common neurodegenerative diseases in the Netherlands with prevalence rates ranging from 1300 to 1400 per 100.000 (1). PD is characterized by resting tremor, bradykinesia, muscle rigidity and a chronic, slowly progressive course of the disease (2). A pharmacological treatment is recommended once the symptoms are severe enough to impede daily life activities (3, 4). Treatments may improve motor functions, but have associated side effects (e.g. dyskinesia) which increase in severity with use. PD is a non-curative chronic disease, and therefore treatment is aimed at alleviating symptoms and improving the patient's quality of life and functional capacity (2).

VALUING TREATMENT OUTCOMES

The impact of PD or treatment outcomes on quality of life can be measured using stated preference methods. Stated preference methods are used to study preferences of patients, health care professionals, caregivers, and the general public regarding the value and relative desirability of a range of medical services, health experiences or health states (5). The elicitation of stated preferences is informative for a wide range of clinical and policy decisions, as they are indicative for how the most important stakeholders value health services.

Public institutions such as the National Institute for Health and Care Excellence (NICE) and Zorginstituut Nederland (ZiNL) use stated preference methods to measure quality-adjusted life-years (QALYs) in health economic evaluations (6-8). The QALY is a single index which incorporates both the effects of health interventions on survival time (life-years) and changes in quality of life (9-11). To calculate a QALY, quality of life needs to be expressed as an utility weight on a scale 0 (dead) to one (perfect health), which reflect the strength of preference for one health state relative to another health state. In welfare economics, it is argued that the desirability of a situation or health state can only be judged based on the utilities provided by individuals in that situation (12). In other words, patients are asked to state their preferences about particular health innovation that could maximize their welfare. These cardinal utility weights can be obtained using standard gamble (SG) (13) or time trade-off (TTO) exercises (14). Both have a sound theoretical basis in utility theory, and ask respondents for a direct utility assessment by asking them how much they would be willing to sacrifice to avoid being in a particular health state (15). Whereas SG is often denoted as the "gold standard", because it is directly based on the axioms of expected utility theory (13, 16), other health economists prefer the TTO, because it is easier to understand and implement (7, 17). A possible alternative technique is the Visual Analogue Scale (VAS), where individuals are asked to

express the strength of their preferences for particular health states by marking a point on a scale with two anchors, usually death (0) and perfect health (100) (18). Because the VAS does not include a trade-off, it is regarded as theoretically inferior to the choice-based SG and TTO methods. However, several options are available to map or convert VAS values into SG and TTO values (19-21).

Beyond Health Outcomes: Process-Utility

While the QALY is used as the preferred method, it is often debated whether health outcomes alone do sufficiently reflect societal values. For that reason, there has been a considerable debate about which other attributes should be included in the evaluation space (22). According to the “traditional” health economist view, utility assessment should be based upon evaluation of health states that describe outcomes, in particular quality of life. This has its origin in the narrowly defined consequentialism, a philosophy developed by Anscombe (23). The focus on health outcomes does not allow individuals to value benefit from other sources, more specifically non-health outcomes such as information provision, autonomy, and dignity and process attributes (e.g. treatment modality, location, waiting time) (24, 25). Although relevant, this view assumes that non-health outcomes only have value in that they are commodities which can be exchanged to derive health gains (26). On the other side of the spectrum, many people advocate widening the evaluation space and a growing body of literature exists in which procedural or process-utility is estimated as supported in welfare economics (12, 27). In addition, the extra-welfarist framework even broadens the evaluative space by not solely focusing on measuring utility, but also admitting other relevant outcomes such as capabilities, health and other non-goods. (28). Secondly, for the source of valuations it also can take into account others (general public, experts) than the affected individual (12).

Evaluating the process of care more explicitly is especially of relevance for evaluations in chronic diseases, where the possibility exists that discomfort of treatment can outweigh treatment benefits. In the case of PD, pharmaceuticals levodopa and dopamine agonist are proven to be effective, but not all desired treatment benefits are always achieved. Leopold et al. (2004) has shown that only 10% of patients with PD fully adhere to drug regimens as intended (29). Hence, besides the effectiveness of a drug, other attributes such as: ease of use, effort and process of care influence its use by patients and -through adherence- its efficacy (30, 31). As PD progresses, an initially effective treatment loses its efficacy over time and important treatment decisions have to

be made to ensure the desired clinical response. This might imply more frequent oral intake of medication, taking a higher dose, adding another drug or starting with 2nd line treatments such as: continuous pump infusion (subcutaneous or intravenous) or neurosurgery (deep brain stimulation) (32). Such 2nd line treatments differ from pharmacological therapy in its process as they require surgery, daily cleaning routines and use of mechanical equipment and may have a detrimental effect on the daily life of patients.

Estimating Process-Utility

The debate about health outcomes also led to the question how the value of non-health outcomes and process attributes should be measured. One way is to add these attributes to the health state descriptions and use a TTO or SG to measure process-related utilities. Brennan et al. (2013) found that only 15 studies reported an empirical measure of process-related utility (mostly measured by TTO or SG) that could be incorporated directly into QALY calculations (27). Ryan et al. (2008) stated that it would not be appropriate to use a TTO or SG to ask individuals whether they e.g. are willing to give up a number of life-years to have waiting time reduced by six months (TTO) (25). They propose a more patient-centered method which is able to show how individuals trade health outcomes alongside non-health outcomes and process attributes: conjoint analysis (33, 34). Beside economic evaluations, societal decision makers might aim to measure the exact influence of process-related and non-health outcomes in health care or reimbursement programs (35). Conjoint analysis methods were exactly therefore intended, as an attempt to go beyond health outcomes and to take account of non-health outcomes and process attributes in the delivery of health care (36, 37).

Conjoint Analysis

Conjoint Analysis (CA) has a long history in marketing, and has gained widespread use as a tool to elicit patient preferences for health care services (38, 39). CA methods are based on the premise that any good or service can be described by its attributes (characteristics) and that the extent to which someone values a good or service depends on the levels of these attributes (37). Both TTO and SG have a holistic approach: the entire health state domain is considered at once by individuals to obtain overall utility estimates (40). In contrast, CA-methods can also estimate the relative importance of included attributes, assess the tradeoffs made between the attributes, and calculate overall utility scores for health states (34, 41). Where SG and TTO ask individuals to express the

utility of health states in numerical terms, CA follows the concept of ordinal utility. Individuals are only asked which health state constitutes a higher value in order to establish an ordering or ranking of preferences (42). Previous research suggests that ordinal preference tasks such as CA are relatively easier to complete and are also less difficult to administer compared to TTO and SG (42-45).

CA can take several forms, but the most often used and well-studied technique is the discrete choice experiment (DCE) (37, 46). In a DCE, individuals are presented with a choice between (usually) two scenarios described in terms of characteristics (attributes) and associated levels. For each choice task, individuals are asked to choose the preferred scenario. More recently, BWS was introduced as a way of obtaining more information by asking respondents to select not only the best alternative (most preferred), but also the least preferred alternative (44). Essentially, respondents are asked to choose the pair that maximizes the difference in value between these two attributes from the set that was offered. BWS choice tasks can appear in three different forms, i.e. BWS case 1 to 3. BWS case 3 is comparable to a DCE choice set and asks respondents to choose the best and the worst from a number of scenarios (usually three). Alternatively, BWS case 2 let respondents consider one scenario at a time, and ask respondents to select the best and worst characteristic within one scenario. Lastly, BWS case 1 does not distinguish levels, but only asks individuals for the most and least preferred attribute (47).

USE OF STATED PREFERENCES IN CLINICAL PRACTICE

More than in policy decisions taking a societal perspective, clinicians have to discuss all important factors with their patients in clinical decisions and there is no debate on whether non-health outcomes or process characteristics should be discussed. The difficulty in clinical decision making, is that often treatment decisions are preference-sensitive decisions, where there is no single best treatment option. Therefore, the patient's opinion regarding the process of care (process-utility) should always be taken into account in treatment decisions. The use of stated-preference methods in shared decision making is more about weighing which strategies are most feasible for clinicians to help individual patients evaluate the desirability of options and identify which option he or she prefers (48). The term stated preference in clinical practice refers to a patient's overall most-favored option, after taking into account his or her attitudes toward the treatment characteristics (health outcomes, non-health outcomes and process characteristics) (49).

Several decision support tools are available to help clinicians and patients to deliberate about treatment options. A recent systematic review found that simple and straightforward decision support tools such as listing pros and cons or rating scales are most common in clinical practice, because they are easier to answer, less complicated to administer and take up less time from clinicians and patients (48). However, preference methods based on a theory or framework can be favored for their methodological strengths. Especially CA-methods have potential in clinical practice because of their ability to estimate the relative value of characteristics within a treatment decision opposed to SG and TTO. CA makes the trade-off between treatment characteristics or treatment options explicit, thereby reducing the use of certain heuristics or mental shortcuts, and preventing errors in reasoning (50). Despite these strengths, few studies have explored the use of CA in clinical practice (51-54).

THESIS AIMS

The objective of this dissertation was to investigate the influence of process-utility on the relative value of treatments. Process-utility is then studied in different chapters with reference to societal health care and (individual) shared clinical decisions. The thesis thereby specifically applied BWS to explore process-utility. This aim is further specified in three main objectives for the three different parts of this thesis.

1. Summarize the available literature on the use of stated preference methods in different healthcare decision contexts.
2. Estimating process-utility for treatments in Parkinson's Disease using different stated preference methods and for different perspectives.
3. Explore the use of BWS to measure process-utility in clinical decision making.

THESIS CONTENTS

Part I Preference Elicitation Methods in Health Care

The choice for a stated preference method is often dependent on the context of decision making and the aim of the decision maker. In **chapter 2**, the aim was to conduct a systematic review to identify which stated preference methods were used to support decisions in different healthcare

contexts, and compare them on their theoretical foundations, practicality/feasibility and technical requirements.

Part II Estimating Process-Utility for Treatments in Parkinson's Disease

In **chapter 3** a BWS case 2 experiment was questioned in a patient population with the aim to measure process-utility for the main treatment options in PD (oral intake of medication, continuous pump infusion of medication, and neurosurgery). Furthermore, in this chapter we assessed the congruence between patients' preferred and perceived involvement in treatment decision making.

Chapter 4 elaborates further on the issue of estimating process-related utility, but now including the debate on who should value health states in economic evaluations: patients or the general population. The objective of this chapter was to compare patient and public valuations for health-states incorporating a process-related attribute regarding treatments in PD using BWS case 2.

Due to a lack of head-to-head comparisons, it is unclear which stated preference methods differentiate best between health-states in estimating process-related utility. Therefore, the objective of **chapter 5** was to compare public valuations for health-states incorporating a process-related attribute regarding treatments in PD using BWS case 2, BWS case 3, TTO, and VAS.

Part III Value Clarification in Clinical Practice

As previously described, CA-methods have the ability to elicit the reasons behind patient choices by determining the relative value of characteristics within a treatment decision. Despite the strengths, only a few studies have used CA as a value clarification method to support individual patients in clinical decision making. The objective of **chapter 6** is to systematically review the use of CA to elicit preferences for clinical decision support at the individual patient level. We aim to identify actual CA design features with regard to the selection of attributes and levels, the design of choice tasks and methods used to clarify values.

To our best knowledge, BWS is never used as a value clarification method in clinical decision making, while literature suggests that BWS has a lower cognitive burden compared to other methods (discrete choice experiment, time trade-off, standard gamble) (42-44). Subsequently, **chapter 7** introduces and critically appraise a structured approach for designing value clarification exercises using BWS.

Following the presented framework in chapter 6, two value clarification exercises for treatment decisions in Parkinson's disease and localized prostate cancer are developed in **chapter 8**. The aim of chapter 8 was to examine the feasibility of these value clarification exercises from the patient's perspective.

Finally in **chapter 9** the results of the different studies are integrated and the practical and methodological consequences of this thesis are discussed.

REFERENCES

1. von Campenhausen S, Bornschein B, Wick R, Botzel K, Sampaio C, Poewe W, et al. Prevalence and incidence of Parkinson's disease in Europe. *European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology*. 2005;15(4):473-90.
2. Lees AJ, Hardy J, Revesz T. Parkinson's disease. *Lancet*. 2009;373(9680):2055-66.
3. Scottish Intercollegiate Guidelines Network (SIGN). Diagnosis and pharmacological management of Parkinson's disease. A national clinical guideline. Edinburgh: Scottish Intercollegiate Guidelines Network (SIGN), 2010 Jan. . Report No.: Contract No.: 113.
4. Connolly BS, Lang AE. Pharmacological treatment of Parkinson disease: A review. *JAMA* 2014;311(16):1670-83.
5. Brennan DF, Baraff LJ. Patient preferences [3]. *Ann Emerg Med*. 1995;26(2):240-1.
6. Banta D. The development of health technology assessment. *Health Policy*. 2003;63(2):121-32.
7. National Institute for Health and Clinical Excellence. Guide to the methods of health technology appraisal. 2013.
8. Zorginstituut Nederland. Guideline for economic evaluation in healthcare. Zorginstituut Nederland,, 2016.
9. Torrance GW, Feeny D. Utilities and quality-adjusted life years. *Int J Technol Assess Health Care*. 1989;5(4):559-75.
10. Briggs A, Claxton K. Decision Modelling for Health Economic Evaluation. New York: Oxford University Press; 2006.
11. Kind P, Lafata JE, Matuszewski K, Raisch D. The Use of QALYs in Clinical and Patient Decision-Making: Issues and Prospects. *Value Health*. 2009;12:S27-S30.
12. Brouwer WB, Culyer AJ, van Exel NJ, Rutten FF. Welfarism vs. extra-welfarism. *J Health Econ*. 2008;27(2):325-38.
13. von Neumann J, Morgenstern O. Theory of games and economic behavior. Princeton (NJ): Princeton University Press; 1944.
14. Torrance GW, Thomas WH, Sackett DL. A Utility Maximization Model for Evaluation of Health Care Programs. *Health Serv Res* 1972;7(2):118-33.
15. Neumann PJ, Goldie SJ, Weinstein MC. Preference-based measures in economic evaluation in health care. *Annu Rev Public Health*. 2000;21:587-611.
16. Torrance GW. Social preferences for health states: An empirical evaluation of three measurement techniques. *Socioecon Plann Sci*. 1976;10(3):129-36.
17. Fox-Rushby J. Economic Evaluation: McGraw-Hill Education; 2005.
18. Aitken RC. Measurement of feelings using visual analogue scales. *Proc R Soc Med*. 1969;62(10):989-93.
19. Dolan P, Sutton M. Mapping visual analogue scale health state valuations onto standard gamble and time trade-off values. *Soc Sci Med*. 1997;44(10):1519-30.
20. Brazier J, Green C, McCabe C, Stevens K. Use of visual analog scales in economic evaluation. *Expert Rev Pharmacoecon Outcomes Res*. 2003;3(3):293-302.
21. Stevens KJ, McCabe CJ, Brazier JE. Mapping between Visual Analog Scale and Standard Gamble data; result from the UK Health Utilities

- Index 2 valuation survey. *Health Econ.* 2006;15(5):527-33.
22. Ryan M, Gerard K. Inclusiveness in the health economic evaluation space. *Soc Sci Med.* 2014;108(0):248-51.
23. Anscombe GEM. Modern Moral Philosophy. *Philosophy.* 1958;33(124):1 - 19.
24. de Bekker-Grob E. Discrete choice experiments in health care: theory and applications 2009.
25. Ryan M, Skåtun D, Major K. Using Discrete Choice Experiments to Go Beyond Clinical Outcomes when Evaluating Clinical Practice. In: Ryan M, Gerard K, Amaya-Amaya M, editors. *Using Discrete Choice Experiments to Value Health and Health Care.* Dordrecht: Springer Netherlands; 2008. p. 101-16.
26. Donaldson C, Shackley P. Does "process utility" exist? A case study of willingness to pay for laparoscopic cholecystectomy. *Soc Sci Med.* 1997;44(5):699-707.
27. Brennan VK, Dixon S. Incorporating process utility into quality adjusted life years: a systematic review of empirical studies. *PharmacoEconomics.* 2013;31(8):677-91.
28. Chimbindi N, Bor J, Newell M-L, Tanser F, Baltussen R, Hontelez J, et al. Time and money: the true costs of health care utilization for patients receiving "free" HIV/tuberculosis care and treatment in rural KwaZulu-Natal. *JAIDS J Acquir Immune Defic Syndr.* 2015;70(2):e52-e60.
29. Leopold NA, Polansky M, Hurka MR. Drug adherence in Parkinson's disease. *Mov Disord.* 2004;19(5):513-7.
30. Wertheimer AI, Radican L, Jacobs MR. Assessing different perspectives on the value of a pharmaceutical innovation. *Southern Med Review.* 2010;3(1):24-8.
31. Arbouw MEL, Movig KLL, Guchelaar HJ, Poels PJE, Van Vugt JPP, Neef C, et al. Discontinuation of ropinirole and pramipexole in patients with Parkinson's disease: Clinical practice versus clinical trials. *Eur J Clin Pharmacol.* 2008;64(10):1021-6.
32. Varanese S, Birnbaum Z, Rossi R, Di Rocco A. Treatment of advanced parkinson's disease. *Parkinsons Dis.* 2010.
33. Ryan M. Discrete choice experiments in health care. *BMJ.* 2004;328(7436):360-1.
34. Ryan M, Farrar S. Using conjoint analysis to elicit preferences for health care. *BMJ.* 2000;320(7248):1530-3.
35. Cleemput I, Devriese S, Kohn L, Devos C, van Til J, Groothuis-Oudshoorn K, et al. Incorporating societal preferences in reimbursement decisions – Relative importance of decision criteria according to Belgian citizens. Brussels: Belgian Health Care Knowledge Centre (KCE), 2014 Contract No.: KCE Reports 234.
36. Ryan M, Hughes J, Economics UCDCfH, Unit UoAHER. *Using Conjoint Analysis to Value Surgical Versus Medical Management of Miscarriage:* University College Dublin, Centre for Health Economics; 1995.
37. Ryan M, Bate A, Eastmond CJ, Ludbrook A. Use of discrete choice experiments to elicit preferences. *Qual Health Care.* 2001;10(SUPPL. 1):i55-i60.
38. Cattin P, Wittink DR. Commercial Use of Conjoint Analysis: A Survey. *J Marketing.* 1982;46(3):44-53.
39. Bridges JF, Hauber AB, Marshall D, Lloyd A, Prosser LA, Regier DA, et al. Conjoint analysis applications in health--a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. *Value Health.* 2011;14(4):403-13.
40. Szeinbach SL, Barnes JH, McGhan WF, Murawski MM, Corey R. Using conjoint analysis to evaluate health state preferences. *Drug Inf J.* 1999;33(3):849-58.
41. Mele NL. Conjoint analysis: using a market-based research model for healthcare decision making. *Nurs Res.* 2008;57(3):220-4.
42. Ali S, Ronaldson S. Ordinal preference elicitation methods in health economics and health services research: using discrete choice experiments and ranking methods. *Br Med Bull.* 2012;103(1):21-44.
43. Potoglou D, Burge P, Flynn T, Netten A, Malley J, Forder J, et al. Best-worst scaling vs. discrete choice experiments: an empirical comparison using social care data. *Soc Sci Med.* 2011;72(10):1717-27.
44. Louviere JJ, Flynn TN, Marley AAJ. *Best-Worst Scaling Theory, Methods and Applications.* Cambridge: Cambridge University Press; 2015. 360 p.

45. Brazier J, Deverill M, Green C, Harper R, Booth A. A review of the use of health status measures in economic evaluation. *Health Technol Assess.* 1999;3(9):iii-163.
46. Clark MD, Determann D, Petrou S, Moro D, de Bekker-Grob EW. Discrete choice experiments in health economics: a review of the literature. *Pharmacoeconomics.* 2014;32(9):883-902.
47. Flynn TN. Valuing citizen and patient preferences in health: recent developments in three types of best-worst scaling. *Expert Rev Pharmacoecon Outcomes Res.* 2010;10(3):259-67.
48. Fagerlin A, Pignone M, Abhyankar P, Col N, Feldman-Stewart D, Gavaruzzi T, et al. Clarifying values: An updated review. *BMC Med Inform Decis.* 2013;13(SUPPL. 2).
49. Llewellyn-Thomas HA, Crump RT. Decision support for patients: values clarification and preference elicitation. *Med Care Res Rev.* 2013;70(1 Suppl):50s-79s.
50. Janis IL, Mann L. *Decision making: a psychological analysis of conflict, choice, and commitment*: Free Press; 1977.
51. Hawley S, Newman L, Griggs J, Kosir M, Katz S. Evaluating a Decision Aid for Improving Decision Making in Patients with Early-stage Breast Cancer. *Patient.* 2015:1-9.
52. Jayadevappa R, Chhatre S, Gallo JJ, Wittink M, Morales KH, Bruce Malkowicz S, et al. Treatment preference and patient centered prostate cancer care: Design and rationale. *Contemp Clin Trials.* 2015;45(Pt B):296-301.
53. Rochon D, Eberth JM, Fraenkel L, Volk RJ, Whitney SN. Elderly patients' experiences using adaptive conjoint analysis software as a decision aid for osteoarthritis of the knee. *Health Expect.* 2014;17(6):840-51.
54. Fraenkel L, Rabidou N, Wittink D, Fried T. Improving informed decision-making for patients with knee pain. *J Rheumatol.* 2007;34(9):1894-8.

2

Systematic Review to Identify the Use of Preference Elicitation Methods in Healthcare Decision Making

Marieke G. M. Weernink*

Sarah I. M. Janus*

Janine A. van Til

Dennis W. Raisch

Jeannette G. van Manen

Maarten J. IJzerman

* contributed equally to this work

Pharm Med. 2014;28(4):175-85. doi:10.1007/s40290-014-0059-1.

ABSTRACT

Introduction: Preference elicitation methods help to increase patient-centred medical decision making(MDM)by measuring benefit and value. Preferences can be applied in decisions regarding reimbursement, including health technology assessment (HTA); market access, including benefit–risk assessment (BRA); and clinical care. These three decision contexts have different requirements for use and elicitation of preferences.

Objectives: This systematic review identified studies using preference elicitation methods and summarized methodological and practical characteristics within the requirements of the three contexts. Methods The search terms included those related to MDM and patient preferences. Only articles with original data from quantitative preference elicitation methods were included.

Results: The selected articles (n = 322) included 379 preference elicitation methods, comprising matching methods [MM] (n = 71, 18.7 %), discrete choice experiments [DCEs] (n = 96, 25.3 %), multi-criteria decision analysis (n = 12, 3.2 %) and other methods (n = 200, 52.8 %; i.e. rating scales, which provide estimates inconsistent with utility theory). Most publications of preference elicitation methods had an intended use in clinical decisions (n = 134, 40 %). Fewer preference studies had an intended use in HTA (n = 68, 20 %) or BRA (n = 12, 4 %). In clinical decisions, rating, ranking, visual analogue scales and direct choice are used most often. In HTA, DCEs and MM are both used frequently, and elicitation of preferences in BRA was limited to DCEs.

Conclusion: Relatively simple preference methods are often adequate in clinical decisions because they are easy to administer and have a low cognitive burden. MM and DCE fulfil the requirements of HTA and BRA but are complex for respondents. No preference elicitation methods with a low cognitive burden could adequately inform HTA and BRA decisions.

INTRODUCTION

In recent years, patients, policy makers and health professionals have expressed a desire for greater patient and public involvement in healthcare decision making (1, 2). Patient involvement in clinical decision making (CDM) has been widely encouraged (3); it is a collaborative process between the physician and the patient. Patient involvement in health technology assessment (HTA) and benefit–risk assessment (BRA) is thought to ensure that any decisions that are made increase patient welfare by meeting patient needs and desires (4–6). Moreover, not only patients but also the public—i.e. citizens and taxpayers—need to be engaged in policy decision making (4). Presently, patient involvement in decision making is usually operationalized through direct involvement of the patient or a patient representative in a group of decision makers, these being either health professionals in clinical care or experts in decision panels deciding on reimbursement or drug market access at the policy level (7). More recently, indirect involvement of the patient and the public perspective on health innovations has received increasing attention. Indirect elicitation of the patient perspective is proposed through various methods that aim to measure the usefulness of health innovations for patients (or the public) in a representative sample of the population. The results are presented to decision panels for consideration during decision making. Indirect involvement is thought to increase the representativeness of the patient perspective in deliberation on the desirability of adopting a particular health innovation (6).

The proposed methods for indirect elicitation of the patient perspective on the value of health innovations vary widely in terms of their methodological characteristics and practical applications. For instance, patient-reported outcomes (PROs) are a well-established method used to evaluate health innovations from the perspective of the patient (8). The simplest PROs use various types of rating scales to measure the effectiveness of an innovation, especially with regard to distinct aspects of a patient's health. The health-related quality-of-life (HRQoL) measurement offers the opportunity to value health innovations in several aggregated domains of health. PROs focus on identifying the state of health or the increase in health as a result of an intervention, from the perspective of the patient. However, most PROs, including the HRQoL measurement, focus on health gains or losses resulting from the healthcare innovation but do not address the 'utility' of an innovation across domains (8). Estimating the utility of outcomes is central to the concept of preference methods and differentiates preference methods from PROs.

Because there is a large variety of preference methods, it is often difficult to determine which concept of utility is being considered by the authors of a study. In the neoclassical utility theory,

utility has ordinal properties and is regarded as a measure of the strength of a preference regarding an outcome. The concept of the expected utility theory, as defined by von Neumann and Morgenstern (9), is often used in health economics. In this theory, the measurement of utility is based on preferences regarding random health outcomes and results in a cardinal measure of utility. In response to this confusion, Carson and Louviere (10) proposed a common nomenclature for preference methods. In their nomenclature, preference elicitation methods are, first and foremost, categorized on the basis of the nature of the information that is collected, and methods are defined as (a) matching methods (MM); (b) discrete choice experiment (DCE); and (c) other methods. Although it was developed from an environmental economics perspective, this nomenclature has been previously applied in healthcare research (11). The key difference between MM and DCE is the valuation response. In MM, respondents are asked to provide a number (or numbers) that will make them indifferent to the good outcome to be valued and the number of an explicit valuation scale. Valuation scales can be monetary, such as willingness to pay (WTP), or may include time spent in a health state, such as in time trade-off (TTO). A specific subset of MM includes standard gamble (SG), in which uncertainty, operationalized as the risk of death, is explicitly incorporated into the valuation task. DCEs belong to another category and are considered as “a general preference elicitation approach that asks respondents to make choice(s) between two or more discrete alternatives where at least one attribute of the alternative is systematically varied across respondents in such a way that information related to preference parameters of an indirect utility function can be inferred” (10). The DCE measurement scale is an implicit cardinal utility scale. Various DCE methods can be distinguished on the basis of the format of the question. For instance, respondents can be asked to provide a binary choice for an alternative (binary choice experiment [BCE]) or a multinomial choice between various scenarios (multinomial choice experiment [MCE]), or they can provide a complete ranking of a set of alternatives based on preferences. In best–worst choice experiments (BWS), respondents are asked to identify both the best and worst alternatives from a set. In addition to the different preference methods identified by Carson and Louviere (10), multi-criteria decision methods (MCDM) are increasingly being used to explicitly consider patient and/or public preferences in decision making in healthcare. While preferences regarding the characteristics of the innovation are calculated compositionally in most preference elicitation methods, i.e. the respondent values the innovation as a whole, and the impact of the various characteristics is calculated from these overall valuations, MCDM offers a compositional approach to preference elicitation. In the compositional methods, respondents evaluate (pairs of) criteria separately, after which an overall composite value for the innovation is estimated (12, 13).

Generally, MM-elicited utilities have cardinal properties, while preferences resulting from DCEs have ordinal properties. Other methods, such as MCDM and those discussed in the article published by Carson and Louviere (10) (e.g. visual analogue scales [VAS] and numerical rating scales), do not provide utilities in the economic sense, because the concept of risk is not incorporated. These methods were developed on pragmatic grounds rather than being underpinned by theoretical assumptions.

Preference methods offer the potential to increase patient-centred healthcare decision making by offering some measure of benefit along with some measure of value. Several reviews of preference methods have already been published (14-17). The studies by Hauber et al. [15, 16] and Dolan (15, 16) also address the context of use. Our paper follows those two reviews but also provides an extensive overview of methods and their use in healthcare decision making. The objective of this paper is to identify studies that used a preference elicitation method, representing the patient's view. This includes methods that (a) measure multiple domains of health or healthcare; (b) require trading between desirable and undesirable aspects of health innovation (also called attributes); and (c) offer a formal approach to valuing the underlying treatment components. The second objective is to identify the intended use of the results of the studies for three types of healthcare decisions: (a) clinical decisions; (b) HTA/ reimbursement decisions; and (c) BRA/market access decisions. Finally, the suggested fit of the preference methods actually used in decision-making contexts is discussed according to the demands of the decision context.

METHODS

Search Strategy and Selection of Articles

A literature search was performed in Scopus and Web of Science in January and February 2014. The basic search string included the following keywords: 'patient preference', 'patient value', 'patient perspective' or 'patient choice' in the paper titles. Several database features, such as truncation, proximity operators and phrase searching, were used (if available in the search engine) to increase the relevancy of retrieval of the free-text search. This basic string was combined with another search string, identifying which healthcare decision was mentioned in the title, abstract or keywords: ('benefit risk' OR 'risk benefit'), ('health technology assessment' OR 'reimbursement'), ('shared decision making' OR 'clinical decision making') or ('medical decision making' OR 'healthcare decision making'). Publications had to be written in English and had to be labelled as

original research articles. The abstracts were read by two authors independently (M.W. and S.J.) so as to minimize selection and data bias. Disagreements were resolved during a discussion between the co-authors. Articles were included if they used a quantitative preference method to elicit patient and/or public preferences regarding healthcare innovations and linked the use of such a method to support decision making. Conversely, articles were excluded if they had not collected preference data or used qualitative methods to collect preferences. In cases of uncertainty of categorization, the entire articles were read. The search strategies are described in detail in Appendix A.

Data Extraction and Interpretation

The data extraction and interpretation consisted of four steps. In step 1, all of the methods identified in the articles were described on the basis of their methodological characteristics. In step 2, the articles were categorized on the basis of the intended practical use of the results in the preference study. In step 3, all of the methods were judged on three criteria related to practical application, to identify the strengths and weaknesses of each method. Finally, in step 4, the practical use of the methods in the various decision contexts (step 2) was matched with the demands of each decision context, in terms of the identified methodological demands (step 1) and the practical demands (step 3). A more detailed description of the steps is provided below.

In step 1, the methods that were found were categorized according to the type of preference method distinguished by the nomenclature provided by Louviere and Carson (10) (MM, DCE, MCDM or other). The methods were ascribed correspondingly with the following methodological properties: direct or indirect weighting of criteria, using single or sequences of questions and outcome measures. Additionally, the need for explicit trading behaviour on the part of the respondent was described. Trading is thought to imitate the real-life context, where no result is obtained without costs, whether monetary costs or potential health loss. Explicit trade-offs can be made by trading outcomes with various characteristics of the innovation, such as a benefit–risk trade-off, or by reference to good versus bad trade-offs to gain health, i.e. the probability of unfavourable outcomes, such as the risk of death or a reduced lifespan, as in TTO. A methodological categorization of preference methods was provided by the use of key references.

In step 2, we distinguished intended from actual use in CDM, HTA and BRA. In CDM, preferences are used to support decision making by offering a measure for patients' values. In HTA, public or patient preferences are used to support reimbursement decisions, and in BRA, patients' values are

elicited concerning the perceived benefits and risks of a health innovation, thus advising market authorization. If no use was described by the original authors or the use could not be deduced by the reviewers, the article was classified as 'decision context not specified'. Subsequently, the attributes used to describe healthcare innovation were extracted. According to Ryan et al. (18), the attributes of health innovations can be classified as either health outcomes, non-health outcomes or process characteristics. Health outcomes describe the impact of an innovation on patients' health in terms of factors such as effectiveness, the risks involved, symptom relief and life expectancy. Non-health outcomes entail other attributes of care that are related to patient benefit, such as information provision and attention given to the patient and their carers. Process characteristics describe the physical characteristics of the innovation, such as the mode of drug administration, waiting time, location, frequency of consultations and costs to the patient or society.

In step 3, three criteria were used to judge the practical application of the methods, identifying their strengths and weaknesses. These criteria were 'cognitive effort on the part of the respondent', 'costs of data collection' and 'skills required in data analysis and interpretation'. 'Cognitive effort' assesses the perceived difficulty of the method for respondents. Methods that require single choices or few rankings or ratings were categorized as requiring low cognitive effort, while methods that require trading between multiple criteria were categorized as requiring high cognitive effort. The costs of data collection entail all practical demands regarding data collection. The need for large sample sizes to calculate reliable estimates, collecting data by interviews and large time investments are seen as practical difficulties. Small sample sizes and administration via online surveys are considered as presenting low or medium practical difficulties, depending on the estimated time investments. The 'skills required in data analysis and interpretation' describe the level of difficulty in applying and/or understanding a method and are categorized as (a) basic (do not require medical/statistical expertise and do not require specialist software to implement); (b) intermediate (do not require extensive medical/statistical expertise but may require implementation of specialist software); or (c) advanced (require extensive medical/statistical expertise and may require implementation of specialist software). Wherever possible, information about these three criteria was extracted from articles comparing specific preference methods. Additional scientific literature regarding comparison of these specific methods was consulted when the reference list was considered to be insufficient. Ultimately, if there was no literature on a particular method, a judgment of the method on the criteria was determined on the basis of agreement between all authors of this paper.

In step 4, the practical use of the methods in the three decision contexts was matched with the requirements of the decision context, in terms of the methodological and practical demands identified in earlier steps. The following hypotheses were formulated for each decision context. In the CDM setting, a simple, hands-on approach is needed, which can be used for elicitation of the preference of a single patient or a group of patients. The methods should allow for inclusion of various types of attributes deemed relevant to patients, and should therefore allow for inclusion of health, non-health and process characteristics. In order to be used in clinical practice, the methods should preferably be low in cognitive effort on the part of the respondent, should be low in costs and should require only basic technical skills in analysis and interpretation. In the clinical context, there is no need for formal calculation of (part-worth) utility or use of the outcomes in other benefit measures, as preferences regarding alternatives (I prefer treatment A to treatment B) will generally suffice in supporting decisions. Compared with clinical decisions, BRA and HTA decisions have more impact on society. These high-impact decisions suggest a preference elicitation method that assists in making the decision-making process transparent, legitimate and accountable. In the policy context, it is important to develop a benefit function; thus, adherence to utility theories is essential. Utilities are preferences measured under conditions of uncertainty, which gives the elicited preferences a solid methodological basis. Furthermore, HTA refers to the evidence of clinical effectiveness, safety and cost-effectiveness across groups of patients; therefore, the methods should account for several patient-relevant attributes and outcome measures. In the BRA context, drugs are approved on the basis of their safety and effectiveness; thus, the method needs to incorporate multiple health outcomes. For both BRA and HTA decision-making contexts, the preference elicitation methods should be easy and simple for patients to understand. In comparison with CDM, there is more time for collecting preferences, because the approval process takes months or even years. Furthermore, there are more resources and time for collection and analysis of preferences. Methods with more advanced statistics are therefore possible.

RESULTS

Results of Search Strategy

The search strategy identified 1,036 unique articles (see Figure 1). The initial selection revealed 322 articles that used one or more quantitative methods to elicit patients' preferences; these are listed in online Appendix B (doi:10.1007/s40290-014-0059-1).

Figure 1. Results of search strategy

Patient Preference Methods

The 322 articles selected for this review were concerned with 379 preference methods (in some articles, multiple methods were described). In accordance with the classification provided by Carson and Louviere (10), 71 of these preference methods were identified as MM, 96 as DCEs, 200 as 'other methods' and 12 as MCDM. The four main categories of preference methods could be further subdivided into 15 conceptually different preference methods (Figure 2).

Matching Methods

Three conceptually different methods within MM were found in the literature. This review identified 47 studies where respondents were asked with a direct question (DQ) to state their WTP to obtain a certain health state. For example, Esfandiari et al. (19) assessed the effect of longterm financing on preferences regarding implant overdentures. Seven articles used the specific WTP method. SG, another variant of DQ, was found 34 times and is used mainly to quantify patient preferences regarding health states, not only for the probability of experiencing those health states but also for decision analysis, as shown by a study by Montgomery et al. (20). DQ is also operationalized using a binary response, instead of a matching variable (time, money), as shown in the study by Gyrd-Hansen and Kristiansen (21), where respondents were presented with

hypothetical therapy scenarios that involved life-year gains and were asked to state whether they should follow the therapy (yes/no).

Figure 2. Patient preference methods found in literature.

* Within the 'direct questions' category, seven articles used willingness to pay and 34 articles used a standard gamble

Another category in MM is the TTO approach, which was identified 19 times in this review. As with SG, TTO is also used to quantify the utilization of patient preferences regarding health states in decision analysis (22). The final MM category is the allocation game, in which respondents are asked to allocate a fixed quantity to different categories. In a study by Ubel et al. (23), respondents were asked how to allocate funds to benefit people with varying levels of disability. Within MM, single-question and sequential-question versions are possible. In most studies, the sequential versions were used in order to assess a value for various goods, i.e. the health states (24). Only three studies were identified using a single question, valuing one single health state. MM uses a direct weighting strategy; individuals are asked to directly provide their utilities for a health state. In MM, trade-offs can be made by trading health gains against several negative outcomes, such as the risk of immediate death (SG), a reduced lifespan (TTO), a monetary equivalent (WTP) or allocation trade-

offs (allocation game). MM is consistent with the expected utility theory, in that outcomes have cardinal properties and result in utilities.

Discrete Choice Experiments

Regarding DCE, four conceptually different methods were found in the literature: BCE, MCE, full ranking exercise and BWS. In this review, the majority (90 studies) used a BCE in their DCE methods. MCE was used only two times. It should be noted that some BCEs included an additional opt-out option, in order to allow respondents to be nondemanders (25). In the BCE category, 19 methods used the variant where one attribute of one of the treatment options is systematically varied until respondents switch their treatment preference to the alternative choice. This variant is used to determine treatment thresholds or switching points as, for example, in a study by Brundage et al. (26), where the treatment threshold represented the minimum survival percentage required by the participant in order to go for the more toxic treatment. One article in our review used a full ranking exercise. According to a study by Singh et al. (27), a full ranking approach is more realistic (i.e. it includes all hypothetical scenarios/products as in a real market); however, the disadvantage is that it yields only ordinal data. A less extensive ranking exercise is BWS, which can be used in several forms. Our review found two Fig. 1 Results of search strategy articles that used the BWS 'profile case', in which respondents judged each scenario at a time and selected the most attractive and the least attractive attribute. Also, one 'multi-profile case' was found where preferences were elicited by selecting the most preferred and least preferred scenarios out of a set of three or more alternatives (28, 29). In contrast to MM, DCEs are mostly used to estimate the marginal value of changing attributes of the health state to be judged through sequential questioning. No preferences regarding single stand-alone scenarios were found by our review. A unique aspect of DCE is that individuals are asked to provide their utilities for each attribute not directly but indirectly by determining relative preferences regarding choice scenarios. Sequential questioning and subsequent regression-based analysis give overall utility and individual part-worth utilities or an estimate of the marginal value of changing attributes.

Multi-Criteria Decision Methods

Three different MCDMs were found in this review: the analytic hierarchy process (AHP), multi-attribute utility theory (MAUT) and direct relative weighting (DRW). All MCDMs are compositional but differ in other characteristics. MAUT, an extension of the expected utility theory of multiple criteria, calculates the expected utility of an innovation by mapping criterion-specific

measurement scales onto a common scale of utility (30). The order of the importance of attributes is determined by presenting explicit trade-offs between gambles to the decision maker, resulting in direct assignment of weights to the decision criteria (overall utility and part-worth utility weights). Only four studies used MAUT, e.g. to model prostate cancer patients' preferences regarding health states (31). AHP does not assign weights directly but presents respondents with pairwise comparisons between criteria. A value scale that preserves the ratio information contained in the pairwise comparisons can then be estimated. Hummel et al. (32) used AHP to elicit patient preferences regarding multiple outcome measures to prioritize antidepressant drug treatment. This review identified seven studies using AHP. In DRW, respondents are asked to assign numbers to every decision characteristic being assessed. Summary scores (typically created using the weighted average method) of direct weighting indicate how well the alternative meets the goal. DRW does not exhibit explicit trading of characteristics, and only criteria weights are calculated. One study employed this preference method (33).

Other Methods

According to Carson and Louviere, this category covers (amongst others) methods that cannot give estimates consistent with utility theory (10). Furthermore, it includes methods that use self-explicated ratings. Those methods are problematic because the resulting data are ordinal but are sometimes used as interval. Methods such as adaptive questioning, rating, ranking, VAS and direct choice fall into this category. Rating, ranking and VAS are compositional methods and use a single question to directly elicit preferences regarding health states (34). A VAS is a sliding scale with anchored end points. In the review, 30 studies used a VAS to value health scenarios. They utilized 1–100 (35) or two statements (no information, full information) as anchored end points (36). Numerical rating scales are used to value health states on a scale ranging from not important (0) to very important (10) (37) or 0–4 (38). Rating scales were used in 93 studies. Ranking exercises (also known as comparative scales) are used to make relative judgments of preference by asking respondents to indirectly weight (rank) health states and put them in their preferred order based on their value. In a study by Guerlain et al. (39), participants were asked to rank the four products by moving them on the table into the preferred ranking order from left to right. The ranking method was used in 29 studies. Adaptive questioning is a combination of self-explicated ratings and BCE, and each question is based on the respondent's answer to previous questions. The indirect weighting strategy of adaptive questioning results in part-worth weights for each attribute. This review identified six studies using adaptive questioning. Direct choice is a binary

choice between two health states to determine which is valued highest. It is compositional, and the weights are drawn up directly. Our review identified 42 studies using direct choice questions in one form or another, ranging from questionnaires consisting only of direct choice questions (40) to studies using direct choice as an additional method—for example, with a VAS (41).

Preference Elicitation Use in Healthcare Contexts

The 322 identified articles were classified according to their intended use in three healthcare decision-making contexts. Four percent of the studies were classified as BRA ($n = 12$), 20 % were classified as HTA ($n = 68$), 40 % were classified as CDM ($n = 134$) and 36 % of the articles specified no intended context of use ($n = 120$). The intended use was not specified among the methods evenly. Eleven studies addressed two contexts of decision making, and these were counted twice. We classified one article as indicating both BRA and HTA, and ten articles indicated both CDM and HTA decision-making contexts (Table 1).

Table 1. *Intended use of preference methods in decision making contexts.*

	Benefit risk assessment	Health technology assessment	Clinical decision making	Decision context not specified
Matching methods	-	19	23	32
Discrete Choice Experiment	12	24	31	30
Multi-Criteria Decision Methods	-	3	3	7
Other methods*	-	36	98	74
Total articles in context	12 (1)**	68 (11)**	134 (10)**	120

* Includes visual analogue scales, direct choice, adaptive questioning, numerical rating and ranking.

**The numbers in parentheses indicate the articles which imply an intended use for two contexts and thus the methods were counted twice.

Benefit–Risk Assessment

All of the articles within BRA used DCEs to elicit preferences (Table 1). Patients' maximum acceptable risk (MAR) for each unit of benefit (42, 43) or the minimal accepted benefit (MAB) (44), were calculated from the preference data obtained by the DCEs. MAR and MAB provide decision makers with information about the payoff that patients require to accept risk or which minimal clinical benefit is required from a healthcare innovation to make it acceptable to patients. Studies designed for or aimed at the BRA context elicit preferences regarding health outcomes (Table 2).

Our literature review showed that process characteristics and non-health outcomes are often not taken into account in BRA.

Table 2. *Attribute valuation in decision making contexts.*

	Health outcomes	Non-health outcomes	Process characteristics	Total articles in context
Benefit risk assessment	12	-	3	12
Health technology assessment	51	28	41	68
Clinical decision making	103	48	44	134

Health Technology Assessment

In the reimbursement context, there is no clear preference for any one method (Table 1). Patient preferences in economic analysis are mainly measured with MM. Preferences measured with TTO and SG are often used in economic evaluations within the cost–utility framework; quality adjusted life-years (QALYs) are used in decision analysis. Examples of studies with economic evaluations of preference methods are those by Sanelli et al., Carroll and Downs, and Netten et al. (45-47). Understanding of patient and societal preferences and resource allocation decisions has gained importance because reimbursement decisions are increasingly based on economic evaluations (48). Our results show that various preference methods for priority setting of resource allocation have been used. The studies have often applied a resource allocation trade-off, which can assist the decision maker in making a choice between competitive fundings of health innovations. For example, Mason et al. (49) used BCE to elicit public preferences regarding prioritization of healthcare interventions in the UK, and Ratcliffe (50) employed rankings to elicit preferences from the recipients of donor liver grafts. Furthermore, patient preference methods are also used to compare patient preferences regarding competing therapies, e.g. two comparing chemotherapies for breast cancer (51) or two insulin types for diabetes (52). Preferences regarding certain process characteristics, such as (out-of-pocket) costs and mode of administration, are elicited more often in HTA than in BRA (Table 2). Process characteristics can have an impact on adherence to a therapy; various articles in this review state that because of non-adherence, potential clinical benefits are lost and that the costs rise per QALY, which is a relevant outcome for policy decision makers (53-55).

The Clinical Decision Context

Fifty percent of CDM articles had rating, ranking or VAS scales to measure patient preferences (Table 1). These scales are used to measure patient preferences to support decision making in the physician's office. For instance, in a study by Witticke et al. (56), rating scales were used by physicians to elicit patient preferences to specify medication-related factors of interest. Rating scales are also utilized as a value elicitation tool in decision aids if and when multiple outcomes of treatment are taken into account, as was done in a study by Fiks et al. (38). Methods such as SG and TTO have rarely been used in the physician's office (22, 57). Some articles have addressed the discrepancies in preferences between physicians and patients (58). The objective of these studies was to identify any discrepancies and to raise awareness among physicians about the relevance of preferences in decision making. Non-health outcomes and process characteristics are very important in the CDM context, as can be seen from the frequency with which they have been included in the preference instrument (Table 2). For instance, mode of administration and (out-of-pocket) costs directly influence decisions made by the patient, which can influence uptake or compliance (56, 59). Preferences regarding non-health outcomes, such as information and participation, are important because of their influence on the patient-physician relationship and how the treatment is delivered to the patient (36, 60, 61).

Comparison of Context Requirements and Method Characteristics

In this section, the requirements of each context as outlined in the 'Methods' section are compared with the methodological and practical characteristics of the methods currently used in the three decision-making contexts. Appendix C shows the scores of each preference method for previously defined methodological and practical characteristics.

The requirements of preference elicitation and their effect on CDM show a good fit with the actual use of preference methods. The choice of relatively simple methods (such as ranking, rating, VAS and direct choice) in the clinical decision context is explained by their ease of use, the low cognitive burden on the patient and the clinician, and the ease of estimating preferences from the value exercise. Ease of administration and fast access to results allow physicians or other healthcare professionals to utilize the results of the preference exercise in clinical practice. In BRA/HTA decisions, MM and DCEs are used the most, and these methods fulfil the theoretical and data collection and analysis requirements in these contexts. MM and DCEs assist in making the decision-making process transparent, legitimate and accountable. Both MM and DCEs have strong

methodological underpinnings in utility and economic theory. However, DCEs can be cognitively complex for respondents if the number of attributes is large or the attributes are difficult to interpret (probabilities). Similarly, SG and TTO are cognitively complex because they deal with abstractions of life and death, and they vary the life-years and health conditions by eliciting preferences. The compositional approach of MCDMs might be easier because preferences are elicited for each attribute at a time. However, MCDM does not provide utilities. MAUT (the only MCDM found by our review that provides utilities) is judged as being exhausting for the respondent because it demands a vast amount of preference information from patients (62).

DISCUSSION

The aim of this systematic review was to identify studies that used a preference elicitation method and to identify the healthcare decision context of the studies. Additionally, the context requirements for eliciting preferences in three separate decision contexts were identified and compared with the actual use of preferences.

In total, 379 methods were identified in 322 articles representing 15 preference elicitation methods. Four percent of the studies were classified as BRA, 20 % were classified as HTA, 40 % were classified as CDM and 36 % of the articles did not specify an intended context of use. Regarding clinical decisions, most preferences are elicited by means of straightforward methods such as rating, ranking, VAS and direct choice. Our findings show a good fit between the actual use of methods and the hypothesized demands for preferences in clinical decisions. Although preferences elicited with relatively simple methods do not result in exact utilities, utilities are not required in clinical decisions, where the results mainly need to inform on relative treatment preferences (e.g. for treatment A versus treatment B). In contrast, the methods used in BRA and HTA need to adhere to the leading paradigm of utilities in health economics (63) to ensure strong methodological underpinnings. For example, DCEs are less prone to bias than rating or VAS (64). MM and DCEs fulfil the theoretical requirements needed in HTA and BRA. However, both MM and DCE can be cognitively complex for the respondents, and this has been noted as a limitation in multiple studies (43, 65, 66).

Although MM is a recognized and long-standing method for preference elicitation, few MM methods were identified in this review. MM is generally used in QALY measurement for performing cost-effectiveness analysis from an HTA perspective, not explicitly to incorporate individual patient values. MCDMs (AHP, MAUT) were the least reported preference methods, even though they can

also be used to accommodate the weighting process of multiple conflicting objectives in the HTA context. MCDM could be used to weigh the importance of the various sources of evidence in the decision process. Hence, employing MCDM can result in greater transparency of the decision process and increased patient-centeredness. However, there is debate as to whether MCDM can be used as a true preference method or whether it would be better used as a decision-support tool (67, 68).

There are some limitations of this review, which influence the interpretation of its outcomes. First, there is a large variety of preference methods in the literature, and several categorizations of those preference methods exist. In this review, we chose to use the nomenclature of Carson and Louviere (10) to frame the preference methods. However, this is not a universally accepted nomenclature. Second, the methods were characterized according to their cognitive and practical difficulties. However, no publications that compared different methods on the basis of those characteristics were found. In any case, both characteristics depend on how they are used in specific studies. The cognitive complexity of BCE and MCE, for example, is significantly affected by an increase in the number of attributes (69). Therefore, a clear categorization depends on one's own evaluation and opinion. Third, as with any systematic review, the search terms and selection process influenced which studies were included. Earlier reviews differed with regard to their search terms or explicitly used the name of the preference method in the search terms. The current review found that only DCEs were used to inform BRA, but Hauber et al. (15) stated that benefit–risk preferences can also be measured through direct elicitation methods (rating scale, threshold technique and SG). However, in agreement with the results of this article, they also stated that DCEs are better tailored to estimate benefit–risk preference weights for multiple outcomes simultaneously.

Conclusion

We identified studies that actually only used their results in practice in the CDM context—for example, for development of a decision aid (33, 70, 71). In most articles, explicit or implicit statements made by the authors referred to just an intended use of the results. In the current state of the art, there is mainly a methodology push; no empirical studies of use of preferences to support decisions in BRA/HTA have been published. The importance of the patient's view when making clinical decisions was acknowledged a long time ago. Healthcare is increasingly concerned with the desires of patients, and patients are encouraged to participate in decision making. This awareness of patient-centeredness facilitates the use of patient preferences in the clinical context.

Preferences regarding health outcomes, and thus the value of outcomes to the patient and society, are also of increasing importance in other decision contexts. Currently, BRA and HTA agencies have established procedures for evaluating (new) health innovations. Inclusion of patient preferences would provide an extra source of evidence for the existing value dossier required for drug approval or reimbursement. Yet, this would require agencies to adapt their procedures to determine how patient preferences can best be measured and how they will fit into this process. Therefore, it is necessary to establish which method best fits the context demands. However, this review indicates a tension between the need to underpin methodological preferences and the cognitive and practical difficulties that come with those preference methods. No method exists that has a low cognitive burden and strong methodological underpinnings and that can, at the same time, deliver enough and adequate information to support the decision.

REFERENCES

1. Litva A, Coast J, Donovan J, Eyles J, Shepard M, Tacchi J, et al. 'The public is too subjective': public involvement at different levels of health-care decision making. *Soc Sci Med*. 2002;54(12):1825-37.
2. Coulter A. Patient engagement--what works? *J Ambul Care Manage*. 2012;35(2):80-9.
3. Smith M, Higgs J, Ellis E. Factors influencing clinical decision making. *Clinical reasoning in the health professions*. Oxford: Butterworth-Heinemann; 2008. p. 89-100.
4. Bridges J, Jones C. Patient-based health technology assessment: a vision of the future. *Int J Technol Assess Health Care*. 2007;23(1):30-5.
5. Gagnon M, Desmartis M, Lepage-Savary D, Gagnon J, St-Pierre M, Rhainds M, et al. Introducing patients' and the public's perspectives to health technology assessment: A systematic review of international experiences. *Int J Technol Assess Health Care*. 2011;27(01):31-42.
6. van Til J, IJzerman M. Why Should Regulators Consider Using Patient Preferences in Benefit-risk Assessment? *PharmacoEconomics*. 2014;32(1):1-4.
7. European Medical Agency. The patient's voice in the evaluation of medicines. How patient can contribute to assessment of benefit and risk. 2013.
8. Bridges J, Onukwugha E, Johnson F, Hauber A. Patient Preference Methods - A Patient Centered Evaluation Paradigm. *ISPOR Connections*. 2007;12(6):4-7.
9. Louviere JJ, Flynn TN, Carson RT. Discrete Choice Experiments Are Not Conjoint Analysis. *J Choice Model*. 2010;3(3):57-72.
10. Carson R, Louviere J. A Common Nomenclature for Stated Preference Elicitation Approaches. *Environ Resour Econ*. 2011;49(4):539-59.
11. Ali S, Ronaldson S. Ordinal preference elicitation methods in health economics and health services research: using discrete choice experiments and ranking methods. *Br Med Bull*. 2012;lds020.
12. Meißner M, Decker R. An empirical comparison of CBC and AHP for measuring consumer preferences. *International Symposium of Analytical Hierarchy Process*; Pittsburgh 2009.
13. Green P, Srinivasan V. Conjoint analysis in consumer research: issues and outlook. *J Consum Res*. 1978(5):103-23.
14. Louviere J. Conjoint analysis modelling of stated preferences: a review of theory, methods, recent developments and external validity. *J Transp Econ Policy*. 1988:93-119.
15. Hauber A, Fairchild A, Johnson F. Quantifying Benefit-Risk Preferences for Medical Interventions: An Overview of a Growing Empirical Literature. *Appl Health Econ Health Policy*. 2013;11(4):319-29.

16. Dolan J. Multi-Criteria clinical decision support. *Patient*. 2010;3(4):229-48.
17. Ryan M, Scott D, Reeves C, Bate A, van Teijlingen E. Eliciting public preferences for healthcare: a systematic review of techniques. *Health Technol Assess*. 2001;5(5):186.
18. Ryan M, Skåtun D, Major K. Using Discrete Choice Experiments to Go Beyond Clinical Outcomes when Evaluating Clinical Practice. In: Ryan M, Gerard K, Amaya-Amaya M, editors. *Using Discrete Choice Experiments to Value Health and Health Care. The Economics of Non-Market Goods and Resources*. 11: Springer Netherlands; 2008. p. 101-16.
19. Esfandiari S, Lund JP, Penrod JR, Savard A, Mark Thomason J, Feine JS. Implant overdentures for edentulous elders: Study of patient preference. *Gerodontology*. 2009;26(1):3-10.
20. Montgomery AA, Harding J, Fahey T. Shared decision making in hypertension: The impact of patient preferences on treatment choice. *Fam Pract*. 2001;18(3):309-13.
21. Gyrð-Hansen D, Kristiansen IS. Preferences for 'life-saving' programmes: Small for all or gambling for the prize? *Health Econ*. 2008;17(6):709-20.
22. Sommers B, Beard C, D'Amico A, Dahl D, Kaplan I, Richie J, et al. Decision analysis using individual patient preferences to determine optimal treatment for localized prostate cancer. *Cancer*. 2007;110(10):2210-7.
23. Ubel P, Spranca M, Dekay M, Hershey J. Public Preferences for Prevention versus Cure. *Med Decis Making*. 18:141-8.
24. Hallan S, Åsberg A, Indredavik B, Widerøe T. Quality of life after cerebrovascular stroke: a systematic study of patients' preferences for different functional outcomes. *J Internal Med*. 1999;246(3):309-16.
25. Marti J. Assessing preferences for improved smoking cessation medications: A discrete choice experiment. *Eur J Health Econ*. 2012;13(5):533-48.
26. Brundage MD, Davidson JR, Mackillop WJ, Feldman-Stewart DEB, Groome P. Using a treatment-tradeoff method to elicit preferences for the treatment of locally advanced non-small-cell lung cancer. *Med Decis Making*. 1998;18(3):256-67.
27. Singh J, Cuttler L, Shin M, Silvers JB, Neuhauser D. Medical Decision-Making and the Patient: Understanding Preference Patterns for Growth Hormone Therapy Using Conjoint Analysis. *Med Care*. 1998;36(8 SUPPL.):AS31-AS45.
28. Swancutt DR, Greenfield SM, Wilson S. Women's colposcopy experience and preferences: A mixed methods study. *BMC Women's Health*. 2008;8.
29. Al-Janabi H, Flynn TN, Coast J. Estimation of a preference-based carer experience scale. *Med Decis Making*. 2011;31(3):458-68.
30. Keeney RL, Raiffa H. *Decisions with Multiple Objectives: Preferences and Value Trade-Offs*: Cambridge University Press; 1993.
31. Chapman GB, Elstein AS, Kuzel TM, Nadler RB, Sharifi R, Bennett CL, et al. A multi-attribute model of prostate cancer patients' preferences for health states. *Qual Life Res*. 1999;8(3):171-80.
32. Hummel MJM, Volz F, Van Manen JG, Danner M, Dintsios CM, Ijzerman MJ, et al. Using the analytic hierarchy process to elicit patient preferences: Prioritizing multiple outcome measures of antidepressant drug treatment. *Patient*. 2012;5(4):225-37.
33. Cunich M, Salkeld G, Dowie J, Henderson J, Bayram C, Britt H, et al. Integrating evidence and individual preferences using a web-based multi-criteria decision analytic tool: An application to prostate cancer screening. *Patient*. 2011;4(3):153-62.
34. McCormack HM, Horne DJ, Sheather S. Clinical applications of visual analogue scales: a critical review. *Psychol Med*. 1988;18(4):1007-19.
35. Heckerling PS, Verp MS, Albert N. Patient or physician preferences for decision analysis: The prenatal genetic testing decision. *Med Decis Making*. 1999;19(1):66-77.
36. Davis M, Hoffman J, Hsu J. Impact of patient acuity on preference for information and autonomy in decision making. *Acad Emerg Med*. 1999;6(8):781-5.
37. Ruland C. Decision Support for Patient Preference-based Care Planning Effects on Nursing Care and Patient Outcomes. *J Am Med Inform Assoc*. 1999;6(4):304-12.
38. Fiks A, Mayne S, Hughes C, DeBartolo E, Behrens C, Guevara J, et al. Development of an Instrument to Measure Parents' Preferences and Goals for the Treatment of Attention Deficit-Hyperactivity Disorder. *Acad Pediatr*. 2012;12(5):445-55.

39. Guerlain S, Hugine A, Wang L. A comparison of 4 epinephrine autoinjector delivery systems: usability and patient preference. *Ann Allerg Asthma Im.* 2010;104(2):172-7.
40. Fagerlin A, Zikmund-Fisher B, Ubel P. Cure me even if it kills me: preferences for invasive cancer treatment. *Med Decis Making.* 2005;25(6):614-9.
41. De Boer A, Stalmeier P, Sprangers M, De Haes J, Van Sandick J, Hulscher J, et al. Transhiatal vs extended transthoracic resection in oesophageal carcinoma: patients' utilities and treatment preferences. *Br J Cancer.* 2002;86(6):851-7.
42. Johnson FR, Hauber AB, Ozdemir S, Lynd L. Quantifying women's stated benefit-risk trade-off preferences for IBS treatment outcomes. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2010;13(4):418-23.
43. Bewtra M, Kilambi V, Fairchild AO, Siegel CA, Lewis JD, Johnson FR. Patient preferences for surgical versus medical therapy for ulcerative colitis. *Inflamm Bowel Dis.* 2014;20(1):103-14.
44. Hauber AB, Johnson FR, Grotzinger KM, Ozdemir S. Patients' benefit-risk preferences for chronic idiopathic thrombocytopenic purpura therapies. *Ann Pharmacother.* 2010;44(3):479-88.
45. Sanelli PC, Gold RL, Greenberg ED, Reichman MB, Ugorec I, Segal AZ, et al. Work-in-progress toward incorporating patients' preferences in practice guidelines for imaging aneurysmal subarachnoid hemorrhage. *Acad Radiol.* 2009;16(5):535-40.
46. Carroll AE, Downs SM. Improving decision analyses: parent preferences (utility values) for pediatric health outcomes. *J Pediatr.* 2009;155(1):21-5.
47. Netten A, Burge P, Malley J, Potoglou D, Towers AM, Brazier J, et al. Outcomes of social care for adults: developing a preference-weighted measure. *Health Technol Assess.* 2012;16(16):1-166.
48. Desser AS, Olsen JA, Grepperud S. Eliciting preferences for prioritizing treatment of rare diseases: The role of opportunity costs and framing effects. *Pharmacoeconomics.* 2013;31(11):1051-61.
49. Mason H, Baker R, Donaldson C. Understanding public preferences for prioritizing health care interventions in England: Does the type of health gain matter? *J Health Serv Res Policy.* 2011;16(2):81-9.
50. Ratcliffe J. Public preferences for the allocation of donor liver grafts for transplantation. *Health Econ.* 2000;9(2):137-48.
51. Beusterien K, Grinspan J, Tencer T, Brufsky A, Visovsky C. Patient preferences for chemotherapies used in breast cancer. *Int J Womens Health.* 2012;4:279-87.
52. Aristides M, Chen J, Schulz M, Williamson E, Clarke S, Grant K. Conjoint analysis of a new Chemotherapy: willingness to pay and preference for the features of raltitrexed versus standard therapy in advanced Colorectal Cancer. *Pharmacoeconomics.* 2002;20(11):775-84.
53. Hiligsmann M, Kanis JA, Compston J, Cooper C, Flamion B, Bergmann P, et al. Health technology assessment in osteoporosis. *Calcified Tissue Int.* 2013;93(1):1-14.
54. Laba TL, Brien JA, Fransen M, Jan S. Patient preferences for adherence to treatment for osteoarthritis: the MEdication Decisions in Osteoarthritis Study (MEDOS). *BMC Musculoskeletal Disord.* 2013;14(160):1471-2474.
55. Benjamin L, Cotte FE, Philippe C, Mercier F, Bachelot T, Vidal-Trecan G. Physicians' preferences for prescribing oral and intravenous anticancer drugs: a Discrete Choice Experiment. *Eur J Cancer.* 2012;48(6):912-20.
56. Witticke D, Seidling H, Klimm H, Haefeli W. Do we prescribe what patients prefer? Pilot study to assess patient preferences for medication regimen characteristics. *Patient Prefer Adher.* 2012;6.
57. Rhodes E, Prosser L, Lieu T, Songer T, Ludwig D, Laffel L. Preferences for type 2 diabetes health states among adolescents with or at risk of type 2 diabetes mellitus. *Pediatr diabetes.* 2011;12(8):724-32.
58. Johnson F, Hauber B, Ozdemir S, Siegel C, Hass S, Sands B. Are gastroenterologists less tolerant of treatment risks than patients? Benefit-risk preferences in Crohn's disease management. *J Manag Care Pharm.* 2010;16(8):616-28.
59. Vermeulen H, Ubbink D, De Zwart F, Goossens A, De Vos R. Preferences of patients, doctors, and nurses regarding wound dressing characteristics: a conjoint analysis. *Wound Repair Regen.* 2007;15(3):302-7.
60. Pardon K, Deschepper R, Stichele RV, Bernheim J, Mortier F, Deliens L. Preferences of advanced lung cancer patients for patient-centred information and decision-making: A prospective

multicentre study in 13 hospitals in Belgium. *Patient Educ Couns*. 2009;77(3):421-9.

61. Pardon K, Deschepper R, Vander R, Bernheim J, Mortier F, Schallier D, et al. Are patients' preferences for information and participation in medical decision-making being met? Interview study with lung cancer patients. *Palliative Med*. 2011;25(1):62-70.

62. Troche CJ, Paltiel AD, Makuch RW. Evaluation of Therapeutic Strategies: A New Method for Balancing Risk and Benefit. *Value Health*. 2000;3(1):12-22.

63. Smith R, Lorgelly P, Al-Janabi H, Venkatapuram S, Coast J. The capability approach: an alternative evaluation paradigm for health economics? The elgar companion to health economics. Cheltenham: Edward Elgar Publishing; 2012. p. 415.

64. Torrance G, Feeny D, Furlong W. Visual Analog Scales Do They Have a Role in the Measurement of Preferences for Health States? *Med Decis Mak*. 2001;21(4):329-34.

65. Johnson F, Ozdemir S, Mansfield C, Hass S, Miller D, Siegel C, et al. Crohn's disease patients' risk-benefit preferences: serious adverse event risks versus treatment efficacy. *Gastroenterology*. 2007;133(3):769-79.

66. Johnson F, Van Houtven G, Ozdemir S, Hass S, White J, Francis G, et al. Multiple sclerosis patients' benefit-risk preferences: serious adverse event risks versus treatment efficacy. *J Neurol*. 2009;256(4):554-62.

67. European Medicines Agency. Benefit-risk methodology project 2010 [cited 2013 10-29]. Available from: http://www.ema.europa.eu/ema/index.jsp?curl=pages/special_topics/document_listing/document_listing_000314.jsp.

68. Diaby V, Goeree R. How to use multi-criteria decision analysis methods for reimbursement decision-making in healthcare: A step-by-step guide. *Expert Rev Pharmacoecon Outcomes Res*. 2014;14(1):81-99.

69. Johnson FR, Lancsar E, Marshall D, Kilambi V, Mühlbacher A, Regier DA, et al. Constructing experimental designs for discrete-choice experiments: Report of the ISPOR conjoint analysis experimental design good research practices task force. *Value Health*. 2013;16(1):3-13.

70. Goel V, Sawka C, Thiel E, Gort E, O'Connor A. Randomized trial of a patient decision aid for choice of surgical treatment for breast cancer. *Med Decis Mak*. 2001;21(1):1-6.

71. Ruland C. Decision Support for Patient Preference-based Care Planning Effects on Nursing Care and Patient Outcomes. *J of the American Med Informatics Association*. 1999;6(4):304-12.

APPENDIX A: SEARCH STRATEGY

Details of the strategies employed for searching Scopus and ISI Web of Science databases in January/February 2014.

Table 1. Search strategy for Scopus

#N	Searches	Results
#1	TITLE=((preference*) OR (patient W/3 value*) OR (patient W/3 perspective*) OR (patient W/3 choice*))	49.956
#2	#1 AND TITLE-ABS-KEY=("health technology assessment" OR reimbursement)	100
#3	#1 AND TITLE-ABS-KEY=("shared decision making" OR "clinical decision making")	281
#4	#1 AND TITLE-ABS-KEY=("benefit risk" OR "risk benefit")	642
#5	#1 AND TITLE-ABS-KEY=("healthcare decision making" OR "medical decision making")	184
#6	#2 OR #3 OR #4 OR #5	1137
#7	#6 AND (LIMIT-TO(DOCTYPE, "ar")) AND (LIMIT-TO(LANGUAGE, "English"))	809

Table 2. Search strategy for ISI Web Of Science

#N	Searches	Results
#1	TITLE=((preference*) OR (patient NEAR/3 value*) OR (patient NEAR/3 perspective*) OR (patient NEAR/3 choice*))	Approximately 76.750
#2	#1 AND TOPIC=("health technology assessment" OR reimbursement)	127
#3	#1 AND TOPIC=("shared decision making" OR "clinical decision making")	487
#4	#1 AND TOPIC=("benefit risk" OR "risk benefit")	74
#5	#1 AND TOPIC=("healthcare decision making" OR "medical decision making")	264
#6	#2 OR #3 OR #4 OR #5	596
#7	#6 refined by: Document Type=(ARTICLE) AND Languages=(ENGLISH)	493

APPENDIX C: SCORES OF PREFERENCE METHODS ON SEVERAL CRITERIA.

This appendix shows the scores of each explicit preference method on “cognitive effort on the part of the respondent”, “costs of data collection”, “skills required in data analysis and interpretation”, and “outcome measures”.

List of abbreviations

WTP: Willingness To Pay

SG: Standard Gamble

TTO: Time Trade-Off

MCE: Multinomial Choice Experiment

BCE: Binary Choice Experiment

BWS (2): Best-Worst Scaling Case 2

BWS (3): Best-Worst Scaling Case 3

VAS: Visual Analogue Scales

MAUT: Multi-Attribute Utility Theory

DRW: Direct Relative Weighting

AHP: Analytical Hierarchy Process

Matching Methods				
	Direct Question	SG	Allocation Game	TTO
	WTP	SG		
Explicit trade-off	No	Yes (health)	Yes (characteristics)	Yes (health)
Attributes*	1,2,3	1,2,3	1,2,3	1,2,3
Cognitive effort on the part of the respondent	High [1, 2]	High [3, 4]	Medium	High [3, 4]
Costs of data collection	High	High [1, 4, 6]	Medium	High [4, 7]
Skills required in data analysis and interpretation	Intermediate	Basic	Basic	Basic
Outcome measures	Overall Value	Overall Utility	Overall Value	Overall Utility
Multi-Criteria Decision Methods				
	AHP	MAUT		DRW
Explicit trade-off	No	Yes (characteristics)		No
Attributes*	1,2,3	1,2,3		1,2,3
Cognitive effort on the part of the respondent	Medium [8]	High [9]		Low
Costs of data collection	Medium [8]	High		Low
Skills required in data analysis and interpretation	Basic [8]	Basic		Basic
Outcome measures	Overall Value; Criteria weights	Overall Utility; Part-Worth Utility		Criteria weights

* 1=health, 2 = non-health, 3 = process

Discrete Choice Experiment				
	BCE	MCE	Full Ranking	Best-Worst Choices
				BWS2 BWS3
Yes (characteristics)	Yes (characteristics)	Yes (characteristics)	Yes (characteristics)	Yes (characteristics)
1,2,3	1,2,3	1,2,3	1,2,3	1,2,3
Medium	Medium	Medium	Medium	Low [5] Medium [5]
Medium	Medium	Medium	Medium	Medium
Intermediate - Advanced	Intermediate - Advanced	Intermediate - Advanced	Intermediate - Advanced	Intermediate
Overall Utility; Part-Worth Utility	Overall Utility; Part-Worth Utility	Overall Utility; Part-Worth Utility	Overall Utility; Part-Worth Utility	Overall Utility; Part-Worth Utility
Other Methods				
	Adaptive questioning	Rating	Ranking	VAS
				Direct Choice
Yes (characteristics)	No	No	No	No
1,2,3	1,2,3	1,2,3	1,2,3	1,2,3
Medium	Low [3]	Low	Low [2]	Low
Medium	Low	Low	Low [1, 4]	Low
Intermediate - Advanced	Basic	Basic	Basic	Basic
Overall Value; Criteria weights	Criteria Rating	Criteria Ranking	Overall Preference	Overall Preference

REFERENCES

1. Ryan, M., et al., Eliciting public preferences for healthcare: a systematic review of techniques. *Health Technol Assess.* 2001. 5(5): p. 186.
2. Franic, D.M. and D.S. Pathak, Assessment of respondent acceptability of preference measures: Discriminatory power of graphic positioning scale versus traditional scaling measures. *Value Health.* 2003. 6(4): p. 483-493.
3. Craig, B.M., J.J.V. Busschbach, and J.A. Salomon, Keep it simple: Ranking health states yields values similar to cardinal measurement approaches. *J Clin Epidemiol.* 2009. 62(3): p. 296-305.
4. Stiggelbout, A.M. and J. De Haes, Patient preference for cancer therapy: an overview of measurement approaches. *J Clin Oncol.* 2001. 19(1): p. 220-230.
5. Flynn, T.N., et al., Best-worst scaling: what it can do for health care research and how to do it. *J Health Econ.* 2007. 26(1): p. 171-189.
6. Froberg, D.G. and R.L. Kane, Methodology for measuring health-state preferences—II: Scaling methods. *J Clin Epidemiol.* 1989. 42(5): p. 459-471.
7. Torrance, G.W., Social preferences for health states: An empirical evaluation of three measurement techniques. *Socioecon Plann Sci.* 1976. 10(3): p. 129-136.
8. IJzerman, M.J., J.A. van Til, and G.J. Snoek, Comparison of two multi-criteria decision techniques for eliciting treatment preferences in people with neurological disorders. *Patient.* 2008. 1(4): p. 265-272.
9. Dolan, J.G., Multi-Criteria clinical decision support. *Patient.* 2010. 3(4): p. 229-248.

3

Involving Patients in Weighting Benefits and Harms of Treatment in Parkinson's Disease

Marieke G.M Weernink
Janine A. van Til
Jeroen P.P. van Vugt
Kris L.L. Movig
Catharina G.M. Groothuis-Oudshoorn
Maarten J. IJzerman

PloS one. 2016;11(8):e0160771. doi:10.1371/journal.pone.016077

ABSTRACT

Introduction: Little is known about how patients weigh benefits and harms of available treatments for Parkinson's Disease (oral medication, deep brain stimulation, infusion therapy). In this study we have (1) elicited patient preferences for benefits, side effects and process characteristics of treatments and (2) measured patients' preferred and perceived involvement in decision-making about treatment.

Methods: Preferences were elicited using a best-worst scaling case 2 experiment. Attributes were selected based on 18 patient-interviews: treatment modality, tremor, slowness of movement, posture and balance problems, drowsiness, dizziness, and dyskinesia. Subsequently, a questionnaire was distributed in which patients were asked to indicate the most and least desirable attribute in nine possible treatment scenarios. Conditional logistic analysis and latent class analysis were used to estimate preference weights and identify subgroups. Patients also indicated their preferred and perceived degree of involvement in treatment decision-making (ranging from active to collaborative to passive).

Results: Two preference patterns were found in the patient sample (N=192). One class of patients focused largely on optimising the process of care, while the other class focused more on controlling motor-symptoms. Patients who had experienced advanced treatments, had a shorter disease duration, or were still employed were more likely to belong to the latter class. For both classes, the benefits of treatment were more influential than the described side effects. Furthermore, many patients (45%) preferred to take the lead in treatment decisions, however 10.8% perceived a more passive or collaborative role instead.

Discussion: Patients weighted the benefits and side effects of treatment differently, indicating there is no "one-size-fits-all" approach to choosing treatments. Moreover, many patients preferred an active role in decision-making about treatment. Both results stress the need for physicians to know what is important to patients and to share treatment decisions to ensure that patients receive the treatment that aligns with their preferences.

INTRODUCTION

Parkinson's disease (PD) is a progressive and degenerative disorder which causes tremors and difficulties with walking, movement, and coordination. Treatment is aimed at maintaining continuous relief of motor symptoms and improving the patient's functional capacity (1). However, a significant number of patients with PD are still undertreated, e.g. because of suboptimal adherence to drug regimens, "levodopa phobia", or fear of deep brain stimulation (2, 3). For example, Leopold et al. (2004) have shown that only 10% of patients with PD fully adhere to drug regimens as intended (4). Besides treatment effectiveness, other attributes such as ease of use, routines, and treatment modality might influence patients' decisions about treatment adherence or advanced treatments (subcutaneous or intraduodenal pump infusion or neurosurgery) (5-9).

Physicians have already begun to pay increased attention to patient-reported outcomes and started to recognise quality of life as the primary treatment goal (10, 11). However, to further optimise treatment outcomes – especially outcomes that matter to the patient – physicians need to better understand how patients weigh up the benefits, side effects and process characteristics of treatment. Although it is a considerable challenge to find ways to elicit patients' preferences, the explicit weighing of treatment characteristics is essential to gather information about the relative desirability of treatment outcomes and modalities from the patient's point of view (12). Since it is often not possible to reduce all symptoms and side effects for a patient with PD, it is important to know whether a patient will, for example, accept the risk of suffering from incidental bouts of dizziness in order to obtain improved motor function.

Patients may weigh these benefits and harms differently to physicians which could influence the treatment patients end up with, depending on who is leading the choice of treatment (13). Patient involvement is essential in deciding which treatment is best tailored to the individual patient's needs (14-16). Moreover, patients are also more likely to be compliant and follow a treatment regimen if they experience greater involvement in the decision-making process (17).

Although patient-centred care is considered important in PD care, little is known about the degree of patient involvement in treatment decisions and the congruence with patients' preferred involvement. Moreover, little is known about the actual trade-offs that patients would prefer to make between the benefits and side effects of treatments. In this study, we aim to elicit patient preferences around motor symptoms, side effects, and factors related to the delivery of care in the main treatments in PD (oral intake of medication, continuous pump infusion of medication, and

neurosurgery). A second objective was to assess the congruence between patients' preferred and perceived involvement in decision-making about treatment.

METHODS

In this study, the best-worst scaling case 2 method was used to elicit patient preferences about the relative desirability of treatment outcomes (18, 19). In best-worst scaling respondents are presented with possible treatment scenarios, described according to their characteristics, and patients have to choose the best and worst features of possible treatments. In the next paragraph the way in which these treatment scenarios were created and presented will be explained.

Best-Worst Scaling: Creating Treatment Scenarios

In a best-worst scaling experiment, treatment scenarios are described by an underlying, basic set of characteristics of care (called attributes) and each attribute is represented by two or more values (called levels). In this study, 18 interviews with patients (qualitative research) provided the basis for identifying the full set of attributes that characterise PD treatments and influence patients' Health Related Quality of Life (HRQoL). The results of this qualitative research are summarised below, but a more extensive report can be found in Appendix A.

Patients for this stage were recruited from the hospital Medisch Spectrum Twente (all 18 patients took oral medication, three patients have had neurosurgery in the past, and one patient received medication via a pump). Patients were first asked to describe their health using the Parkinson-Disease Questionnaire-39 (PDQ-39) (20) and the EuroQoL-5D-5L (EQ5D-5L) (21). The interviewer used this information to structure the interviews and to decide which health domains were important to discuss (semi-structured interviews). Next, patients were questioned about the impact of symptoms, side effects, and process characteristics on their HRQoL. The interviews were tape recorded and transcribed verbatim and coded and analysed in ATLAS.ti (22). The final result was a table ranking the frequency of symptoms and side effects (attributes) and their impact on HRQoL according to these patients. Yet best-worst scaling is typically limited to a small number of attributes (four to eight). Discussion with a project team consisting of a neurologist, a hospital pharmacist, a rehabilitation specialist, three health sciences researchers, and two patients was used to narrow the list of attributes. The task of the project team was to select a balanced set of symptoms, side effects, and process characteristics based on the qualitative and quantitative

importance of these attributes of treatment in patients' daily life (interview results), and to make sure that the selected attributes were typical symptoms and side effects of PD and not too general.

After deliberation, one process attribute (treatment modality), three motor symptoms (tremor, slowness of movement, and posture and balance problems) and three side effects (dizziness, drowsiness, dyskinesia) were selected for inclusion in the preference task. The selection of the process attribute reflected interviewed patients expressing concerns regarding the impact of surgery and pump procedures on their daily life, despite the expected reduction in symptoms. This highlights an important trade-off in expressing preference for treatment. For motor symptoms, the project team followed the interview results: tremor, slowness of movement, and posture and balance problems were the problems most frequently reported and all had a major impact on patient's HRQoL. Less common symptoms were: problems with writing, crying, drooling, swallowing, loss of smell, and constipation. To keep a balance between the negative and positive effects of treatments, three side effects were also selected. The interview results showed that dyskinesia occurred the most often and had the largest impact on patients' daily life, because of the duration, the unpredictable character, and the obstruction of daily tasks. Secondly, side effects were most often reported in the sleeping domain and the selected attribute 'drowsiness' was defined as extensive daytime sleepiness. Nausea, stomach pain, vomiting, diarrhoea, and constipation also occurred frequently, but the project team concluded that these side effects were too general. Lastly, dizziness (lightheadedness caused by orthostatic hypotension) was selected instead of 'hallucinations and paranoia'. Although the latter had a larger impact on HRQoL, it was only reported by three patients.

The next step was to describe the variation in possible outcomes for each attribute. In order to reduce the cognitive difficulty, three qualitative levels were chosen to represent the burden of symptoms and side effects in the treatment scenarios (seldom to never, sometimes, and often suffer from). The attribute treatment modality was described as the oral intake of tablets, continuous pump infusion of medication, and neurosurgery. The attribute-levels were systematically combined into treatment profiles and in theory 2187 (3^7) hypothetical treatment scenarios were possible. Since it is impossible to ask patients to answer best-worst questions for each of these scenarios, experimental design software was used to select the smallest subset that identified all necessary parameters (23). Experimental design software uses algorithms to construct D-optimal designs to approximate, so that each attribute-level appears the same number of times and each attribute-level appears an equal number times with another attribute-level (no

correlation). Our final subset consisted of 36 profiles, which we divided over four versions of the questionnaire (nine per version). For each hypothetical treatment profile, patients were asked to indicate which aspect of treatment they found most and least desirable (Fig 1). Thus patients were asked to make a trade-off between treatment modality, symptoms, and side effects of treatments. By offering several of these scenarios to multiple patients, the data can be used to predict the relative desirability of treatment characteristics from the patients' point of view (as a group). The best-worst scaling questions were introduced by a detailed explanation and a clear example. Respondents were provided with information sheets about the advanced treatments.

Figure 1. Example of a treatment profile in which the patient had to indicate the least and most desirable characteristic of treatment.

Most desirable characteristic	What do you find the...	Least desirable characteristic
Treatment profile 1
☐	I receive medication via tablets	☐
☐	I often suffer from tremors	☒
☒	I seldom to never suffer from posture and balance problems	☐
☐	I seldom to never suffer from slowness in motion	☐
☐	I often suffer from dizziness	☐
☐	I sometimes suffer from drowsiness	☐
☐	I often suffer from rapid uncontrolled movements	☐

Control Preference Scale

In the second part of the questionnaire, patients were asked about their perceived and preferred involvement in the choice of treatment. A modified version of the validated Control Preferences Scale was used to determine the level of agreement between the preferred and perceived decision role of the patient (24, 25) (Table 1). Since preferred and perceived decision roles were operationalised differently, the measurement of agreement also included patients who experienced only a slight difference in decision roles (e.g. patients who said they preferred an active shared decision role were included if they picked either the active or collaborative decision

role as their perceived decision role). In addition, it was decided to only measure extreme discordance between decision roles. The validated Control Preferences Scale does not study the role of the caregiver or partner. Therefore, we questioned (in addition to benefits and harms) to what extent advice from family and friends and burden of treatment for partner/caregiver influenced decision-making (5-point rating scale: no influence to very influential). Lastly, patients were asked for socio-demographics, quality of life (PDQ-8 (26) and EQ5D-5L (21)), PD-related medication, and current/previous treatments.

Table 1. Questions and answer-categories used to assess patient's preferred and perceived decision role in treatment decision making.

Preferred Decision Role To what extent do you want to be involved in the treatment decision?	Perceived Decision Role To what extent were you involved in the previously made treatment decision?
<ul style="list-style-type: none"> • I prefer to make the final treatment decision • I prefer to make the final treatment decision after seriously considering my doctor's opinion • I prefer that my doctor and I share responsibility for deciding which treatment is best • I prefer that my doctor makes the final treatment decision, but seriously considers my opinion • I prefer to leave all treatment decisions to my doctor 	<ul style="list-style-type: none"> • Treatment was largely chosen by myself • Treatment was chosen in collaboration with the doctor • Treatment was largely chosen by the doctor

Study Population

A paper version of the questionnaire was sent to all patients diagnosed with idiopathic PD at the outpatient department of Medisch Spectrum Twente (except for patients who were registered as having atypical parkinsonism or dementia). Medisch Spectrum Twente is a large teaching hospital in the Netherlands that provides all three advanced therapies in PD (deep brain stimulation, subcutaneous apomorphine infusion and intraduodenal levodopa infusion). The data collection was expanded to include a non-hospital based patient sample by recruiting respondents through online Parkinson communities, the website and paper magazine of the Dutch Parkinson's Disease Society and PD forums in both the Netherlands and Belgium. It could not be verified whether these participants were genuinely diagnosed with idiopathic PD or had atypical parkinsonism or dementia, because there was no access to hospital records. These patients were invited to complete the online survey or to request the paper version. According to the Medical-Ethics

Twente Committee, this study did not have to be assessed against the medical research involving human subjects legislation. According to Dutch law, written informed consent was not required; by returning the completed questionnaire patients agreed to the use of anonymised data for the research purposes of this study.

Data Analysis

First, patients with incomplete data on both main parts (Control Preference Scale and best-worst scaling questions) were excluded. Subsequently, the data from the hospital and non-hospital patient samples were merged and socio-demographic, health outcomes, and treatment characteristics were described using frequencies, percentages, means, and standard deviations (SD). ticked multiple best/worst boxes in one scenario, chose only the most desirable characteristics, or left multiple questions blank. For the best-worst scaling analysis, patients who had multiple response errors in their paper-questionnaire (ticked multiple best/worst boxes in one scenario, chose only the most desirable characteristics, or left multiple questions blank) were excluded. The online submitted surveys contained no missing data nor response errors, because error messages were presented during completion. Next, a conditional logit model estimated part-worth utilities for each attribute-level. Due to a latent scale, part-worth utilities cannot be directly compared. Importance weights were therefore calculated, based on the difference between minimum and maximum part-worth utilities within an attribute. The largest difference value received an importance weight of one, representing the attribute which had the highest impact on treatment desirability and the other difference values were divided by the largest difference value, resulting in a relative distance of all attributes to the attribute with the highest impact. Latent Class analysis was conducted to explore heterogeneity in preferences. Latent class analysis sorts patients into classes based on similarity in preferences and then identifies patient characteristics that are significantly associated with class membership. The model provides information about the likelihood of patients falling into as specific class on the basis of their characteristics (27). The number of classes was pre-specified by fit indices (Bayesian and Akaike Information Criterion BIC/AIC) which measure the quality of statistical models. In this study, the AIC indicated a three-class model while the BIC indicated a two-class model. A two-class model was selected because using three classes did not reveal an additional interpretable subgroup. To determine whether certain patients were more likely to belong to one of the classes, several potential predictors were entered and tested for their significance (socio-demographics, health outcomes, experience with symptoms/side effects, treatment characteristics and quality of life measures). Finally, for the

Control Preference Scale questions a chi-square test was used to evaluate the level of agreement between the perceived and preferred decision role of patients in decision-making about treatment (2-sided, p -value < 0.05). All data was analysed using Stata version 14 (StataCorp, College Station, TX).

RESULTS

132 patients from the Medisch Spectrum Twente sample returned their paper-questionnaire (response rate: 48%). Eight questionnaires were returned with incomplete data on all main parts of the questionnaire and were not used for further analysis. The non-hospital based patient sample consisted of 105 complete surveys online, with no paper versions requested. Most questionnaires (80%) were completed by patients independently, only 20% of patients were assisted by their partner or caregiver.

Patient Sample

Background characteristics and health care data for the total patient sample (N=229) are summarised in Table 2. The sample consisted predominantly of men (66.5%), with a mean age of 65.4 years (SD 10.0) and a mean disease duration of 7.6 years (SD 6.9). Only 15% of patients were still employed. Concerning the clinical data, most patients faced mild to moderate impairments in motor symptoms: tremors (59.6%), slowness of movement (85.0%), and posture and balance problems (78.8%). About 81.3% sometimes or often suffered from drowsiness, 46.2% from dizziness (orthostatic hypotension), and 44.4% from dyskinesia. Most patients indicated moderate impairments in quality of life (PDQ-8: 31.0 (16.4), EQ5D-5L: 0.69 (0.17)), but 73% of patients were still able to live independently or only required help from their partner. However, almost three-quarters of patients reported one or more co-morbidities. Almost all patients took oral medication (96.1%), and advanced treatments were seen in only 17.5% of patients, with about 13.5% having had neurosurgery in the past and 4% currently wearing a pump. Apart from the non-hospital based sample being significantly younger and more often employed, both patient samples were comparable and they are therefore merged in Table 2.

Table 2. Background, socio-demographic and clinical characteristics (N=229).

Variables	N (%) or mean $\pm$ SD (min, max)
Background Characteristics & Socio-Demographics	
Sample	
Medisch Spectrum Twente (hospital based sample)	124 (54.1)
Non-hospital based sample	105 (45.9)
Gender	
Man	151 (66.5)
Woman	76 (33.5)
Marital state	
Single	31 (13.7)
With partner (no children living at home)	152 (81.0)
With partner and children living at home	34 (15.0)
Other	9 (4.0)
Age (years)	65.4 $\pm$ 10.0 (33, 88)
Education level*	
Low	84 (37.0)
Medium	65 (28.6)
High	78 (34.4)
Employment status	
Employed (full or part-time)	34 (14.8)
Disabled / unable to work	52 (22.7)
Retired	135 (59.0)
House maker or housewife	8 (3.5)
Questionnaire completion by	
Patient	182 (80.5)
Patient and partner	33 (14.6)
Patient and caregiver (other than partner)	11 (4.8)
Health Outcomes and Treatment Characteristics	
Disease duration (years)	7.6 $\pm$ 6.9 (0.25, 31)
General health experience	
(Very) good	69 (30.2)
Moderate	135 (59.2)
(Very) poor	27 (10.6)
Current treatment*	
Oral medication	220 (96.1)
Levodopa monotherapy	90 (39.3)
Dopamine agonist monotherapy	23 (10.0)
Combined levodopa/dopamine agonist treatment	107 (46.7)
Continuous pump infusion (subcutaneous)	2 (0.9)
Continuous pump infusion (intraduodenal)	7 (3.1)
Neurosurgery (one-sided)	7 (3.1)
Neurosurgery (two-sided)	24 (10.6)
Severity of motor symptoms	
No visible symptoms of PD	30 (13.3)
Symptoms are one-sided	113 (50.0)
Symptoms are two-sided	83 (36.7)
Suffering from motor symptoms	
Tremor	136 (59.6)
Posture and balance problems	180 (78.8)
Slowness of movement	195 (85.0)

Table 2. Continued

Variables	N (%) or mean $\pm$ SD (min, max)
Suffering from side effects	
Dizziness	106 (46.2)
Drowsiness	186 (81.3)
Dyskinesia	101 (44.4)
Independency	
Independent in self-care tasks	70 (30.7)
Assistance from family member (partner)	96 (42.1)
Home health care attendant	53 (23.2)
Living in assisted living facility	7 (3.1)
Other	2 (0.9)
Patients with co-morbidities*	143 (74.5)
Hypertension	38 (19.8)
Heart disease	36 (18.8)
Sleep disorder	36 (18.8)
Arthritis	24 (12.5)
Cancer	10 (5.2)
Diabetes	7 (3.6)
Other	41 (21.4)
Felt depressed**	15 (6.7)
Had concentration problems**	45 (19.8)
Was unable to communicate properly**	36 (15.8)
Quality of Life Measures	
PDQ-8 Summary Index Score (0-100)	31.0 $\pm$ 16.4 (0, 71.9)
EQ5D-5L Index Value (0-1)	0.69 $\pm$ 0.17 (0.1, 1)
Rating scale current health (0-100)	64.9 $\pm$ 17.6 (1.0, 100)

* Low educational level: lower technical and vocational training and lower general secondary education; Medium education level: intermediate vocational training and advanced secondary education; High education level: higher vocational education and university.

** This was a single item in the PDQ-8. Here the patients are reported who sometimes, often or always experienced this item during the last month.

¥ Multiple replies possible.

Treatment Desirability in Parkinson's Disease

Due to incomplete data or incoherent responses in the best-worst scaling questions, an additional 37 patients were excluded from this analysis. These excluded patients were significantly older than 75 and were less educated.

The conditional logit model included data for 192 patients and showed that patients perceived the option of neurosurgery as least desirable and 'oral intake of medication' as most desirable (Table 3, column 3). Figure 2 shows that treatment modality had the greatest impact on the perceived desirability of treatment (importance weight of 1). This was followed by two attributes related to treatment efficacy: the effect of treatment on reducing posture and balance problems (0.69) and

slowness of movement (0.65). Of the side effects, suffering from dyskinesia was perceived as least desirable (0.53), followed by dizziness (0.45) and drowsiness (0.39). However, the occurrence of described side effects (dizziness, drowsiness and dyskinesia) had less impact on the perceived desirability of treatment than the treatment's effect on motor symptoms. These findings are confirmed by the results of the direct assessment, where 67% of patients thought the perceived benefits of treatments were (very) influential in treatment choice while only 35% of patients considered the side effects of treatments to be (very) influential.

Columns 4-5 in Table 3 reveal the preference data of the two subgroups which were identified through latent class analysis. Both classes follow the previous observation that the benefits of treatment were more influential than the possible side effects. The importance weights of class 1 resemble the importance weights of the average estimate, except that it is even more important to optimise the treatment modality (Fig. 2). The preference patterns of class 2 are very different from the average estimate and class 1, because the attributes related to motor symptoms had the greatest impact on treatment desirability. The focus of class 2 was on slowness of movement (importance weight of 1.0), posture and balance problems (0.97), and tremors (0.85). Furthermore, a notable difference was that these patients preferred neurosurgery over pump infusion while the average estimate and patients in class 1 preferred pump infusion over neurosurgery. Next, potential predictors were added to the model to determine whether certain patients are more likely to belong to class 2. In comparison with class 1, class 2 includes an above-average number of patients who had experienced advanced treatments (OR: 3.1) and who were still employed (OR: 1.3). In addition, a small significant difference was found between disease duration in both classes, with patients with a shorter disease duration more likely to belong to class 1 (OR: 0.19). Based on the highest probability to belong to a class, 57 patients were assigned to class 1 and 135 patients were assigned to class 2. Prevalence of significant predictors were then: employment (27.2% in class 1 versus 13.1% in class 2), experience with advanced treatment (32.7% versus 9.56%), and disease duration (6.8 years versus 7.3 years). We found no evidence that age, gender, education, independence, or quality of life scores significantly influenced the preference patterns in the calculated latent class model.

Table 3. Treatment desirability in Parkinson's Disease based on conditional logit analysis and latent class analysis (N=192).

Attribute	Level	Average Estimate [§]	Class 1: Patients preferring 'Process optimisation' [§]	Class 2: Patients preferring 'Symptom optimisation' [§]
Treatment modality	Oral tablets	1.83 (0.08)*	2.10 (0.11)*	1.20 (0.20)*
	Pump	-1.91 (0.08)*	-2.07 (0.11)*	-1.52 (0.17)*
	Neurosurgery	-2.74 (0.09)*	-4.36 (0.21)*	0.60 (0.24)*
Tremor	Seldom to never	1.66 (0.09)*	1.64 (0.11)*	1.77 (0.17)*
	Sometimes	0.21 (0.10)*	0.32 (0.12)*	0.08 (0.19)
	Often	-1.17 (0.09)*	-1.03 (0.14)*	-1.69 (0.17)*
Posture and balance problems	Seldom to never	1.70 (0.09)*	1.70 (0.11)*	1.82 (0.16)*
	Sometimes	-0.03 (0.10)	0.08 (0.12)	-0.24 (0.19)
	Often	-1.46 (0.09)*	-1.23 (0.12)*	-2.11 (0.17)*
Slowness of movement	Seldom to never	1.70 (0.09)*	1.59 (0.11)*	2.03 (0.17)*
	Sometimes	0.26 (0.10)*	0.44 (0.12)*	-0.15 (0.19)
	Often	-1.26 (0.09)*	-0.84 (0.13)*	-2.03 (0.17)*
Dizziness	Seldom to never	1.10 (0.09)*	1.11 (0.11)*	1.20 (0.18)*
	Sometimes	0.08 (0.10)	0.15 (0.12)	-0.07 (0.19)
	Often	-0.95 (0.10)*	-0.98 (0.13)*	-1.09 (0.18)*
Drowsiness	Seldom to never	1.16 (0.09)*	1.18 (0.12)*	1.21 (0.18)*
	Sometimes	0.23 (0.10)*	0.37 (0.12)*	-0.03 (0.19)
	Often	-0.62 (0.10)*	-0.55 (0.13)*	-0.83 (0.19)*
Dyskinesia	Seldom to never	1.36 (0.09)*	1.28 (0.11)*	1.66 (0.17)*
	Sometimes	-0.12 (0.10)	-0.02 (0.13)	-0.28 (0.19)
	Often (ref) ^{§§}	-1.06 (0.09)*	-0.89 (0.13)*	-1.53 (0.18)*
Class probability model^{§§§}				
Constant			(ref)	0.56 (0.41)
Employment			-	0.48 (0.23)*
Experience with advanced treatments			-	1.13 (0.27)*
Disease duration			-	-0.07 (0.04)*
Average class probability			0.704	0.296
Class membership			N=135	N=57

[§] A high positive part-worth utility reflects that the attribute-level is likely to be preferred relative to other attribute levels (the opposite is true for negative part-worth utilities).

^{§§} The data were entered using the effect-coding system. The part-worth utility of the reference category can be calculated as -1 * (the sum of the estimated part-worth utilities)

^{§§§} The membership predictors significantly improved the model fit compared to a model including no predictors (log likelihood -4626 -4533, P < 0.001).

* p < 0.05

Figure 2. Importance weights of the attributes estimated from the conditional logit analysis and latent class analysis.

Patient Involvement

The roles patients preferred in decision-making about treatment were 8% active, 37% active-shared, 30% collaborative, 24% passive-shared and 1% passive (N=212) (Table 4). The roles that patients reported actually experiencing were 12% active, 59% collaborative, and 29% passive. Patients' preferred and perceived roles were significantly associated ($\chi^2(8) = 40.1, P < .001$). For 78% of patients it can be concluded that they approximately experienced their preferred role during decision-making. In total, relatively few patients (11.3%) experienced extreme discordance between their preferred role and their perceived role (Table 4). The majority of patients (45%) preferred to take the lead in treatment decisions, however 10.8% preferred a more passive or collaborative decision role. In addition, the direct questions revealed that only 10% of patients take into account the advice of family or friends when deciding about treatment, and assistance needed from partner was only relevant for 23% of patients.

Table 4. Congruence between the patient's preferred and perceived decision role in treatment decision making (N=212).

Perceived Decision Role → Preferred Decision Role ↓	Active: Treatment was largely chosen by myself	Collaborative: Treatment was chosen in collaboration with the doctor	Passive: Treatment was largely chosen by the doctor	Total
Active: I prefer to make the final treatment decision.	5*	9**	3**	17 (8%)
Active-shared: I prefer to make the final treatment decision after seriously considering my doctor's opinion.	17*	51*	11**	79 (37%)
Collaborative: I prefer that my doctor and I share responsibility for deciding which treatment is best.	3	41*	20	64 (30%)
Passive-shared: I prefer that my doctor makes the final treatment decision, but seriously considers my opinion.	1**	23*	26*	50 (24%)
Passive: I prefer to leave all treatment decisions to my doctor.	-	-	2*	2 (1%)
Total	26 (12%)	124 (59%)	62 (29%)	212 (100%)

$\chi^2(8) = 40.1, P < .001$

* Patients who approximately had their preferred role during decision making.

** Patients who experienced extreme discordance in their preferred and perceived decision role during decision making

DISCUSSION

The objective of this study was to elicit patient preferences about motor symptoms, side effects, and process characteristics of the main treatments in PD and to measure patients' preferred and perceived involvement in decision-making about treatment. The best-worst scaling data indicated that the occurrence of described side effects (dizziness, drowsiness, and dyskinesia) has less impact on the perceived desirability of treatment than the treatment's effect on motor symptoms. In contrast, a study of Hattori et al. in Japan (N=371) found that more than half of the study participants preferred to avoid side effects (dyskinesia, drowsiness, constipation etc.) rather than obtain effective relief from bradykinesia (28). However, this number gradually decreased with increasing symptom severity. The difference between studies might be explained by cultural differences or by the way the questions were posed (direct/indirect). In our study dyskinesia was identified as the most influential side effect for treatment preference; however compared to

dizziness and drowsiness, dyskinesia was the side effect from which the fewest patients suffered. This might be explained by findings in other studies that patients who had not yet experienced dyskinesia were more concerned about avoiding dyskinesia than patients who had already developed dyskinesia (28, 29).

Furthermore, two preference patterns were found in the patient sample. One class focused largely on optimising the process of care, while the other class focused more on controlling motor symptoms. Patients who had experienced advanced treatments, who were still employed, or who had a shorter disease duration were more likely to belong to the latter class. One can only speculate about the reasons why this was the case. It may be that experience with advanced treatment increases its perceived value (or reduces the disutility of its negative aspects) and patients are less interested in process optimisation. On the other hand, these patients may have had different preferences than the average patient before choosing advanced interventions (30). Second, Murphy et al. (2013) found that employed patients reported slowness and tremor as their greatest occupational challenges and therefore these motor symptoms might be perceived to be more important by this class in our study (31). Although disease duration significantly differs between the two preferences groups one it is arguable whether this difference is clinically relevant (6.8 years versus 7.3 years). However, for patients with a recent onset (short disease duration), it may be assumed that treatment still has a direct, apparent effect on motor symptoms and thus they may be perceived as more important. We did not find evidence that age influenced preference patterns in our model. However, it is unlikely that the decision-making process is the same in a 45-year-old man as in an 80-year-old man, though we did not find evidence for this in the current model.

Previous studies have indicated that it is important to provide care that is tailored to each PD patient's individual values and preferences (14, 15). Our study revealed subgroups with varying preferences, indicating that there is no "one-size-fits-all" treatment and that treatment decisions have to be individualised. Fortunately, we found that many patients (45%) preferred an active role in the treatment decision-making process. Only 1% of patients prefer to leave the treatment decision entirely up to the doctor, while 24% of patients reported that in their case the treatment decisions were largely made by their doctor. If patients desire more involvement in the decision-making process, it is essential for the doctor to know what is important to patients. We are currently studying the possibility of developing a value clarification tool to facilitate individual treatment decisions for advanced treatments in PD. This tool will help physicians to propose individualised treatment strategies which are tailored to the patient's preferences and enhance collaborative or

shared decision-making. Understanding the advantages and disadvantages of the available treatment options will help patients feel more involved and more certain about the relevance and likely efficacy of their treatment and may therefore enhance adherence to treatment (17).

Limitations

Although the attribute selection process in this study can be criticised because we only interviewed a limited number of patients through convenience sampling (selection bias), we did rely on a process of open, semi-structured and structured interview questions. We feel that one can attach value to all answers, as the answers patients gave spontaneously highlighted their main daily concerns (which might not be related to specific drugs), while providing a detailed list of symptoms and side effects might have brought other concerns to the surface. With a different group of patients, the definite selection of attributes might have been altered. However, the background characteristics of this sample and the final patient sample were comparable and we do feel that this article highlights important trade-offs between attributes that patients make in expressing preferences for treatment. However, this does not mean that our attribute selection process was comprehensive: the results of our study only hold true for the importance of the symptoms and side effects that were part of the treatment scenario. Lastly, the physical aspects of PD are the defining characteristics of the disease and, understandably, patients focused on those during the interviews. Our final selection of attributes did not include non-motor symptoms. However, literature suggests that non-motor symptoms have, as a whole, a greater impact on quality of life than motor symptoms (32, 33). Further research should focus on the relative value of non-motor symptoms versus motor symptoms.

In this study, the method best-worst scaling was selected to elicit preferences. Compared to other trade-off methods such as discrete choice experiments, time trade-off, and standard gamble, literature suggests that best-worst scaling has a lower cognitive burden (18, 34, 35). Rating scales would probably have been easier for the patient to complete, but then no statements could have been made about the relative desirability of treatment attributes. However, overall 45 of the 132 submitted paper-questionnaires had to be removed from the sample (34%). Older (>75) and less educated patients were excluded more often, which might be explained by the initial and declining cognitive abilities as a result of age, or their disease (36). Completion of the questionnaire together with a researcher or during an interview would probably have improved the reliability of the

answers, but would have introduced the risk of interviewer bias and, in any case, was not possible due to time and budget constraints.

For the data collection, there was more uncertainty regarding the non-hospital based sample, e.g. it was explicitly stated that patients with atypical parkinsonism should not fill out the questionnaire, but we cannot be sure this did not happen. In the analysis phase, the hospital and non-hospital sample were merged. Although in the latter sample the patients were younger and more frequently employed, the two samples were almost equally divided among the two estimated latent classes (48-52%).

For the Control Preference Scale questions, the different operationalisation of the perceived and preferred decision roles led to difficulties with estimating the level of (dis)agreement between them. Therefore we only estimated extreme discordance between the roles and approximated the level of agreement. However, this might have led to a bias in the results found. Had the same number of answer categories been used, it would have been possible to calculate the level of (dis)agreement using Cohen's kappa. In addition, the Control Preferences Scale does not recognise the decision role of the partner or caregiver, while in (later stages of) PD this role becomes very important (37). In our study, 20% of the patients completed the questionnaire together with their partner or caregiver, which also shows their interest in the subject.

Conclusions

Our study revealed differences in the way subgroups of patients weigh the different attributes of treatment, indicating there is no "one-size-fits-all" treatment possible and decision-making about treatment needs to be individualised. It was found that patients who had experienced advanced treatments, who were still employed, or who had a shorter disease duration were more likely to focus their treatment on controlling motor symptoms rather than on optimising the process of care. In addition, our study has shown that many patients prefer an active role in decision-making about treatment. Both results stress the need for physicians to know what is important to patients and to share treatment decisions to ensure that patients receive the treatment that aligns with their preferences.

REFERENCES

1. Lees AJ, Hardy J, Revesz T. Parkinson's disease. *Lancet*. 2009;373(9680):2055-66.
2. Hamberg K, Hariz G-M. The decision-making process leading to deep brain stimulation in men and women with parkinson's disease - an interview study. *BMC Neurology*. 2014;14(1):89.
3. Kurlan R. "Levodopa phobia": a new iatrogenic cause of disability in Parkinson disease. *Neurology*. 2005;64(5):923-4.
4. Leopold NA, Polansky M, Hurka MR. Drug adherence in Parkinson's disease. *Mov Disord*. 2004;19(5):513-7.
5. Wertheimer AI, Radican L, Jacobs MR. Assessing different perspectives on the value of a pharmaceutical innovation. *Southern Med Review*. 2010;3(1):24-8.
6. Arbouw MEL, Movig KLL, Guchelaar HJ, Poels PJE, Van Vugt JPP, Neef C, et al. Discontinuation of ropinirole and pramipexole in patients with Parkinson's disease: Clinical practice versus clinical trials. *Eur J Clin Pharmacol*. 2008;64(10):1021-6.
7. Ryan M, Kinghorn P, Entwistle VA, Francis JJ. Valuing patients' experiences of healthcare processes: towards broader applications of existing methods. *Soc Sci Med*. 2014;106:194-203.
8. Watt S. Clinical decision-making in the context of chronic illness. *Health Expect*. 2000;3(1):6-16.
9. Clarke CE, Worth P, Grosset D, Stewart D. Systematic review of apomorphine infusion, levodopa infusion and deep brain stimulation in advanced Parkinson's disease. *Parkinsonism Relat Disord*. 2009;15(10):728-41.
10. Martinez-Martin P, Kurtis MM. Health-related quality of life as an outcome variable in Parkinson's disease. *Ther Adv Neurol Disord*. 2012;5(2):105-17.
11. Martinez-Martin P, Rodriguez-Blazquez C, Frades-Payo B. Specific patient-reported outcome measures for Parkinson's disease: analysis and applications. *Expert Rev Pharmacoecon Outcomes Res*. 2008;8(4):401-18.
12. Warren C, McGraw AP, Van Boven L. Values and preferences: defining preference construction. *Wiley Interdiscip Rev Cogn Sci*. 2011;2(2):193-205.
13. Mühlbacher AC, Juhnke C. Patient preferences versus physicians' judgement: Does it make a difference in healthcare decision making? *Appl Health Econ Health Policy*. 2013;11(3):163-80.
14. Jankovic J. Parkinson's disease therapy: tailoring choices for early and late disease, young and old patients. *Clin Neuropharmacol*. 2000;23(5):252-61.
15. van der Eijk M, Nijhuis FAP, Faber MJ, Bloem BR. Moving from physician-centered care towards patient-centered care for Parkinson's disease patients. *Parkinsonism Relat Disord*. 2013;19(11):923-7.
16. Barry MJ, Edgman-Levitan S. Shared decision making - The pinnacle of patient-centered care. *N Engl J Med*. 2012;366(9):780-1.
17. Grosset KA, Grosset DG. Patient-perceived involvement and satisfaction in Parkinson's disease: effect on therapy decisions and quality of life. *Mov Disord*. 2005;20(5):616-9.
18. Louviere JJ, Flynn TN, Marley AAJ. *Best-Worst Scaling Theory, Methods and Applications*. Cambridge: Cambridge University Press; 2015. 360 p.
19. Flynn TN, Louviere JJ, Peters TJ, Coast J. Best-worst scaling: What it can do for health care research and how to do it. *J Health Econ*. 2007;26(1):171-89.
20. Jenkinson C, Fitzpatrick R, Peto V, Greenhall R, Hyman N. The Parkinson's Disease Questionnaire (PDQ-39): development and validation of a Parkinson's disease summary index score. *Age Ageing*. 1997;26(5):353-7.
21. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727-36.
22. ATLAS.ti Scientific Software Development Company. ATLAS.ti. Berlin, Germany: GmbH.
23. Johnson FR, Lancsar E, Marshall D, Kilambi V, Mühlbacher A, Regier DA, et al. Constructing experimental designs for discrete-choice experiments: Report of the ISPOR conjoint analysis experimental design good research practices task force. *Value Health*. 2013;16(1):3-13.
24. Strull WM, Lo B, Charles G. Do patients want to participate in medical decision making? *Jama*. 1984;252(21):2990-4.

25. Degner LF, Sloan JA, Venkatesh P. The Control Preferences Scale. *Can J Nurs Res*. 1997;29(3):21-43.
26. Jenkinson C, Fitzpatrick R, Peto V, Greenhall R, Hyman N. The PDQ-8: Development and validation of a short-form Parkinson's disease questionnaire. *Psychol Health*. 1997;12(6):805-14.
27. Swait J. A structural equation model of latent segmentation and product choice for cross-sectional revealed preference choice data. *Journal of Retailing and Consumer Serv*. 1994;1(2):77-89.
28. Hattori N, Fujimoto K, Kondo T, Murata M, Stacy M. Patient perspectives on Parkinson's disease therapy in Japan and the United States: results of two patient surveys. *Patient Relat Outcome Meas*. 2012;3:31-8.
29. Hung SW, Adeli GM, Arenovich T, Fox SH, Lang AE. Patient perception of dyskinesia in Parkinson's disease. *J Neurol Neurosurg Psychiatry*. 2010;81(10):1112-5.
30. Kim MR, Yun JY, Jeon B, Lim YH, Kim KR, Yang HJ, et al. Patients' reluctance to undergo deep brain stimulation for Parkinson's disease. *Parkinsonism Relat Disord*. 2016;23:91-4.
31. Murphy R, Tubridy N, Kevelighan H, O'Riordan S. Parkinson's disease: How is employment affected? *Ir J Med Sci*. 2013;182(3):415-9.
32. Martinez-Martin P, Rodriguez-Blazquez C, Kurtis MM, Chaudhuri KR. The impact of non-motor symptoms on health-related quality of life of patients with Parkinson's disease. *Mov Disord*. 2011;26(3):399-406.
33. Müller B, Assmus J, Herlofson K, Larsen JP, Tysnes OB. Importance of motor vs. non-motor symptoms for health-related quality of life in early Parkinson's disease. *Parkinsonism Relat Disord*. 2013;19(11):1027-32.34. Potoglou D, Burge P, Flynn T, Netten A, Malley J, Forster J, et al. Best-worst scaling vs. discrete choice experiments: an empirical comparison using social care data. *Soc Sci Med*. 2011;72(10):1717-27.
35. Ali S, Ronaldson S. Ordinal preference elicitation methods in health economics and health services research: using discrete choice experiments and ranking methods. *Br Med Bull*. 2012;103(1):21-44.
36. Muslimović D, Post B, Speelman JD, Schmand B. Cognitive profile of patients with newly diagnosed Parkinson disease. *Neurology*. 2005;65(8):1239-45.
37. Abendroth M, Lutz BJ, Young ME. Family caregivers' decision process to institutionalize persons with Parkinson's disease: A grounded theory study. *Int J Nurs Stud*. 2012;49(4):445-54.

APPENDIX A: SELECTION OF ATTRIBUTES AND LEVELS.

Introduction

In a preliminary study, the identification of relevant symptoms, side effects and process characteristics in Parkinson's Disease (PD) took place through qualitative interviews with 18 patients. This provided the basis for identifying the full set of attributes that characterise PD treatments and influence patients' Health Related Quality of Life (HRQoL). Subsequently, these results were discussed in a project meeting and based on the qualitative and quantitative importance of these characteristics in patients' daily life attributes were selected for the preference experiment.

Methods

The interviews with patients were semi-structured using a framework of important themes influencing HRQoL in PD (table 1). Patients were first asked to describe their health using the Parkinson-Disease Questionnaire-39 (PDQ-39) and the EuroQol-5D-5L (EQ5D-5L). With this information the interviewer got an overall impression of the health status of the patients and used this information to structure the interviews and to decide which health domains were important to discuss (semi-structured interviews). Hereafter, patients were asked about the main characteristics of their treatment, the positive and negative outcomes, the influence on HRQoL and other themes listed in table 1. The interviews were tape recorded and the expected duration of the interviews was about an hour.

Table 1. Structured themes during the interviews.

Global themes	Points of discussion
Personal characteristics	Age, marital status, children, employment
Global health status	Diagnosis, duration, medication use, health problems, PDQ-39 and EQ5D-5L assessment.
Effect of symptoms on HRQoL	Efficacy, symptom suppressing (duration)
Effect of side-effects or motor-fluctuations on HRQoL	Side effects (frequency, duration), influence of daily life, experience with dyskinesia, wearing off, on-off effects and/or freezing
Impact of treatment (process) characteristics on HRQoL	Treatment regime, experience, difficulties, satisfaction, quitting/adhering to a drug, advanced treatment (brain surgery, pump)

Data Analysis

For the PDQ-39 dimension scores were obtained and a summary index was calculated. For the EQ5D-5L an overall index utility value was retrieved. The interviews were transcribed verbatim and coded and analysed in ATLAS.ti (software for the qualitative analysis of large bodies of textual and audio data). The final result was a table ranking the frequency of symptoms and side effects (attributes) and their impact on HRQoL according to these patients. Discussion with a project team consisting of a neurologist, a hospital pharmacist, a rehabilitation specialist, three health sciences researchers, and two patients was used to narrow the list of attributes. The project team selected attributes based on: the qualitative and quantitative importance of these attributes of treatment in the daily life of patients (interview results), and the limitations of the chosen stated preference method, which only allowed for inclusion of between four and eight attributes. More-over, we aimed to reach attribute balance (i.e. include equal positive and negative outcomes of treatment choice and to make sure that the selected attributes were typical symptoms and side effects of PD and not too general.). At the end of the project meeting the goal was to have selected attributes that highlights important trade-offs between attributes that patients make in expressing preference for treatment.

Interview Results

Table 2 shows the sociodemographic and clinical characteristics of the 18 patients enrolled in the interviews. Table 3 shows the results of the PDQ-39: patients suffered the most from problems in the domains of 'activities of daily living and mobility'. Only few problems were identified in the stigma and social support domains.

Table 2. Socio-demographics and clinical characteristics (N=18).

Sample characteristics (N=255)	N (%) or mean $\pm$ SD (min, max)
Gender	
Man	13 (72.2)
Woman	5 (11.1)
Age (years)	66 $\pm$ 7.0 (54, 79)
Employment status	
Employed	3 (16.7)
Disabled / unable to work	4 (22.2)
Retired	11 (61.1)
Marital state	
Single	2 (11.1)
With partner (no children living at home)	15 (83.3)
With partner and children living at home	1 (5.6)
Disease duration	8.0 $\pm$ 7.15 (1, 22)
Current treatment*	
Oral medication	18 (100.0)
Continuous pump infusion (subcutaneous)	-
Continuous pump infusion (intraduodenal)	1 (5.6)
Neurosurgery	3 (16.6)
Current oral medication treatment	
Levodopa	7 (38.9)
Dopamine agonist	2 (11.1)
Combination therapy	9 (50.0)

Table 3. Health-related quality of life questionnaires (N=18).

Variables	Mean (standard deviation)
PDQ-39	
Mobility	32.78 (25.67)
Activities of daily living	38.89 (23.31)
Emotional well-being	23.61 (17.21)
Stigma	9.03 (13.08)
Social support	12.96 (18.35)
Cognitive impairment	37.5 (15.46)
Communication	30.06 (15.19)
Bodily discomfort	30.39 (15.54)
PDQ-39 Summary Index	30.38 (15.5)
EQ5D-5L Index Value	0.69 (0.25)
EQ5D-VAS	68.72 (11.47)

Table 4. Symptoms and side effects and their influence on HRQoL(N=18).

	no. of patients suffering	Influence on HRQoL (according to most patients)
Symptoms		
Postural instability	15	Major
Tremors	13	Major
Slowness of movement (reduced gait, leg dragging, reduced arm swing)	10	Major
Freezing	8	Major
Rigidity, muscle stiffness	7	Major
Anxiety – Depression	3	Major
Constipation	2	Minor
Writing	2	Minor
Loss of smell	1	Minor
Crying	1	Moderate
Swallowing	1	Moderate
Drooling	1	Moderate
Side effects		
Dyskinesia	10	Major
Drowsiness (sleepy, daytime sleepiness)	8	Major
Dizziness (orthostatic hypotension)	8	Moderate
Wearing off	8	Moderate
Constipation, diarrhea and incontinence	7	Moderate
On-off response	5	Major
Nausea, stomach pain and or vomiting	5	Minor
Falling asleep at night	4	Moderate
Headache	3	Moderate
Confusion, hallucinations and paranoia	3	Major
Waking up too early	2	Minor
More frequent dreams and nightmares	1	Moderate
Dislocation intestinal tube	1	Minor

Table 4 shows the final result of the interviews: a table ranking the frequency of symptoms and side effects (attributes) and their impact on HRQoL according to most patients. Since not all patients had experience with advanced treatment, this theme was not quantified. However, the qualitative results will be reported.

Impact of Treatment (process) Characteristics on HRQoL

The last theme of the interview was how the process characteristics of different treatments in PD affect patient's daily life and HRQoL. All patients took oral medication and thought this only had a minor influence on their HRQoL. There were only four patients who had experienced brain surgery or were currently wearing a pump. At first instance, these patients had doubts about starting advanced treatments, but in the end they adapted to their situation because it brought them relief

of symptoms. The fourteen non-experienced patients were asked to think of undergoing brain surgery or using a pump (and likewise to consider all routines, operations and factors involved). All patients were very firm in their opinion and were certain that starting with advanced treatments would have a major influence on their daily life and HRQoL, despite the benefit of expected symptom reduction. With respect to brain surgery, most patients were aware of the risk of complications, but the prospect of being awake during deep brain stimulation provoked most anxiety and fear. Besides the surgery itself, patients expected the use of the neurostimulator to be complicated and difficult, as well as with a pump. With a pump patients stood negatively against carrying the pump at all times, changing the drug cassette, rinsing the intestinal tube or changing the infusion needle/line. All in all, the advanced treatment modalities will have a major influence on HRQoL.

Project Meeting

After deliberation, one process attribute (treatment modality), three motor symptoms (tremor, slowness of movement, and posture and balance problems) and three side effects (dizziness, drowsiness, dyskinesia) were selected for inclusion in the preference task.

The process attribute:

The selection of the process attribute reflected interviewed patients expressing concerns regarding the impact of surgery and pump procedures on their daily life, despite the expected reduction in symptoms. This highlights an important trade-off in expressing preference for treatment. Therefore the process attribute was operationalized as treatment modality: whether the treatment required brain surgery, continuous infusion of medication via a pump, or oral medication.

Motoric symptoms:

The physical aspects of PD (motoric symptoms) are the defining characteristics of the disease and, understandably, patient's focus most on those during the interviews. The three selected symptoms for the preference instrument were: tremor, posture/balance problems, and slowness of movement. These symptoms occurred most, are easily recognized and had a major impact on HRQoL according to interviewed patients. In the project team the exclusion of rigidity / muscle stiffness was heavily debated, but eventually the project team selected the attributes which were most often reported by the patients in the interviews.

Side effects:

To keep a balance between symptoms and side effects, also three side effects were selected: dyskinesia, drowsiness, and dizziness. We aimed to select side effects which were recognizable to most patients, and side-effects which were typical to PD and not too general. The interview results showed that dyskinesia occurred the most often and had the largest impact on patient's daily life, because of the duration, the unpredictable character, and the obstruction of daily tasks. Secondly, side effects were most often reported in the sleeping domain and the selected attribute drowsiness was defined as extensive daytime sleepiness. Nausea, stomach pain, vomiting, diarrhea, constipation also frequently occurred, but most had only a limited to moderate impact on HRQoL. Besides, nausea and stomach pain are examples of side effects which are common in taking oral medication. Lastly, dizziness (lightheadedness caused by orthostatic hypotension) was selected at the cost of 'hallucinations and paranoia'. Although the latter had a greater impact on HRQoL, it was only reported by three patients.

Level assignment

The next step was to describe the variation in possible outcomes for each attribute. In order to reduce the cognitive difficulty, three qualitative levels were chosen to represent the burden of symptoms and side effects in the treatment scenarios (seldom to never, sometimes, and often suffer from). The attribute treatment modality was described as the oral intake of tablets, continuous pump infusion of medication, and neurosurgery. Table 5 displays the final selection of attributes, their descriptions and levels, which served as the input for the preference experiment.

Table 5. *Selected attributes and levels (descriptions)*

Attributes	Description	Level operationalization
Treatment modality	The path by which a treatment enters (or is applied to) the body.	Oral tablets Medication through pump infusion Brain surgery
Tremor	Rhythmic muscle contraction	Seldom to never suffer from Sometimes suffer from Often suffer from
Posture and balance problems	Abnormal axial postures and loss of postural reflexes.	Seldom to never suffer from Sometimes suffer from Often suffer from
Slowness of movement	Difficulties with planning, initiating and executing movement	Seldom to never suffer from Sometimes suffer from Often suffer from
Drowsiness	Excessive daytime sleepiness	Seldom to never suffer from Sometimes suffer from Often suffer from
Dizziness	Lightheadedness, effect of orthostatic hypotension	Seldom to never suffer from Sometimes suffer from Often suffer from
Dyskinesia	Repetitive, involuntary muscle movement	Seldom to never suffer from Sometimes suffer from Often suffer from

4

Comparison of patient and public preferences for treatment attributes in Parkinson's Disease

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

Submitted

ABSTRACT

Background: Patient and public preferences for therapeutic outcomes or medical technologies are often elicited, and discordance between the two is frequently reported. Our main objective is to compare patient's and public preferences for treatment attributes in Parkinson's Disease (PD).

Methods: A representative sample from Dutch PD patients and the general public were invited to complete a best-worst scaling case 2 experiment consisting of six health-related outcomes and one attribute describing the specific treatment (brain surgery, pump, oral medication). Data were analyzed using mixed logit models, and attribute impact was estimated and compared between populations (and population subgroups).

Results: Both the public (N:276) and patient population (N:198) considered treatment modality the most important attribute, although patients attached higher relative importance. Both attached high disutility to pump infusion and brain surgery and prefer drug treatment. Most health outcomes were valued equally by patients and public, except for reducing dizziness (more important to the public) and improving slow movement (more important to patients).

Discussion: Although this data does not support definite conclusions on whether patients are less likely to undergo invasive treatments, its (predicted) choice probability would be low based upon the (un) desirability of the attribute-levels. Patients with PD might have adapted to their condition and are not willing to undergo advanced treatments in order to receive health improvements. Both public and patients' preferences entail information that is potentially relevant for decision makers, patient preferences can inform decision makers about the likelihood of adaptation to a condition.

INTRODUCTION

Parkinson's Disease (PD) is a chronic neurodegenerative disease, and no curative treatment is available. Many symptoms and complications accompany PD, such as slower movement (bradykinesia), resting tremor, muscular rigidity, and dyskinesia. Symptomatic treatment aims to improve the patient's quality of life and functional capacity (1-3). Early-stage PD is predominantly treated with drugs. However, recent developments to extend therapeutic options for patients with more advanced PD include novel approaches such as continuous pump infusion technology (intraduodenal or subcutaneous) and brain surgery techniques (deep brain stimulation) (4). The device-aided therapies are effective in reducing motor symptoms, but currently no direct comparative data exist to support the use of one therapy over another (5).

In addition, new technologies are systematically evaluated on their social, economic, organizational and ethical consequences in health care (6-8). While such health technology assessment is performed from a public perspective on the need for reimbursement of these technologies (7, 8), it is sometimes argued that patient's preferences should be considered. Patients have experience with the condition and can provide unique knowledge that can support more informed decision making on the need for reimbursement (9).

Although discordance is frequently reported between patient and public preferences, no preference studies are available that compare patient and public valuations for clinical outcomes of PD technologies (10, 11). Therefore, the first objective of this study is to compare patient and public preferences for clinical outcomes (symptom reduction, side effects) of PD treatment. It is hypothesized that patients will place higher importance on the burden of treatment, because they have experience with side effects (in contrast to the public).

Besides maximization of the treatment's effect on health outcomes, it can be of paramount importance to also evaluate the process of care (12, 13). Especially in evaluations in chronic diseases, the possibility exists that discomfort of treatment can outweigh treatment benefits. Hence, besides the effectiveness of a drug, other attributes such as: ease of use and treatment modality influence its use by patients and -through adherence- its efficacy (14, 15). For the newer technologies in PD, it is particularly interesting to study the relative trade-offs between the process of care (taking pharmaceuticals, having pump infusion, or undergoing brain surgery) and improved health outcomes. Therefore, the second objective of this study is to compare the impact of treatment modality on treatment desirability from the patient and public perspective. It is

hypothesized that the patients have a more positive view of pump infusion and brain surgery, because they might be more willing to adapt in order to receive improved symptom reduction.

Lastly, preferences have the tendency to be influenced by socio-demographics and experience (16-18). In an accompanying study, we already identified different preference patterns for patients with PD who had experience with advanced treatments compared to patients who had no experience (19). In general, public preferences are regarded as ex-ante preferences which are obtained without any experience (20). However, individuals may have more or less experience with PD depending on their social environment, or may find it easier to relate to PD patients depending on their age or declining health. Hence, the last objective is to explore whether public preferences differ based on sociodemographic variables or their familiarity with the disease.

METHODS

Selection of Attributes and Levels

The first step in each preference study is to select attributes and levels. The identification of relevant clinical outcomes took place through qualitative interviews with 18 patients. Patients were asked which symptoms, side effects or actions were mostly impeding their daily life (e.g. difficulties with walking, dizziness and taking pills). The interviews provided the basis for identifying the full set of attributes that characterize PD treatments and influence patient's Health Related Quality of Life (HRQoL). A project team consisting of a pharmacist, neurologist, rehabilitation specialist, two patients and researchers discussed the interview results and selected less than eight attributes, because the human mind is limited in its ability to include many attributes and levels in decision making. The selected attributes had to be a balanced set of symptoms, side effects, and process characteristics based on the qualitative and quantitative importance of these treatment attributes in the patients' daily life (interview results). Furthermore, the project team had to make sure that the selected attributes were typical symptoms and side effects of PD and not too general (e.g. nausea). Eventually, they selected seven clinical outcomes: three symptoms (tremor, slowness of movement, posture and balance problems), three side-effects (dizziness, drowsiness and dyskinesia). To minimize the cognitive difficulty for respondents and to frame the treatment profiles as comprehensive as possible (especially important for the patients) the choice was made to have a similar and comprehensive level designation for clinical outcomes: seldom – sometimes – often suffering from. Furthermore, one process attribute was added to refer to the treatment

modalities of the main treatment options (oral medication, continuous infusion of medication via a pump, and undergoing neurosurgery). Table 1 displays the selected attributes and levels (a more detailed description of this preliminary research can be found in Appendix A of Chapter 3).

Table 1. *Selected attributes and levels (descriptions)*

Attributes	Description	Level operationalization
Treatment modality	The path by which a treatment enters (or is applied to) the body.	Oral medication Continuous infusion of medication via a pump Brain surgery
Tremor	Rhythmic muscle contraction	Seldom to never suffer from Sometimes suffer from Often suffer from
Posture and balance problems	Abnormal axial postures and loss of postural reflexes.	Seldom to never suffer from Sometimes suffer from Often suffer from
Slowness of movement	Difficulties with planning, initiating and executing movement	Seldom to never suffer from Sometimes suffer from Often suffer from
Drowsiness	Excessive daytime sleepiness	Seldom to never suffer from Sometimes suffer from Often suffer from
Dizziness	Lightheadedness, effect of orthostatic hypotension	Seldom to never suffer from Sometimes suffer from Often suffer from
Dyskinesia	Repetitive, involuntary muscle movement	Seldom to never suffer from Sometimes suffer from Often suffer from

Design of the Preference Instrument

Multiple techniques are available to measure the relative importance of treatment aspects in PD. More recently, Best-Worst Scaling (BWS) was introduced as a way of obtaining more information by asking respondents to select not only the best alternative (most preferred), but also the least preferred alternative (21). This best-worst question is repeated several times in certain combinations of attributes and levels to determine the relative importance of each item. Different types of BWS exist. Best–worst scaling case 1 was not a relevant type of choice here as the attributes were not equivalent to item statements, but were in need of attribute-levels to indicate the effect of treatment or the occurrence of side effects (e.g. seldom, sometimes or often). BWS case 2 (the profile case) lets respondents evaluate one treatment scenario in one choice task and asks respondents to identify their most and least preferred attribute-level. BWS case 3 is comparable to

a DCE choice set and asks respondents to choose the best and the worst from a number of scenarios (usually three). Five pilot-tests were conducted in our vulnerable patient group to compare the feasibility of BWS case 2 and BWS case 3. Patients reported it was difficult to compare multiple treatment profiles in one choice task and suffered from information overload (BWS case 3). Subsequently, the project team had a preference for BWS case 2 (which was well-received by the patients in the pilot-tests).

The next step was to select the hypothetical treatment scenarios. It is impossible to let respondents evaluate all possible 2187 scenarios (3⁷). Therefore Sawtooth Software was used to select the smallest subset of treatment scenarios that is able to identify all necessary parameters. A D-efficient design of 36 unique treatment scenarios was selected, meaning that the design was balanced (each attribute-level appears twelve times), but not completely orthogonal (meaning there is slight correlation between attribute-levels). The 36 scenarios were then divided into four blocks (versions) of nine scenarios per respondent.

For each treatment scenario, respondents were asked to select the two aspects that they perceived as most and least desirable (Figure 1). Treatment scenarios were formulated so that patients can recognize themselves in the description and the public can evaluate what life would be like living whilst undergoing this treatment. The public sample was also specifically asked to imagine a situation in which they have been diagnosed with PD. Before answering these questions, both patients and public respondents received an extensive explanation and illustration of the selected attributes, levels, and treatment for PD. They received a standard sheet with clinical background information about PD and factual information about the main treatment options (pharmaceutical treatment, neurosurgery, and receiving medication via a pump). Lastly, respondents were asked to answer sociodemographic questions regarding their gender, age, education, and fill out the EuroQol-5D-5L health-related quality of life instrument (22).

Figure 1: Best-Worst Scaling exercise

Most desirable characteristic	What do you find the...	Least desirable characteristic
Treatment profile 1
☐	I receive medication via tablets	☐
☐	I often suffer from tremors	☒
☒	I seldom to never suffer from posture and balance problems	☐
☐	I seldom to never suffer from slowness in motion	☐
☐	I often suffer from dizziness	☐
☐	I sometimes suffer from drowsiness	☐
☐	I often suffer from rapid uncontrolled movements	☐

Study Sample

PD patients who were diagnosed with idiopathic PD at the outpatient department of the hospital Medisch Spectrum Twente in the Netherlands (N = 287) received a paper version of the BWS questionnaire at home. Patients who were registered as having atypical parkinsonism or dementia did not receive an invitation. The data collection was expanded by recruiting respondents through online Parkinson communities, the website and paper magazine of the Dutch Parkinson's Disease Society and PD forums. It could not be verified whether these participants were genuinely diagnosed with idiopathic PD or had atypical parkinsonism or dementia, because there was no access to hospital records. These patients were invited to complete the online survey or to request the paper version.

A Dutch public sample was recruited via an online survey sample coordinated by Survey Sampling International (Rotterdam, the Netherlands). Respondents, aged 18 to 65, were invited via e-mail to respond to the web-based questionnaire.

No literature is available to determine sample sizes for BWS experiments. Therefore sample sizes were determined based on a rule of thumb for conjoint analysis which state that estimate precision increases quickly at sample sizes over 150 and flattens out at +/-300 observations (23, 24). Taking into consideration the number of patients who received an invitation via the Medisch Spectrum

Twente, we aimed at least 150 complete responses for both study samples. According to the Medical-Ethics Committee of the Medisch Spectrum Twente our study did not require assessment according to the Medical Research (Human Subjects) Act.

Data Analysis

Public respondents who did not complete the survey or completed it within five minutes were excluded from analysis. Patient respondents were excluded if their paper questionnaire contained too many blank answers or response errors in the BWS choice tasks (ticked multiple best/worst). Differences in socio-demographic variables between patient and general population were identified with independent sample T-tests and Chi-Square tests.

For the BWS analysis, the maximum difference (maxdiff) model was adopted in which it is assumed that respondents pick both best and worst items at the same time. Part-worth utilities for each attribute-level were obtained using a mixed (random parameter) logit model for each sample to account for the panel nature of the data (25). Dummy coding was applied to each attribute-level. The model estimate for each attribute-level is interpreted as a relative utility (relative to the reference level). The p-value indicates whether the attribute-level significantly ($P < 0.05$) differs from the selected reference level.

The estimated utility weights for patient and public population cannot be directly compared, because of the confound between mean and variance in the latent scale. Importance weights were therefore calculated, based on the difference between minimum and maximum part-worth utilities within an attribute. The largest difference value received an importance weight of one, representing the attribute which had the highest impact on treatment desirability and the other difference values were divided by the largest difference value, resulting in a relative distance of all attributes to the attribute with the highest impact (26). In addition, to present a graphical insight in the value differences, we rescaled the part-worth utilities on a 0-1 scale and presented a scatterplot. The attribute-level with the lowest part-worth utility was anchored at 0, and the attribute-level with the highest part-worth utility was anchored at 1 (Equation 1). Correlation between patient's and public preferences was assessed using Pearson correlation coefficient.

Equation 1. Formula used to rescale BWS part-worth utilities on a 0-1 scale.

$$\text{Rescaled value } (A_i) = \frac{U(A_i) - U(A_{\min})}{U(A_{\max}) - U(A_{\min})}$$

$U(A_i)$ represents the part-worth utility of an attribute level, $U(A_{\min})$ is the attribute-level with the lowest part-worth utility and $U(A_{\max})$ is the attribute-level with the highest part-worth utility.

Furthermore, explorative subgroup analyses were performed to examine whether public respondents are shaping and (re) constructing their preferences on the basis of their experiences (with PD). The variables age, quality of life (eq5d scores), and familiarity with PD were tested. Familiarity with PD was operationalized as being close to someone (i.e. family, friends) who suffers from PD. All data was analyzed using Stata version 14 (StataCorp, College Station, TX).

Please note that this article focuses on the differences between patient and public preferences, whereas the extensive clinical implications of patient preferences have been published elsewhere (19). In addition, the public dataset has been used in an article to compare the outcomes of different preference methods in measuring process-related utilities (27).

RESULTS

In total, 228 patients returned the paper-questionnaire or completed the online survey. Some paper-questionnaires were incomplete or contained multiple response errors and were excluded from analysis ($N=30$, 13.2%). 285 complete questionnaires were filled out online by the general public respondents (drop-out rate: 35%), but nine respondents (3.1%) were excluded because of the time-criterion. The differences between the patient and public population in the distribution of age, gender, education and EQ5D-5L index value are displayed in Table 2.

Table 2. Demographic details of study participants

Variable	Patients ($n = 198$)	Public Population ($n = 276$)	<i>P</i> value
Gender Male (n , %)	133 (67.2)	98 (35.5)	<0.001
Education (n , %)*			
Low	70 (35.4)	52 (18.9)	<0.001
Middle	55 (27.8)	119 (43.3)	
High	73 (36.9)	104 (37.8)	
Age, yr (mean $\pm$ SD)	64.7 (10.1)	43.0 (15.6)	<0.001
EQ5D-5L index value	0.70 (0.17)	0.85 (0.19)	<0.001

* Low education level: lower technical and vocational training and lower general secondary education; Medium education level: intermediate vocational training and advanced secondary education; High educational level: higher vocational education and university.

Preference for the Attribute Treatment Modality

Table 3 shows the results of the BWS analysis: the part-worth utilities could be interpreted relative to the selected reference level: often suffering from dyskinesia. The results show that treatment modality had the greatest impact on the perceived desirability of treatment in both the public and patient population (importance weight of 1) (Table 3). However, the patients attached higher importance to treatment modality in comparison to the public population (demonstrated by the lower attribute impacts for the other attributes). We did not have prior expectations with regard to the relative ordering of treatment modalities, yet the scatterplot with the rescaled BWS data (Figure 2) shows that the relative order of treatment modalities is similar in both populations, with neurosurgery being perceived as most negative by both patients and the public (coordinates 0,0). Noteworthy is that patients valued the infusion of medication via a pump more negative and the oral intake of medication slightly more positive compared to the public view.

Table 3. Results of mixed logit analysis for BWS case 2 – patient and general population.

Attribute	Levels	Patient N: 198		General population N: 276	
		Coefficient (SE)	Attribute impact (rank)	Coefficient (SE)	Attribute impact (rank)
Treatment modality	<i>Oral tablets</i>	3,51 (0.25)*	1.0 (1)	2,72 (0.16)*	1.0 (1)
	<i>Pump</i>	-1,41(0.26)*		1,23 (0.18)*	
	<i>Neurosurgery</i>	-4,93 (0.62)*		-2,06 (0.21)*	
Posture and balance problems	<i>Seldom to never</i>	3,47 (0.19)*	0.49 (2)	3,63 (0.13)*	0.84 (3)
	<i>Sometimes</i>	1,52 (0.20)*		1,24 (0.13)*	
	<i>Often</i>	-0,68 (0.23)*		-0,37 (0.14)*	
Slowness of movement	<i>Seldom to never</i>	3,44 (0.19)*	0.46 (3)	2,83 (0.13)*	0.49 (7)
	<i>Sometimes</i>	1,86 (0.20)*		1,66 (0.13)*	
	<i>Often</i>	-0,41 (0.23)*		0,49 (0.13)	
Tremor	<i>Seldom to never</i>	3,37 (0.20)*	0.42 (4)	3,45 (0.13)*	0.80 (4)
	<i>Sometimes</i>	1,8 (0.20)*		1,33 (0.13)*	
	<i>Often</i>	-0,21 (0.23)		-0,39 (0.13)*	
Dyskinesia	<i>Seldom to never</i>	3,05 (0.22)*	0.36 (5)	3,20 (0.13)*	0.67 (5)
	<i>Sometimes</i>	1,41(0.20)*		1,46 (0.13)*	
	<i>Often</i>	0 (reference)		0 (reference)	
Dizziness	<i>Seldom to never</i>	2,75 (0.20)*	0.32 (6)	3,51 (0.13)*	0.90 (2)
	<i>Sometimes</i>	1,53 (0.21)*		1,34 (0.13)*	
	<i>Often</i>	0,01 (0.22)*		-0,78 (0.13)	
Drowsiness	<i>Seldom to never</i>	2,83 (0.2)*	0.27 (7)	2,91 (0.13)*	0.58 (6)
	<i>Sometimes</i>	1,83 (0.20)*		1,90 (0.13)*	
	<i>Often</i>	0,57 (0.22)*		0,12 (0.13)	

* Coefficients differing significantly from the value of the reference-level $P < 0.05$.

Figure 2. Rescaled Best-Worst Scaling part-worth utilities for patient and public population.

If the rescaled coefficient is on or close to the center line in this graph, patients and public share the same opinion on the specific attribute-level (e.g. seldom to never suffer from drowsiness). The further the rescaled coefficient is from the center line, the more disagreement there exists between patients and public about this attribute-level (e.g. pump). The overall agreement between the two samples based on Pearson correlation coefficient is 0.897 ($P < 0.001$).

Preference for Clinical Outcomes

All health outcome attributes followed prior expectations: leading to decreasing scores as the level of impairment on each dimension increased. The two main differences in health outcomes between patient and public population were found in the perceived importance of the attributes slow movement and dizziness. Whereas for patients the effect of treatment on dizziness had less impact on their preference for treatment, the public found dizziness the second-most important outcome to consider in choosing treatment. Conversely, slow movement which is important to patients (3rd rank), was the least important according to the public (7th ranked). The differences in importance scores for the attributes drowsiness, tremor, posture and balance problems and dyskinesia were only minor.

Subgroup Analysis of the Public Sample

Fifty-five respondents (20%) were familiar with PD, because someone in their family or a close friend suffers from this condition. Notable here was that respondents who were familiar with PD attached higher relative importance to tremors compared to respondent who were not familiar (Figure 3A). Figure 3B shows that respondents younger than the age of 30 identified dizziness as the attribute which had the greatest impact on the perceived desirability of treatment instead of treatment modality (which was favoured by the older age groups). Furthermore, older respondents attached higher importance to dyskinesia, while the younger respondents attached higher importance to slow movement, dizziness and drowsiness. An interesting finding with regards to subgroups based on quality of life was that all respondents stated that treatment modality had the greatest impact on the perceived desirability of treatment (Figure 3C). However, respondents with a quality of life <0.75 attached relatively higher importance to treatment modality in comparison to respondents with higher quality of life scores (lower attribute impacts for other attributes).

Figure 3. Subgroup analysis of the preferences of public respondents. Subgroup analysis of the preferences of public respondents. 3A: familiarity with PD, 3B: age of the respondents, 3C: quality of life of the respondents.

3B **3C**

DISCUSSION

The first objective of this study was to compare patient and public preferences for clinical outcomes (symptom reduction, side effects) of PD treatment. Most health outcomes were valued equally by patients and public, except for reducing dizziness (more important to the public) and improving slow movement (more important to patients).

In economic theory it is assumed that importance of improving an outcome is directly linked to the perceived severity of that outcome. Thus, it seems that relative to patients, the public

underestimates the severity of slowness of movement and overestimates the severity of dizziness. A limitation of our experiment is that we did not quantify frequency of events. This might have influenced results, for instance because slowness of movement happens more frequent than dizziness in patients (and patients are aware of this), while dizziness is a more familiar experience to the public (28). Conversely, slow movement - which was described as difficulties with planning, initiating and executing movement - is an unfamiliar concept for non-PD patients and thus its impact on life is difficult to determine for public respondents.

The second objective of this study was to compare the impact of treatment modality on treatment desirability in the patient and public perspective. This study clearly shows that the process of care is an important driver of preferences, in both patients and in the general public. However, the patients attached higher relative importance to treatment modality than the public population. Especially, the infusion of medication via a pump was valued more negatively by patients. In contrast to our expectations and hypothesis, the treatment modality can outweigh the benefits of PD treatments. Patients with PD might have adapted to their condition (29) and are not willing to undergo advanced treatments in order to receive health improvements.

Our last objective was to explore whether public preferences differ based on sociodemographic variables or their familiarity with the disease. Subgroup analyses have shown that respondents who are familiar with PD attached higher importance to the resting tremor than respondents who are not familiar with PD. A possible explanation is that tremor is one of the most prominent and common symptoms of PD and is therefore easily noticed. Furthermore, older respondents and respondents with a lower quality of life (<0.75) have preferences that are more similar to the patient's preferences. They attach higher importance to the attribute treatment modality compared to younger respondents and respondents with a higher quality of life. Just as patients, they are more concerned about process optimization and less focused on symptom relief and prevention of side effects. Previous research has also shown that the expected burden of invasive treatments is an important reason for elderly patients to refuse medical interventions (30). Older respondents and respondents with declining health may find it is easier to relate to PD patients and consequently the preferences of these subgroups are more similar to the patient's preferences.

Implications of Results

Our study indicated differences between values of the patients and the general population. Although this study was not designed to recommend whether societal decision makers should add

patients values to their analysis, both public and patients' preferences entail information that is potentially relevant for societal decision makers (31, 32). For instance, knowing both perspectives can inform decision makers about the extent to which patients adapt to their condition, by comparing patient and public utility estimates. Moreover, in this study predicted uptake of invasive treatment is lower in patients compared to the public, which could indicate that desirability of a treatment is overestimated if its value is based on public preferences. These information could inform reimbursement and investment decision making (33). Since, patient and public preferences are not necessarily mutually exclusive, more initiatives are being employed to use patient preferences alongside public preferences in societal decision making (20, 34). The use of a stated preference survey in a representative sample of patients would be a valid method of doing so (35). The results presented in this article (and previous research (19)) suggest that there are indications for preference heterogeneity in the patient and population data. The tendency for preferences to be determined by socio-demographics and experience is not new and has been found in other research (16-18). For decision makers, the problem of preference heterogeneity can be added to the problem of whose values should count. More specifically: is preference heterogeneity relevant to measure, and if so, how will it impact decisions?

Strengths and Limitations

To our knowledge, this was the first study to compare patients' and public preferences of the relative importance of treatment aspects in PD. One difficulty encountered in pilot testing of the questionnaire was the cognitive difficulty of the choice task (especially for patients). In order to reduce cognitive burden we used colour coding of the attributes' levels, omitted quantitative frequency of symptoms and side-effects, and did not randomize the order of attributes within or over the tasks. While these decisions may have influenced the elicitation of 'true' preferences, it is likely that it's effect is similar in both populations. Furthermore, a limitation of our study was that public respondents received general information about PD, while patients considered their own situation. This may have led to differences in the perception of (the benefit of) treatments and to variance both within patients and between patients and the public respondents.

In the analysis of our BWS data we adopted the max diff model, in which it is assumed that respondents make a simultaneous choice of the pair of attribute-levels that maximizes the difference between them on a utility scale. However, respondents might have provided best-worst data in a particular order, which should be analyzed using a different model (sequential model of

best-worst choices) (21). Furthermore, we did not choose to conduct latent class analysis to explore subgroups, because several fit indices indicated a seven-class model which would be difficult to interpret given the aim of this study. Lastly, BWS experiments result in outcomes on a latent scale, therefore no direct comparisons can be made between two separate models without rescaling the data. In contrast to comparing two discrete choice models, no method is yet available to test for significant differences in utility estimations between two BWS models (36). This issue and rescaling the BWS data on the health utility (0-1) scale are important areas for future research, because the health utility scale is still favored in economic evaluations.

Conclusion

Our study has shown that patients and public judgments of the relative importance of treatment aspects in PD slightly differ. Patients attached higher importance to treatment modality compared to the public population. Most health outcomes were valued equally by patients and public, except for reducing dizziness (more important to the public) and improving slow movement (more important to patients). Public and patients' preferences entail information that is potentially relevant for decision makers. In a context of limited resources, knowing the patient's alongside public preferences for health outcomes and process of care can inform decision makers about actual patient priorities and thus effective allocation of limited resources.

REFERENCES

1. Lees AJ, Hardy J, Revesz T. Parkinson's disease. *Lancet*. 2009;373(9680):2055-66.
2. Martinez-Martin P, Kurtis MM. Health-related quality of life as an outcome variable in Parkinson's disease. *Ther Adv Neurol Disord*. 2012;5(2):105-17.
3. Schrag A, Jahanshahi M, Quinn N. What contributes to quality of life in patients with Parkinson's disease? *J Neurol Neurosurg Psychiatry*. 2000;69(3):308-12.
4. Clarke CE, Worth P, Grosset D, Stewart D. Systematic review of apomorphine infusion, levodopa infusion and deep brain stimulation in advanced Parkinson's disease. *Parkinsonism Relat Disord*. 2009;15(10):728-41.
5. Volkmann J, Albanese A, Antonini A, Chaudhuri KR, Clarke CE, de Bie RMA, et al. Selecting deep brain stimulation or infusion therapies in advanced Parkinson's disease: an evidence-based review. *J Neurol*. 2013;260(11):2701-14.
6. Banta D. The development of health technology assessment. *Health Policy*. 2003;63(2):121-32.
7. National Institute for Health and Clinical Excellence. Guide to the methods of health technology appraisal. 2013.
8. Zorginstituut Nederland. Guideline for economic evaluation in healthcare.: Zorginstituut Nederland;; 2016.
9. Stamuli E. Health outcomes in economic evaluation: who should value health? *Br Med Bull*. 2011;97:197-210.
10. Peeters Y, Stiggelbout AM. Health state valuations of patients and the general public analytically compared: a meta-analytical comparison of patient and population health state utilities. *Value Health*. 2010;13(2):306-9.

11. De Wit GA, Busschbach JJ, De Charro FT. Sensitivity and perspective in the valuation of health status: whose values count? *Health Econ.* 2000;9(2):109-26.
12. Coleman CI, Limone B, Sobieraj DM, Lee S, Roberts MS, Kaur R, et al. Dosing frequency and medication adherence in chronic disease. *J Manag Care Pharm.* 2012;18(7):527-39.
13. Brennan VK, Dixon S. Incorporating process utility into quality adjusted life years: a systematic review of empirical studies. *Pharmacoeconomics.* 2013;31(8):677-91.
14. Wertheimer AI, Radican L, Jacobs MR. Assessing different perspectives on the value of a pharmaceutical innovation. *Southern Med Review.* 2010;3(1):24-8.
15. Arbouw MEL, Movig KLL, Guchelaar HJ, Poels PJE, Van Vugt JPP, Neef C, et al. Discontinuation of ropinirole and pramipexole in patients with Parkinson's disease: Clinical practice versus clinical trials. *Eur J Clin Pharmacol.* 2008;64(10):1021-6.
16. Ryan M, Ubach C. Testing for an experience endowment effect in health care. *Appl Econ Lett.* 2003;10(7):407-10.
17. Dolan P. The effect of experience of illness on health state valuations. *J Clin Epidemiol.* 1996;49(5):551-64.
18. Schaarschmidt ML, Umar N, Schmieder A, Terris DD, Goebeler M, Goerdt S, et al. Patient preferences for psoriasis treatments: impact of treatment experience. *J Eur Acad Dermatol Venereol.* 2013;27(2):187-98.
19. Weernink MGM, van Til JA, van Vugt JPP, Movig KLL, Groothuis-Oudshoorn CGM, Ijzerman MJ. Involving Patients in Weighting Benefits and Harms of Treatment in Parkinson's Disease. *PloS one.* 2016;11(8):e0160771.
20. Versteegh MM, Brouwer WBF. Patient and general public preferences for health states: A call to reconsider current guidelines. *Soc Sci Med.* 2016;165:66-74.
21. Louviere JJ, Flynn TN, Marley AAJ. *Best-Worst Scaling Theory, Methods and Applications.* Cambridge: Cambridge University Press; 2015. 360 p.
22. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res.* 2011;20(10):1727-36.
23. Johnson FR, Lancsar E, Marshall D, Kilambi V, Mühlbacher A, Regier DA, et al. Constructing experimental designs for discrete-choice experiments: Report of the ISPOR conjoint analysis experimental design good research practices task force. *Value Health.* 2013;16(1):3-13.
24. Hensher DA, Greene WH. The mixed logit model: the state of practice. *Transportation.* 2003;30(2):133-76.
25. Lancsar E, Louviere J, Flynn T. Several methods to investigate relative attribute impact in stated preference experiments. *Soc Sci Med.* 2007;64(8):1738-53.
26. Weernink MGM, Groothuis-Oudshoorn CGM, Ijzerman MJ, van Til JA. Valuing Treatments for Parkinson Disease Incorporating Process Utility: Performance of Best-Worst Scaling, Time Trade-Off, and Visual Analogue Scales. *Value Health.* 2016;19(2):226-32.
27. Ricci F, De Caterina R, Fedorowski A. Orthostatic Hypotension: Epidemiology, Prognosis, and Treatment. *J Am Coll Cardiol.* 2015;66(7):848-60.
28. Dakof GA, Mendelsohn GA. Patterns of adaptation to Parkinson's disease. *Health Psychol.* 1989;8(3):355-72.
29. Rothman MD, Van Ness PH, O'Leary JR, Fried TR. Refusal of Medical and Surgical Interventions by Older Persons with Advanced Chronic Disease. *J Gen Intern Med.* 2007;22(7):982-7.
30. Drummond M, Brixner D, Gold M, Kind P, McGuire A, Nord E. Toward a consensus on the QALY. *Value Health.* 2009;12(SUPPL. 1):S31-S5.
31. Brazier J, Akehurst R, Brennan A, Dolan P, Claxton K, McCabe C, et al. Should patients have a greater role in valuing health states? *Appl Health Econ Health Policy.* 2005;4(4):201-8.
32. Dowie J, Kjer Kaltoft M, Salkeld G, Cunich M. Towards generic online multicriteria decision support in patient-centred health care. *Health Expect.* 2015;18(5):689-702.
33. Abelson J, Wagner F, DeJean D, Boesveld S, Gauvin FP, Bean S, et al. Public and Patient Involvement in Health Technology Assessment: A Framework for Action. *Int J Technol Assess Health Care.* 2016:1-9.

Chapter 4

34. Danner M, Hummel JM, Volz F, Van Manen JG, Wiegard B, Dintsios CM, et al. Integrating patients' views into health technology assessment: Analytic hierarchy process (AHP) as a method to elicit patient preferences. *Int J Technol Assess Health Care*. 2011;27(4):369-75.

35. Swait J, Louviere J. The role of the scale parameter in the estimation and comparison of multinomial logit models. *J Mark Res*. 1993:305-14.

5

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

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

Value Health. 2016;19(2):226-32. doi:10.1016/j.jval.2015.11.011.

ABSTRACT

Objective: The objective of this study was to compare treatment profiles including both health outcomes and process characteristics in Parkinson disease using best-worst scaling (BWS), time trade-off (TTO), and visual analogue scales (VAS).

Methods: From the model comprising of seven attributes with three levels, six unique profiles were selected representing process-related factors and health outcomes in Parkinson disease. A Web-based survey (N = 613) was conducted in a general population to estimate process-related utilities using profile-based BWS (case 2), multiprofile-based BWS (case 3), TTO, and VAS. The rank order of the six profiles was compared, convergent validity among methods was assessed, and individual analysis focused on the differentiation between pairs of profiles with methods used.

Results: The aggregated health-state utilities for the six treatment profiles were highly comparable for all methods and no rank reversals were identified. On the individual level, the convergent validity between all methods was strong; however, respondents differentiated less in the utility of closely related treatment profiles with a VAS or TTO than with BWS. For TTO and VAS, this resulted in nonsignificant differences in mean utilities for closely related treatment profiles.

Conclusions: This study suggests that all methods are equally able to measure process-related utility when the aim is to estimate the overall value of treatments. On an individual level, such as in shared decision making, BWS allows for better prioritization of treatment alternatives, especially if they are closely related. The decision-making problem and the need for explicit trade-off between attributes should determine the choice for a method.

INTRODUCTION

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

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

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

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

Standard gamble (SG), time trade-off (TTO), and visual analogue scales (VAS) are common methods to value health outcomes. SG is mostly used in economic evaluation studies and clinical decision analyses. SG is considered to be the "criterion standard" for utility measurement because it involves making choices under conditions of uncertainty and has rigorous foundations in expected utility

theory (11). Alternatively, TTO forces trade-offs between length of life and quality of life under conditions of certainty. TTO is well known for being the standard approach to derive health utility weights for EuroQol five-dimensional questionnaire (EQ-5D) health states (EQ-5D tariff) (12, 13). A VAS asks respondents to rate health states on a scale anchored at “worst imaginable health state” and “best imaginable health state” under conditions of certainty. They are easy and are often used in combination with the EQ-5D (14, 15).

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

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

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

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

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

METHODS

BWS Experiments

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

Treatment Profile Valuations

Six hypothetical core treatment profiles were defined: both extreme profiles and four intermediate (closely related) treatment profiles (Table 1). A full variation of the process attribute was not included in the six core scenarios because we also wanted to test the effect of small changes in health outcomes on utility scores. All respondents valued both the best and worst treatment profiles and were randomly assigned to value two of the four intermediate profiles with both a TTO and a VAS. The end points of the VAS were labeled as “best imaginable treatment” and “worst imaginable treatment.” A computer-based TTO was executed to determine how many years of life lived in perfect health followed by death would be equivalent to 10 years of life lived in the particular PD treatment profile followed by death. Respondents were asked to put themselves in the hypothetical scenarios and were asked whether they were willing to give up any life-years to return to full health again. If they did, a second question determined whether the respondent was willing to give up 5 life-years. This was followed by an iteration procedure in which the number of life-years was increased or decreased until the point of indifference was reached.

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

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

Sampling

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

Data Analysis

Respondents who completed the survey within 5 minutes were excluded from analysis. Furthermore, BWS data were checked using the Empirical Scale Parameter (ESP), which shows the extent to which people are consistent in their choices (29). Respondents were excluded if they either chose randomly ($ESP < 2$) or used an elimination-by-aspect strategy ($ESP > 5.5$). For some respondents the profiles were all of equal value, the so-called nondifferentiators. Nondifferentiators include respondents who did not want to trade any life-years (TTO), respondents who traded the same number of life-years for each of the four evaluated treatment profiles (TTO), and respondents who assigned equal values to each profile on the scale from 0 to 100 (VAS). Inherent to the BWS method, respondents were forced to make choices, which in principle leads to a difference in utilities for the treatment profiles. Nondifferentiators were excluded from the main analysis to investigate whether the remainder of respondents, as a group and as individuals, made significant differences in value between treatment profiles. The main analysis consisted of two components: 1) comparison of aggregated utility scores and 2) comparison of individual scores (within-person comparison). From the final data set, aggregated TTO and VAS utilities for each of the six treatment profiles were calculated using the mean. BWS-2 and BWS-3 utility weights (β) for each attribute level were estimated using a conditional logit model. These β weights allow for utility estimation of the six selected treatment profiles. Because of the latent scale of BWS, the derived BWS-2 and BWS-3 utility weights could not be directly compared with each other and compared with TTO/VAS utilities. Therefore, a rank order

comparison was used to observe whether the same order from best to worst treatment profile existed for each method. The convergent validity between the aggregated utilities of all methods was assessed using the squared Pearson correlation coefficient (across-respondents means) (30). To study whether the hypothesis that VAS and TTO use a smaller part of their scale to value small changes in treatments was true, it was calculated which part of each method's scale was needed to differentiate between the value of the worst, intermediate, and best profiles. Furthermore, the importance of each attribute according to BWS-2 and BWS-3 was compared by estimating the difference between minimum and maximum β weights within an attribute divided by the total variance summed across all attributes. In addition, the relative size of BWS-2 and BWS-3 differences was adjusted by using TTO and VAS valuations of the best and worst treatment profiles to "anchor" BWS valuations on the VAS/TTO scale of 0 to 1 (31). Subsequently, scatterplots were used to identify (scaling) differences and outliers. Individual utilities were analyzed to study whether the respondents were able to differentiate between the values of the six treatment profiles with all methods. Mixed logit models were used to estimate the individual BWS-2 and BWS-3 utility weights for the best and the worst profiles, and for the two intermediate profiles for which the respondent also has provided a TTO and VAS valuation. For each pair of treatment profiles, the number of nondifferentiators was identified. The difference with previously excluded nondifferentiators is that these respondents at least made one differentiation between the four evaluated profiles; for example, respondents might have indicated a difference between the best and the worst profile but not between two intermediate profiles. Individual differences between all pairs of two treatment profiles were calculated and tested with a paired t test. The convergent validity between the individual utilities was studied using the squared Pearson correlation coefficient (within-respondents mean) (30). Data were analyzed using Stata version 13 (StataCorp, College Station, TX). P values of less than 0.05 were considered statistically significant.

RESULTS

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

Nondifferentiators

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

Aggregated Utility Scores of Hypothetical Profiles

Table 2 presents the aggregated mean utility for each of the treatment profiles. Each of the mean utility scores follows a monotonic relationship: both VAS and TTO utilities increase if BWS utilities increase. There were no rank reversals, and the group of respondents was able to distinguish between the best, the worst, and all intermediate profiles. Squared Pearson correlation coefficients also show that the convergent validity between all methods is very strong (VAS-BWS-2 0.97; VAS-BWS-3 0.98; TTO-BWS-2 0.97; TTO-BWS-3 0.98; TTO-VAS 0.95; BWS-2-BWS-3 0.98; $P = 0.000$, $N = 6$). Overall, respondents gave higher TTO utilities for treatment profiles compared with VAS utilities. In TTO, respondents used 49% $[(0.88 - 0.59)/(0.93 - 0.34)]$ of the scale to differentiate between the four intermediate treatment profiles, whereas with BWS-2 this was only 30%, with BWS-3 41%, and with VAS 32%. The largest difference in utility was found between the utility of the worst profile and the utility of intermediate treatment 1: BWS-2 50%, BWS-3 49%, VAS 43%, and TTO 42%. According to the attribute importance calculated with both BWS-2 and BWS-3, process of care had the largest impact on overall treatment valuation in patients with PD. When the results of BWS-2 and BWS-3 were compared, there was a difference in ranking of attributes: there were rank reversals between the attributes dizziness and resting tremor and the attributes dyskinesia and drowsiness (Fig. 1). The rescaled BWS data (Fig. 2A 2B) show that the shape of both BWS-2 and BWS-3 utilities is similar, except for the fact that respondents assigned slightly higher utilities to treatment profiles with BWS-3 than with BWS-2. Furthermore, the scatterplots show that TTO utilities for intermediate profiles were more in line with BWS_{TTO} rescaled utilities (2A) than with VAS utilities compared with BWS_{VAS} rescaled utilities (2B).

Table 2. Mean-utilities for the six core treatment profiles (mean/SE).

	TTO	VAS	BWS-2	BWS-3
Worst*	0.34 (0.01)	0.25 (0.01)	-2.41 (0.85)	-1.10 (0.04)
Intermediate 1**	0.59 (0.02)	0.48 (0.01)	11.16 (0.83)	1.75 (0.07)
Intermediate 2**	0.69 (0.02)	0.52 (0.01)	14.5 (0.84)	2.22 (0.07)
Intermediate 3**	0.78 (0.02)	0.55 (0.01)	17.24 (0.97)	2.65 (0.08)
Intermediate 4**	0.88 (0.02)	0.65 (0.01)	19.41 (0.98)	4.17 (0.11)
Best*	0.93 (0.01)	0.78 (0.01)	24.89 (0.99)	4.74 (0.12)

* 422 observations

** 200 observations (each respondent only valued 2 of the 4 intermediate treatment profiles)

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

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

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

Individual Scores of Hypothetical Profiles (within-Person Comparison)

Table 3 presents the mean difference in individual utilities for all pairs of treatment profiles. In addition, the number of respondents who did not differentiate the utility of pairs of profiles is presented for TTO and VAS. Overall, when respondents used a TTO to state their preference for the treatment alternatives, 30% of 2477 pairs of treatment profiles did not result in a preference for one of the two profiles. Conversely, for VAS valuations this was only 2.5%. Because of the forced choice, all respondents stated their preference for treatment profiles with BWS-2 and BWS-3. Most respondents gave a higher preference to the best-case scenario than to the worst-case scenario with all methods. Only a few respondents (TTO 5%; VAS 1%) were indifferent between the extreme profiles, but gave different valuations for at least one other pair of treatment profiles. More respondents were indifferent between the value of pairs of intermediate profiles (TTO 33%–58%; VAS 1%–13%). As a result, only 8% of respondents had a ranking of treatment profiles that was similar to the overall ranking with TTO (VAS 56%; BWS-2 94%; BWS-3 91%). Furthermore, respondents were not able to make a significant difference in value between intermediate profiles 3 and 2 with both TTO and VAS, and intermediate profiles 2 and 1 (only TTO). For all methods, the mean difference in estimated utility between the intermediate profiles and either extreme profile decreased when the intermediate profile was more similar to one of the extreme profiles. At the

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

Profiles	N	TTO Mean-difference (SE)	TTO non- differentiators N (%)	VAS Mean-difference (SE)
<i>Best & worst</i>	422	0.59 (0.016)*	20 (5%)	0.53 (0.014)*
<i>Best & intermediate 4</i>	215	0.06 (0.015)*	154 (72%)	0.14 (0.010)*
<i>Best & intermediate 3</i>	199	0.14 (0.024)*	107 (54%)	0.22 (0.013)*
<i>Best & intermediate 2</i>	204	0.24 (0.025)*	93 (46%)	0.25 (0.012)*
<i>Best & intermediate 1</i>	204	0.35 (0.024)*	68 (33%)	0.30 (0.014)*
<i>Intermediate 4 & 1</i>	72	0.34 (0.039)*	24 (33%)	0.18 (0.018)*
<i>Intermediate 4 & 2</i>	70	0.18 (0.039)*	34 (49%)	0.14 (0.017)*
<i>Intermediate 4 & 3</i>	62	0.22 (0.044)*	40 (55%)	0.06 (0.016)*
<i>Intermediate 3 & 1</i>	73	0.08 (0.040)*	21 (34%)	0.11 (0.018)*
<i>Intermediate 3 & 2</i>	64	0.05 (0.043)	37 (58%)	0.01 (0.018)
<i>Intermediate 2 & 1</i>	70	0.06 (0.033)	37 (53%)	0.07 (0.015)*
<i>Intermediate 4 & worst</i>	215	0.55 (0.023)*	14 (7%)	0.42 (0.016)*
<i>Intermediate 3 & worst</i>	199	0.45 (0.024)*	21 (11%)	0.29 (0.017)*
<i>Intermediate 2 & worst</i>	204	0.35 (0.025)*	32 (16%)	0.27 (0.017)*
<i>Intermediate 1 & worst</i>	204	0.23 (0.022)*	51 (25%)	0.22 (0.014)*

Profiles	N	VAS non- differentiators N (%)	BWS-2 Mean-difference (SE)	BWS-3 Mean-difference (SE)
<i>Best & worst</i>	422	4 (1%)	34.58 (0.280)*	7.47 (0.064)*
<i>Best & intermediate 4</i>	215	7 (3%)	6.29 (0.028)*	0.78 (0.029)*
<i>Best & intermediate 3</i>	199	5 (3%)	9.84 (0.150)*	2.71 (0.027)*
<i>Best & intermediate 2</i>	204	3 (1%)	12.65 (0.102)*	3.15 (0.030)*
<i>Best & intermediate 1</i>	204	3 (1%)	16.91 (0.125)*	3.68 (0.034)*
<i>Intermediate 4 & 1</i>	72	1 (1%)	10.40 (0.185)*	2.88 (0.062)*
<i>Intermediate 4 & 2</i>	70	3 (4%)	6.45 (0.180)*	2.40 (0.057)*
<i>Intermediate 4 & 3</i>	62	5 (7%)	6.60 (0.353)*	0.99 (0.060)*
<i>Intermediate 3 & 1</i>	73	6 (10%)	3.41 (0.218)*	1.96 (0.042)*
<i>Intermediate 3 & 2</i>	64	5 (8%)	2.93 (0.368)*	0.45 (0.067)*
<i>Intermediate 2 & 1</i>	70	9 (13%)	4.34 (0.167)*	0.57 (0.030)*
<i>Intermediate 4 & worst</i>	215	3 (1%)	28.13 (0.379)*	6.65 (0.087)*
<i>Intermediate 3 & worst</i>	199	1 (1%)	24.35 (0.383)*	4.86 (0.092)*
<i>Intermediate 2 & worst</i>	204	4 (2%)	21.78 (0.404)*	4.18 (0.083)*
<i>Intermediate 1 & worst</i>	204	3 (1%)	17.95 (0.363)*	3.81 (0.092)*

* p value < 0.05

same time, the rate of nondifferentiators for TTO increased when the intermediate profile was more similar to one of the extreme profiles (not observed for VAS). Squared Pearson mean within-respondent correlation showed that the convergent validity between all methods is strong (VAS-BWS-2 0.67; VAS-BWS-3 0.68; TTO-BWS-2 0.58; TTO-BWS-3 0.59; TTO-VAS 0.56; BWS-2-BWS-3 0.97; $P = 0.000$, $N = 422$).

DISCUSSION

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

With the VAS, most of the respondents in our study were able to differentiate between the value of treatment profiles. The results of a treatment valuation with VAS allows for prioritization of outcomes and treatment choice on the individual level. A VAS can be used in clinical decision making if there is only a need to inform on relative treatment preferences (e.g., choice for treatment A vs. treatment B). VAS is a simple and cognitively undemanding method. However, it does not involve uncertainty or any trade-off between gains and possible harms (i.e., length of life, higher risk of). Because this is not in agreement with utility theory, the use of a VAS as a valuation tool is not useful in economic evaluations (32).

With TTO, 25% of respondents were not willing to trade-off life-years with any of the treatment scenarios. This might have resulted from the unrealistic choice a TTO poses, religious beliefs, or the abstract reasoning of life and death. Yet, 25% of non-differentiators is not seen as a high rate in

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

Inherent to the method, BWS is more able to disaggregate the health-state utility to a single attribute level, providing evidence on those characteristics relevant to respondents. The results of BWS allow for estimation of treatment valuation on the individual level, thereby supporting clinical decision making. Compared with the number of nondifferentiators in TTO, more respondents engaged with the BWS task. Although VAS is a simple method, BWS overcomes the measurement bias of VAS by using discrete choices and trading of characteristics (40). For the use of BWS as a valuation tool on a societal level, the method lacks scaling of the treatment utilities on a general, disease-exceeding level (health utility scale 0–1). Several options are available to rescale BWS data on the quality-of-life scale (24, 31). However, as long as a sound methodological framework for mapping DCE/BWS utilities is lacking, SG and TTO will remain the criterion standard. The establishment of an ISPOR task force on mapping utilities with the aim to develop good practice guidelines can be seen as an important step toward the use of BWS and DCE in health technology assessment (25). However, SG continues to have the advantage of letting respondents make choices under conditions of uncertainty. Furthermore, our results have shown that utilities derived from BWS-2 or BWS-3 are highly comparable. Our recommendation is that the characteristics of the respondents (such as cognitive state) should influence the choice for a specific BWS case.

The results of this study also showed that the process of care is an attribute that was relevant to respondents. Whether or not it is relevant to include it in a measure of health depends on its

contexts. We feel it is very relevant to include in clinical decision making. However, in societal decision making it is debatable whether valuations should include process factors (6, 8). Yet, the Belgium Health Care Knowledge Centre has included discomfort/ inconvenience of current treatment as an attribute in a DCE to measure societal preferences in reimbursement decisions in Belgium (41).

Strengths and Limitations

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

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

Conclusions

The value of treatments in chronic diseases such as PD greatly depends on both clinical outcomes and process characteristics. If one wants to explicitly include process characteristics in scenarios, the decision-making problem, and the need for explicit trade-off between attributes, should

determine the choice for a method to measure process-related utility. We conclude that on a societal level, when group preferences are relevant, BWS, TTO, and VAS perform equally well with regard to distinguishing treatment alternatives on the basis of process-related utility. In clinical decision making, BWS allows for better prioritization of treatment alternatives and provides evidence on those characteristics that are most important to respondents.

REFERENCES

1. Beaglehole R, Ebrahim S, Reddy S, Voûte J, Leeder S. Prevention of chronic diseases: a call to action. *Lancet*. 2007;370(9605):2152-7.
2. Stewart AL, Greenfield S, Hays RD, Wells K, Rogers WH, Berry SD, et al. Functional status and well-being of patients with chronic conditions. Results from the Medical Outcomes Study. *JAMA*. 1989;262(7):907-13.
3. Coleman CI, Limone B, Sobieraj DM, Lee S, Roberts MS, Kaur R, et al. Dosing frequency and medication adherence in chronic disease. *J Manag Care Pharm*. 2012;18(7):527-39.
4. Brennan VK, Dixon S. Incorporating process utility into quality adjusted life years: a systematic review of empirical studies. *PharmacoEconomics*. 2013;31(8):677-91.
5. Dunbar-Jacob J, Erlen JA, Schlenk EA, Ryan CM, Sereika SM, Doswell WM. Adherence in chronic disease. *Annu Rev Nurs Res*. 2000;18:48-90.
6. Ryan M, Kinghorn P, Entwistle VA, Francis JJ. Valuing patients' experiences of healthcare processes: towards broader applications of existing methods. *Soc Sci Med*. 2014;106:194-203.
7. Ryan M, Gerard K. Inclusiveness in the health economic evaluation space. *Social Science & Medicine*. 2014;108(0):248-51.
8. Culyer AJ, Newhouse JP. *Handbook of Health Economics*: Elsevier Science; 2000.
9. Mooney G. Communitarian claims and community capabilities: furthering priority setting? *Soc Sci Med*. 2005;60(2):247-55.
10. Donaldson C, Shackley P. Does "process utility" exist? A case study of willingness to pay for laparoscopic cholecystectomy. *Soc Sci Med*. 1997;44(5):699-707.
11. von Neumann J, Morgenstern O. *Theory of games and economic behavior*. Princeton (NJ): Princeton University Press; 1944.
12. Viney R, Norman R, King MT, Cronin P, Street DJ, Knox S, et al. Time trade-off derived EQ-5D weights for Australia. *Value Health*. 2011;14(6):928-36.
13. Longworth L, Rowen D. Mapping to Obtain EQ-5D Utility Values for Use in NICE Health Technology Assessments. *Value Health*. 2013;16(1):202-10.
14. Gafni A. The standard gamble method: what is being measured and how it is interpreted. *Health Serv Res*. 1994;29(2):207-24.
15. Weernink MGM, Janus SLM, van Til JA, Raich DW, van Manen JG, IJzerman MJ. A Systematic Review to Identify the Use of Preference Elicitation Methods in Healthcare Decision Making. *Pharm Med*. 2014;28(4):175-85.
16. Flynn TN. Valuing citizen and patient preferences in health: recent developments in three types of best-worst scaling. *Expert Rev Pharmacoecon Outcomes Res*. 2010;10(3):259-67.
17. Flynn TN, Louviere JJ, Peters TJ, Coast J. Best-worst scaling: What it can do for health care research and how to do it. *J Health Econ*. 2007;26(1):171-89.
18. Ali S, Ronaldson S. Ordinal preference elicitation methods in health economics and health services research: using discrete choice experiments and ranking methods. *British medical bulletin*. 2012;103(1):21-44.
19. Louviere JJ, Woodworth GG. *Best-Worst Scaling: A Model for Largest Difference Judgments*. University of Alberta: Working Paper, Faculty of Business. 1990.
20. Rowen D, Brazier J, Van Hout B. A comparison of methods for converting DCE values onto the full health-dead QALY scale. *Med Decis Making*. 2015;35(3):328-40.
21. Ratcliffe J, Couzner L, Flynn T, Sawyer M, Stevens K, Brazier J, et al. Valuing child health utility 9D health states with a young adolescent sample: A feasibility study to compare best-worst scaling discrete-choice experiment, standard gamble and time trade-off methods. *Applied*

- Health Economics and Health Policy. 2011;9(1):15-27.
22. Salomon JA. Reconsidering the use of rankings in the valuation of health states: a model for estimating cardinal values from ordinal data. *Popul Health Metr.* 2003;1(1):12.
23. Brazier J, Rowen D, Yang Y, Tsuchiya A. Comparison of health state utility values derived using time trade-off, rank and discrete choice data anchored on the full health-dead scale. *The European journal of health economics : HEPAC : health economics in prevention and care.* 2012;13(5):575-87.
24. Flynn T, Louviere J, Marley A, Coast J, Peters T. Rescaling quality of life values from discrete choice experiments for use as QALYs: a cautionary tale. *Popul Health Metr.* 2008;6(1):6.
25. International Society for Pharmacoeconomics and Outcomes Research. Mapping to estimate health state utility values from non-preference based outcomes measures for cost per QALY economic analysis - Good practices task force 2014 [cited 2015 27-10-2015]. Available from: <http://www.ispor.org/taskforces/mapping-utility-values-from-non-pref-based-outcomes-grp.asp>.
26. Bijlenga D, Birnie E, Bonsel GJ. Feasibility, Reliability, and Validity of Three Health-State Valuation Methods Using Multiple-Outcome Vignettes on Moderate-Risk Pregnancy at Term. *Value Health.* 2009;12(5):821-7.
27. Johnson FR, Lancsar E, Marshall D, Kilambi V, Mühlbacher A, Regier DA, et al. Constructing experimental designs for discrete-choice experiments: Report of the ISPOR conjoint analysis experimental design good research practices task force. *Value Health.* 2013;16(1):3-13.
28. Sawtooth Software. Sawtooth Software. Orem, UT: Sawtooth Software; 2009.
29. Swait J, Louviere J. The role of the scale parameter in the estimation and comparison of multinomial logit models. *J Mark Res.* 1993:305-14.
30. Nickerson CAE. Assessing Convergent Validity of Health-state Utilities Obtained Using Different Scaling Methods. *Med Decis Making.* 1999;19(4):487-98.
31. Gu NY, Botteman MF, Gerber RA, Ji X, Postema R, Wan Y, et al. Eliciting health state utilities for Dupuytren's contracture using a discrete choice experiment. *Acta Orthop.* 2013;84(6):571-8.
32. Parkin D, Devlin N. Is there a case for using visual analogue scale valuations in cost-utility analysis? *Health Econ.* 2006;15(7):653-64.
33. Arnesen T, Trommald M. Roughly right or precisely wrong? Systematic review of quality-of-life weights elicited with the time trade-off method. *Journal of health services research & policy.* 2004;9(1):43-50.
34. Boye KS, Matza LS, Feeny DH, Johnston JA, Bowman L, Jordan JB. Challenges to time trade-off utility assessment methods: when should you consider alternative approaches? *Expert Rev Pharmacoecon Outcomes Res.* 2014;14(3):437-50.
35. Attema AE, Edelaar-Peters Y, Versteegh MM, Stolk EA. Time trade-off: one methodology, different methods. *The European journal of health economics : HEPAC : health economics in prevention and care.* 2013;14:53-64.
36. Morimoto T, Shimbo T, Orav JE, Matsui K, Goto M, Takemura M, et al. Impact of social functioning and vitality on preference for life in patients with Parkinson's disease. *Mov Disord.* 2003;18(2):171-5.
37. Palmer CS, Schmier JK, Snyder E, Scott B. Patient preferences and utilities for 'off-time' outcomes in the treatment of Parkinson's disease. *Qual Life Res.* 2000;9(7):819-27.
38. Gray A, McNamara I, Aziz T, Gregory R, Bain P, Wilson J, et al. Quality of life outcomes following surgical treatment of Parkinson's disease. *Mov Disord.* 2002;17(1):68-75.
39. Torrance GW, Feeny D. Utilities and quality-adjusted life years. *Int J Technol Assess Health Care.* 1989;5(4):559-75.
40. Torrance GW, Feeny D, Furlong W. Visual analog scales: Do they have a role in the measurement of preferences for health states? *Med Decis Making.* 2001;21(4):329-34.
41. Cleemput I, Devriese S, Kohn L, Devos C, van Til J, Groothuis-Oudshoorn K, et al. Incorporating societal preferences in reimbursement decisions – Relative importance of decision criteria according to Belgian citizens. Brussels: Belgian Health Care Knowledge Centre (KCE), 2014 Contract No.: KCE Reports 234.
42. Norman R, King MT, Clarke D, Viney R, Cronin P, Street D. Does mode of administration matter? Comparison of online and face-to-face administration of a time trade-off task. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation.* 2010;19(4):499-508.

APPENDIX A

Example Best-Worst Scaling Case 2 – profile task

Most desirable characteristic	What do you find the...	Least desirable characteristic
Treatment profile 1
☐	I receive medication via tablets	☐
☐	I often suffer from tremors	☒
☒	I seldom to never suffer from posture and balance problems	☐
☐	I seldom to never suffer from slowness in motion	☐
☐	I often suffer from dizziness	☐
☐	I sometimes suffer from drowsiness	☐
☐	I often suffer from rapid uncontrolled movements	☐

Example Best-Worst Scaling Case 3 – choice task

	Treatment 1	Treatment 2	Treatment 3
I receive medication via tablets	I receive medication via pump infusion	I have to undergo brain surgery
I often suffer from tremors	I sometimes suffer from tremors	I often suffer from tremors
I seldom to never suffer from posture and balance problems	I sometimes suffer from posture and balance problems	I often suffer from posture and balance problems
I seldom to never suffer from slowness in motion	I often suffer from slowness in motion	I seldom to never suffer from slowness in motion
I often suffer from dizziness	I often suffer from dizziness	I seldom to never suffer from dizziness
I sometimes suffer from drowsiness	I sometimes suffer from drowsiness	I sometimes suffer from drowsiness
I often suffer from rapid uncontrolled movements	I seldom to never suffer from rapid uncontrolled movements	I often suffer from rapid uncontrolled movements
The most desirable treatment	☒	☐	☐
The least desirable treatment	☐	☐	☒

6

Exploration of Values Clarification Methods based on Conjoint Analysis Techniques: A Systematic Review

Marieke G.M Weernink

Janine A. van Til

Holly O. Witteman

Liana Fraenkel

Maarten J. IJzerman

Submitted

ABSTRACT

Background: The majority of conjoint analysis (CA) studies are undertaken to support organisational and regulatory decision making, but in shared decision making, individual preferences are the main interest. Our objective was to systematically review the use of CA to elicit preferences for clinical decision support at the individual patient level. We aimed to identify design features regarding the selection of attributes and levels, choice tasks, and methods used to clarify values.

Methods: We searched Scopus, Pubmed, PsycInfo and Web of Science to identify studies which developed a CA-exercise to elicit individual patient's preferences related to medical decisions. We extracted data on the above-mentioned items.

Results: Eight studies were identified. All used adaptive conjoint analysis (ACA). Most studies resembled each other in design choices made, but patients received different feedback across studies. All studies included a fixed set of attributes. Across studies, most patients reported that the ACA exercise was easy to do, clear and interesting. However, patients who were less well educated and unfamiliar with computers had greater difficulty performing the ACA-task. Furthermore, time and effort needed from clinicians to facilitate the ACA exercise was identified as the main barrier to implementation.

Discussion: In this review, all included studies used ACA. ACA has potential to enhance shared decision making, but for these tools to succeed, further research should first focus on the feedback patients want to or should receive, the interchangeability of attributes and levels in the exercise, and how best to incorporate CA exercise within the clinic workflow.

INTRODUCTION

Conjoint Analysis (CA) has a long history in marketing, and has gained widespread use as a tool to elicit patient preferences for health care services (1, 2). CA allows patients to think about complex treatment decisions by letting them evaluate scenarios through rating, ranking or choice tasks, enabling a mathematical model to algorithmically derive the relative value of treatment characteristics (3). The majority of CA studies are undertaken to elicit preferences at the population level to support organisational and regulatory decision making (4). As Kaltoft et al. (2015) argued, the results of these studies have limited clinical relevance to individual patients in decision making (5). What is important to one patient may not be the same as what is important to others. In shared decision making, approaches used must be able to generate part-worth utilities at the individual patient level.

In shared decision making, patient decision aids are often used to support patients' understanding of the process of care and the subsequent evidence-based outcomes (6). Values clarification methods are used in decision aids to help patients evaluate the desirability of attributes or options, with the aim that choice of treatment reflects personal preferences and values (7). Based on its theoretical axioms (8), CA may be an effective value clarification method because it can present patients with estimates of the relative impact of each attribute on their decision as well as provide an estimate of their preferences for available treatment options (9). However, a review by Witteman et al. (2016) showed that values clarification methods such as rating scales or providing the pros and cons were most commonly used. Furthermore, only 38% of the articles found described a values clarification method explicitly or implicitly based on any theory, framework, model or mechanism (10).

In this study, we systematically review published studies describing the use of CA to elicit preferences for clinical decision support at the individual patient level. We aim to identify actual design decisions (theoretical and practical considerations) with regard to the selection of attributes and levels, the design of choice tasks and the instrument used to clarify values. Finally, we present how the use of CA was evaluated in these studies.

METHODS

Search Strategy and Screening Process

We conducted a systematic literature search in Scopus, Pubmed, PsycInfo and Web of Science in May 2016 to identify studies which have described the use of CA to elicit preferences for clinical decision support at the individual patient level. Studies were included if they contained sufficient information for data extraction on the relevant criteria detailed below. Appendix A contains a full description of the search strategy. Next, two authors (MW, JT) independently screened all articles. Primary exclusion was based on title, abstract and keywords. Subsequently, potentially relevant articles were reviewed in full. Discrepancies were resolved by discussion between MW and JT until consensus was reached.

Extraction of Relevant Data

For each included study, two independent researchers (MW, JT) systematically extracted data using a standardized abstraction form. The extracted data included five categories of information: (1) general study information, (2) selection of attributes and levels, (3) choice task design, (4) instrument design and (5) study outcomes / evaluation.

1. We derived general information of the included studies such as its aim, type and clinical condition.
2. We reported design decisions regarding the selection of attributes and levels: methods used, number of attributes/levels, possibility for the patient to add extra attributes and the type of attributes selected such as health outcomes, non-health outcomes (e.g. information provision, autonomy) and process attributes (e.g. treatment modality, waiting time) (11, 12).
3. Next, choice task design elements were derived from the studies. Depending on the method of CA used, some of the choice task design components are more or less applicable to each study: type of choice questions (rate, rank, choice), experimental design, and number of choice tasks.
4. We then described the instrument used to clarify values: the medium and setting used, the preference estimation procedure (analytical method, use of software packages), and whether and how CA results were presented to patients (form and setting).
5. Lastly, we examined the study outcomes and evaluation of the CA-exercise: reported positive and negative outcomes with respect to difficulty of CA-exercise, reliability of its outcomes, behavioral outcomes (actions, adherence) and affective and cognitive outcomes (knowledge, decisional conflict, satisfaction).

Included studies had to present sufficient data on criterion 1 to 3, but not on 4 and 5 (in the case of clinical trial protocols). We derived information about design choices from the main text, the appendix, or the decision tool itself if it was accessible on a website. If information about the design characteristics was lacking or insufficiently described, we requested the authors to send a copy of the tool used. Descriptive summaries and graphs were used for all categories, except for point 5, where narrative synthesis methods were used.

RESULTS

Search Results

In total 1057 articles were screened, 994 articles were excluded by title and abstract review. The remaining 63 articles were subjected to full-text review. This led to exclusion of an additional 55 articles, mainly because the focus of the article was elicitation of preferences at the population level. In total, eight articles used CA to support preference elicitation in clinical decisions in individual patients (Figure 1).

Figure 1. Selection of studies

Characteristics of Included Studies

The included articles discussed the application of CA in curative treatment decisions (e.g. breast cancer and prostate cancer) and treatment decisions for chronic conditions (e.g. osteoporosis). The focus of most studies was to examine the use of CA to support preference elicitation in a clinical setting. One article only discussed the clinical trial protocol and not results (13). One study compared the additional value of an interactive CA-exercise to a printed booklet and a video booklet on reducing decision conflict (14) and Pieterse et al. (2010) (15) examined whether the mode of administration (Web-based, local computer) influenced importance scores. We concluded that the studies of Fraenkel (2010) (16) and Rochon et al. (2012) (17) described the development of the same values clarification exercise and are therefore discussed as one study. Appendix B shows the complete evidence matrix for all five categories of extracted data.

Selection of Attributes and Levels

In our review, we found a range of practices to identify attributes including literature review (15, 18), consulting clinicians (14) or more in-depth qualitative research techniques such as interviews and/or focus groups with patients and/or clinicians (13, 19, 20). In most studies the final selection of attributes was made by the researchers or a facilitated discussion with the project team. Only one study specifically stated that the patients had to rank and rate the attributes, from which the top attributes were determined (20). None of the studies allowed individual patients the possibility to add (or remove) attributes prior to the actual start of the CA-exercise.

Each study included attributes concerning health outcomes and often a balance was found between benefits and harms (Figure 2). In addition, almost all studies included a measure of process of care e.g. treatment modality (18), costs (16), or days in hospital (14). The studies selected between four and eight attributes for inclusion in the CA-exercise. Attribute-levels reflect the range of actual variation within attributes. Across all studies, 7 out of 34 attributes were two-level attributes 18 were three-level attributes and 9 were four-level attributes. Three studies had equal levels over all attributes (14, 17, 20), while two studies had a combination of two and three-level attributes (15, 19), and for two studies the number of levels over all attributes was unknown.

Most of the articles reported that pilot tests were conducted to test the feasibility of the exercise. Patient understanding was enhanced by using frequencies (graphs) to describe risks (1 in 10, 1 in 1000) (14, 15, 17, 19) or visual pictographs of attributes or icon arrays (18, 19).

Figure 2. Type of attributes selected in the studies.

* The study of Jayadevappa et al. (2015) did not reported the number of selected attributes.

Choice Task Design

All included studies used Adaptive Conjoint Analysis (ACA). ACA is an interactive computerized way of using CA that enables patients to construct their preferences through self-explicated rating and ranking tasks followed by pairwise comparisons. In the first series of questions, patients are asked to rank the levels of the attributes that do not have a natural ranking. In the reviewed studies, these were primarily process attributes. ACA allows patients to handle a large number of attributes while avoiding information overload. The highest number found in this review was a study with eight attributes, each having four levels (14).

In the second set of questions, patients are asked to rate the importance of moving from the most preferred level to the least preferred level of each attribute on a Likert-type scale. However, three studies used a modified (simplified) version of ACA which has been developed by Fraenkel (2010). It asks patients to first choose the attribute which is most important to them. Next, they rate the remaining attributes relative to the one they indicated as being the most important one (16).

Both the original ACA and modified ACA are followed by a series of pairwise comparisons in which patients have to use a nine-point Likert scale to indicate to which extent one scenario is preferred to the other. ACA is interactive, which means that after each pairwise comparison the utility estimates are updated through regression analysis and a new pair of scenarios is selected which are estimated to be of equal utility. The number of paired comparisons across the ACA studies

ranged from 12 to 18 and is determined prior to the preference elicitation (when using Sawtooth Software). The formula which is used to determine the minimum number of paired comparisons is: $3 * (N - n - 1) - N$, where N is the number of levels across all attributes and n is the number of attributes (15). Also, the number of attributes in each subset can be chosen by the investigator. Pieterse et al. (2010) included two out of the four attributes in pairwise comparisons 1-4, three attributes in pairs 6-10 and only included all attributes in final pairs 11 and 12.

Instrument Design

All studies used Sawtooth software to collect and analyse the data. All ACA studies were conducted via a computer, because the pairwise comparisons had to be adapted to the patient's previous answers.

Three ACA studies were web-based, and patients received a link to a website containing the ACA (20) or a direct link to the exercise (13, 15). Four studies used CAPI-based interviews (computers not connected to the Internet) to elicit patient preferences. These patients had to visit the hospital or the research facility to complete the exercise, where in one case patients received prior oral explanation (19) and in another study a brief training session (14). In a few cases, a research assistant was present during the exercise to answer procedural questions (14, 18, 19). Based on the self-explicated rating and ranking tasks and the pairwise comparisons a final set of part-worth utilities is estimated using ordinary least square regression analysis. All studies provided patients with the output in real-time, immediately after the exercise was performed.

Three slightly different formats for feedback on attribute importance were found (Figure 3: reconstructions based on the information in the articles). Figure 3A and 3B are comparable in their format: a longer bar indicates higher importance of an attribute. The difference is in the description and the title of the x-axis. Figure 3C, displays the format used in the study of Pieterse et al. (2010), which is not a standard format in the commercial software, but is the result of custom programming with Sawtooth Software (15). In addition to presenting attribute importance, three studies converted the part-worth utilities into overall preference scores for specific treatment options using Sawtooth Software Market Research Tools (SMRT) (18). Based on the assumption that patients prefer the option which provides them with the highest utility, these studies made full use of the capabilities of CA-exercises to predict shares of preference. The studies used different formats to present the outcomes. For instance, in the study of Hawley et al. (2015) patients only received a treatment recommendation in words (Figure 4A) (20). In the other two studies (Figure

4B 4C), a vertical bar was used to display the relative ranking of options against the worst and best (hypothetical) treatment options, only the lay-out and format of the vertical bar slightly differ.

In the studies of de Achaval et al. (2012) and Fraenkel et al. (2007) a research assistant explicitly explained the results to patients. Additionally, four studies explicitly mentioned that patients received a hand-out or a printed sheet with their results in order to discuss them with their health care provider (13, 14, 18, 20).

Figure 3. Results presentation to patients: attribute importance

3A

3B

3C

Outcome	minimal probability	maximal probability	Importance
Probability of Attribute A	4 out of 10	6 out of 10	20%
Probability of Attribute B	2 out of 10	8 out of 10	40%
Probability of Attribute C	1 out of 10	4 out of 10	30%
Probability of Attribute D	2 out of 10	9 out of 10	10%

3A Abraham et al. (2015), de Achaval et al. (2012), 3B Fraenkel et al. (2007), Fraenkel, (2010)-Rochon et al. (2012)
3C Pieterse et al. (2010)

* In the clinical trial protocol of Jayadevappa et al. (2015) it is described that "patients receive a graph and a list of the top five attributes most preferred by the participant".

* For the study of Hawley et al. (2015) it is unclear whether patients receive feedback about attribute importance.

Evaluation of Adaptive Conjoint Analysis

In most of the included studies, self-administered questionnaires or surveys were used to evaluate the ACA exercises except for the studies of Rochon et al. (2012) and Abraham et al. (2015), which used qualitative in-depth interviews and/or focus groups. Most patients across studies reported that the ACA exercise was easy to do, clear and interesting (15, 18). However, self-reported difficulty of the task was higher in patients who were less well educated (15) and older patients who were unfamiliar with computers (17). In one study, patients commented on the hypothetical nature of the task, which made them think they were being manipulated (17). In two studies, patients commented on the preselection of attributes, and that there was no possibility to add or exclude attributes or treatment options (17, 20).

Figure 4. Result presentation to patients: treatment recommendation

4A Hawley et al. (2015)
4B Fraenkel (2010)-Rochon et al. (2012)
4C Fraenkel et al. (2007)

With regard to reliability of preferences, Pieterse et al. (2010) asked patients to do a retest after 7-10 days and found that preferences were unstable in one-third of the sample. Hawley et al. (2015) found a concordance of 96% between predicted and revealed preferences (treatment actually received or planning to receive). Rochon et al. (2012) found that 66% of patients agreed with their prediction of preferred treatment and in the study of Fraenkel et al. (2007) 68% thought the feedback “very much” reflected their values.

Patients reported having greater self-confidence and self-efficacy, greater knowledge of risk and benefits and greater awareness of treatment preferences (18-20). Pieterse et al. (2010) reported specifically that patients were satisfied with the feedback received on attribute importance. However, de Achaval et al. (2012) showed that a video booklet decision aid provided the largest reduction in decision conflict compared to a group receiving an education booklet and a group receiving the videobooklet + ACA exercise.

In general, patients were positive with regards to the usefulness of results. Pieterse et al. (2010) reported that most patients would discuss results with their doctor and that 62% thought the ACA exercise would be helpful in deciding about treatment (15). Fraenkel et al. (2007) and Abraham et al. (2015) also reported increased patient activation and increased patient engagement with clinicians after the use of the CA exercise (18, 19). Abraham et al. (2015) studied the effect of the CA exercise on adherence and found that patients who used the ACA exercise and were prescribed a therapy that was concordant with their preferences had a 15% increase in adherence to their antithrombotic therapy (19). However, none of included studies went into depth about how patient engagement should take place and how the results of the exercise should be incorporated into the patient-physician discussion. Additionally, a barrier reported across articles for successful implementation is the current infrastructure and time and effort needed from the clinical personnel to discuss the result in a “shared decision making” consultation (15, 19, 20).

DISCUSSION

The aim of this study was to systematically review the use of CA to support preference elicitation at the individual patient level in a clinical setting. Eight studies met our inclusion criteria. The principal findings related to our aims are discussed below.

Selection of Attributes and Levels

All studies worked from a fixed set of attributes determined by focus groups or literature reviews or expert consultations. None of the included studies allowed individual patients to determine the relevant attributes prior to the start of the ACA exercise (either from a predefined set or come up with attributes themselves). However, identification of attributes should be supported by evidence on the potential range of preferences and values that the specific patient holds (2). Predetermination of attributes and levels might omit values which are important to specific patients. However, CA exercises cannot be easily adapted to include additional attributes for each patient at the time of decision making.

Choice Task Design

In this review, all included studies used ACA as their preference elicitation method. Other CA methods, such as discrete choice experiments, and full profile conjoint were not observed. This may be because not for all CA-methods commercially software is available to generate reliable preference data based on the responses of a single patient in real time (21). Another plausible explanation is the unstudied effect of using less efficient designs (not fully balanced and orthogonal) on the reliability of individual preference estimations. However, ACA designs (in hindsight) usually have good statistical efficiency, although they are often not strictly orthogonal (22). Nevertheless, it might be interesting for researchers to have greater flexibility in CA-methods to construct decision support tools in healthcare settings. Fraenkel (2013) already described some pilot-test work with the best-worst scaling method in patients with rheumatoid arthritis (21). However, an article describing this work in detail is not yet published and therefore could not be included in this review.

Instrument Design

Individualized treatment recommendation

Only three studies fully exploited the abilities of CA exercises by presenting patients with a treatment recommendation or a ranking of treatment options. Presenting patients with an algorithmic treatment recommendation raises a number of issues. First, patients may not want to have a computer tell them what their best option is. Second, physicians might feel hesitant in case the treatment recommended is not the best treatment option according to the physician's experience or the evidence available (21). Third, shared decision making has historically shied away

from making treatment recommendations (23, 24). Providing the patient with results on attribute importance leaves more room for interpretation and discussion during shared decision making; however, it is arguable that helping patients to better understand how each treatment option aligns or does not align with what matters to them is a critical—and too often overlooked—step in supporting evidence-informed, values-congruent decisions (9). However, the question remains whether this should be done by an algorithmic treatment recommendation or by the doctor.

Feasibility of implementation (evaluation of ACA)

One of the main barriers we often came across was the implementation of ACA-exercises within the clinical workflow. For these tools to succeed it is of paramount importance that thought is given as to when is the appropriate time to elicit preferences, what effort is asked from physicians, and how will the results be discussed in consultations. Web-based ACA exercises offer flexibility to patients who have internet access to use the ACA exercise at home, at their own pace, and thus reduce time demands on clinicians. However, as study results have been shown, people who are older or who have lower levels of education may need additional support.

Limitations of This Review

To the best of our knowledge, this is the first study to systematically review the literature regarding individual preference estimation based on CA methods. However, there are some limitations of this review which have influenced the interpretation of its outcomes. Some studies lacked complete information on instrument design and it was difficult to get an unambiguous picture. Authors were approached for clarification, but few responded and clarified our questions. Furthermore, there might have been selection bias in our review. The studies of Cheung et al. (2010) and Imaeda et al. (2010) addressed the theme of individual preference estimation but were excluded because the experimental design of this ranking conjoint and maximum difference scaling experiment were designed to elicit population preferences (25, 26). Moreover, one study used a discrete choice experiment to improve patient knowledge and adjust high patient expectations of treatment outcomes (27). The study was excluded, because the discrete choice experiment was not analyzed, nor did patients receive results.

Conclusion

Eight studies met our inclusion criteria and used ACA to estimate individual preferences in clinical decision making. ACA has the potential to support preference elicitation in clinical decision making: patients had a positive attitude about the need to actively think about the relevant trade-offs and they reported ACA exercises to be useful and informative. However, for these tools to truly succeed further research should first focus on a more flexible set of attributes and levels, the feedback patients want to and should receive, and the fit within clinical workflow. Furthermore, it might be interesting for researchers to have greater flexibility in the choice of CA-methods for decision support tools, because the optimal approach will vary depending on the needs of the clinical setting and the capabilities of patients.

REFERENCES

1. Cattin P, Wittink DR. Commercial Use of Conjoint Analysis: A Survey. *J Marketing*. 1982;46(3):44-53.
2. Bridges JF, Hauber AB, Marshall D, Lloyd A, Prosser LA, Regier DA, et al. Conjoint analysis applications in health—a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. *Value Health*. 2011;14(4):403-13.
3. Ryan M, Farrar S. Using conjoint analysis to elicit preferences for health care. *BMJ*. 2000;320(7248):1530-3.
4. Weernink MGM, Janus SLM, van Til JA, Raisch DW, van Manen JG, IJzerman MJ. A Systematic Review to Identify the Use of Preference Elicitation Methods in Healthcare Decision Making. *Pharm Med*. 2014;28(4):175-85.
5. Kaltoft MK, Neilsen JB, Salkeld G, Dowie J. Can a Discrete Choice Experiment contribute to person-centred healthcare? *Eur J Pers Cent Healthc*. 2015;3(4):431-7.
6. O'Connor AM, Rostom A, Fiset V, Tetroe J, Entwistle V, Llewellyn-Thomas H, et al. Decision aids for patients facing health treatment or screening decisions: Systematic review. *BMJ*. 1999;319(7212):731-4.
7. Fagerlin A, Pignone M, Abhyankar P, Col N, Feldman-Stewart D, Gavaruzzi T, et al. Clarifying values: an updated review. *BMC Med Inform Decis Mak*. 2013;13(2):1-7.
8. Lancaster KJ. A New Approach to Consumer Theory. *J Polit Econ*. 1966;74(2):132-57.
9. Witteman HO, Gavaruzzi T, Scherer LD, Pieterse AH, Fuhrel-Forbis A, Chipenda Dansokho S, et al. Effects of Design Features of Explicit Values Clarification Methods: A Systematic Review. *Med Decis Making*. 2016;36(6):760-76.
10. Witteman HO, Scherer LD, Gavaruzzi T, Pieterse AH, Fuhrel-Forbis A, Chipenda Dansokho S, et al. Design Features of Explicit Values Clarification Methods: A Systematic Review. *Med Decis Making*. 2016.
11. de Bekker-Grob E. Discrete choice experiments in health care: theory and applications 2009.
12. Ryan M, Skåtun D, Major K. Using Discrete Choice Experiments to Go Beyond Clinical Outcomes when Evaluating Clinical Practice. In: Ryan M, Gerard K, Amaya-Amaya M, editors. *Using Discrete Choice Experiments to Value Health and Health Care*. Dordrecht: Springer Netherlands; 2008. p. 101-16.
13. Jayadevappa R, Chhatre S, Gallo JJ, Wittink M, Morales KH, Bruce Malkowicz S, et al. Treatment preference and patient centered prostate cancer care: Design and rationale. *Contemp Clin Trials*. 2015;45(Pt B):296-301.
14. de Achaval S, Fraenkel L, Volk RJ, Cox V, Suarez-Almazor ME. Impact of educational and patient decision aids on decisional conflict associated with total knee arthroplasty. *Arthritis Care Res*. 2012;64(2):229-37.

15. Pieterse AH, Berkers F, Baas-Thijssen MC, Marijnen CA, Stiggelbout AM. Adaptive Conjoint Analysis as individual preference assessment tool: feasibility through the internet and reliability of preferences. *Patient Educ Couns*. 2010;78(2):224-33.
16. Fraenkel L. Feasibility of Using Modified Adaptive Conjoint Analysis Importance Questions. *Patient*. 2010;3(4):209-15.
17. Rochon D, Eberth JM, Fraenkel L, Volk RJ, Whitney SN. Elderly patients' experiences using adaptive conjoint analysis software as a decision aid for osteoarthritis of the knee. *Health Expect*. 2014;17(6):840-51.
18. Fraenkel L, Rabidou N, Wittink D, Fried T. Improving informed decision-making for patients with knee pain. *J Rheumatol*. 2007;34(9):1894-8.
19. Abraham NS, Naik AD, Street RL, Jr., Castillo DL, Deswal A, Richardson PA, et al. Complex antithrombotic therapy: determinants of patient preference and impact on medication adherence. *Patient Prefer Adherence*. 2015;9:1657-68.
20. Hawley S, Newman L, Griggs J, Kosir M, Katz S. Evaluating a Decision Aid for Improving Decision Making in Patients with Early-stage Breast Cancer. *Patient*. 2015:1-9.
21. Fraenkel L. Incorporating patients' preferences into medical decision making. *Med Care Res*. 2013;70(1 Suppl):80s-93s.
22. Sawtooth Software. ACA 6.0 Technical Paper. 2007.
23. Frongillo M, Feibelman S, Belkora J, Lee C, Sepucha K. Is there shared decision making when the provider makes a recommendation? *Patient Educ Couns*. 2013;90(1):69-73.
24. Fried TR. Shared Decision Making — Finding the Sweet Spot. *N Engl J Med*. 2016;374(2):104-6.
25. Cheung SW, Aranda D, Driscoll CL, Parsa AT. Mapping clinical outcomes expectations to treatment decisions: an application to vestibular schwannoma management. *Otol Neurotol*. 2010;31(2):284-93.
26. Imaeda A, Bender D, Fraenkel L. What is most important to patients when deciding about colorectal screening? *J Gen Intern Med*. 2010;25(7):688-93.
27. Dowsey MM, Scott A, Nelson EA, Li J, Sundararajan V, Nikpour M, et al. Using discrete choice experiments as a decision aid in total knee arthroplasty: study protocol for a randomised controlled trial. *Trials*. 2016;17(1):1-10.

APPENDIX A: DETAILED SEARCH STRATEGY

A systematic literature search was conducted in May 2016 to identify published articles which have used a method of conjoint analysis to elicit individual preference weights. Several conjoint methods were specified in the table below under #1. The number of hits in search #1 was restricted by keywords which indicate individual preference estimation (#3 #4 #5 #6 #7 #8 #9) and by excluding other stated-preference methods (#10) and English language (#11). The tables below show the number of hits per combination per database. The different databases led to a total of 1057 unique articles which were screened.

Database Scopus:

Search	Keywords	Number of hits
#1	"conjoint analysis" OR "conjoint measurement" OR "conjoint studies" OR "conjoint choice experiment" OR "discrete choice conjoint experiment" OR "discrete choice experiment" OR "pairwise choices" OR "Best-Worst Scaling" OR "Best Worst Scaling" OR "MaxDiff Scaling" OR "Maximum Difference Scaling" OR "ranking conjoint" OR "rating conjoint" OR "adaptive conjoint analysis" OR "adaptive choice based conjoint" OR "choice based analysis" OR "full profile conjoint" OR "choice based conjoint" OR "choice set" OR "relative preference weight" OR "hypothetical scenario" OR "stated preference"	10.119
#2	#1 AND "decision aid"	14
#3	#1 AND "individual* utilit*"	15
#4	#1 AND value* AND clarif*	21
#5	#1 AND "decision analysis"	79
#6	#1 AND individual* w/2 preference (proximity operator used)	252
#7	#1 AND "shared decision" or "medical decision" or "clinical decision" or "treatment decision"	372
#8	#1 AND treatment w/2 choice (proximity operator used)	118
#9	#1 AND individual w/2 analysis (proximity operator used)	46
#10	AND NOT "multi-criteria decision" OR "MCDA" OR "analytical hierarchy process" OR "ahp" OR "time trade-off" or "standard gamble" or TTO Exclusion of other stated preference methods	- 47
#11	English language restriction	- 17
Total number of unique articles included from Scopus		732

Database Psychinfo:

Search	Keywords	Number of hits
#1	"conjoint analysis" OR "conjoint measurement" OR "conjoint studies" OR "conjoint choice experiment" OR "discrete choice conjoint experiment" OR "discrete choice experiment" OR "pairwise choices" OR "Best-Worst Scaling" OR "Best Worst Scaling" OR "MaxDiff Scaling" OR "Maximum Difference Scaling" OR "ranking conjoint" OR "rating conjoint" OR "adaptive conjoint analysis" OR "adaptive choice based conjoint" OR "choice based analysis" OR "full profile conjoint" OR "choice based conjoint" OR "choice set" OR "relative preference weight" OR "hypothetical scenario" OR "stated preference"	1878
#2	#1 AND "decision aid"	5
#3	#1 AND "individual* utilit*"	7
#4	#1 AND value* AND clarif*	2
#5	#1 AND "decision analysis"	9
#6	#1 AND individual* n2 preference (proximity operator used)	44
#7	#1 AND "shared decision" OR "medical decision" OR "clinical decision" OR "treatment decision"	61
#8	#1 AND treatment n2 choice (proximity operator used)	27
#9	#1 AND individual n2 analysis (proximity operator used)	9
#10	AND NOT "multi-criteria decision" OR "MCDA" OR "analytical hierarchy process" OR "ahp" OR "time trade-off" OR "standard gamble" OR TTO Exclusion of other stated preference methods	- 10
#11	English language restriction	- 1
Total number of unique articles included from Psychinfo		136

Database Pubmed*

Search	Keywords	Number of hits
#1	"conjoint analysis" OR "conjoint measurement" OR "conjoint studies" OR "conjoint choice experiment" OR "discrete choice conjoint experiment" OR "discrete choice experiment" OR "pairwise choices" OR "Best-Worst Scaling" OR "Best Worst Scaling" OR "MaxDiff Scaling" OR "Maximum Difference Scaling" OR "ranking conjoint" OR "rating conjoint" OR "adaptive conjoint analysis" OR "adaptive choice based conjoint" OR "choice based analysis" OR "full profile conjoint" OR "choice based conjoint" OR "choice set" OR "relative preference weight" OR "hypothetical scenario" OR "stated preference"	1876
#2	#1 AND "decision aid"	6
#3	#1 AND "individual* utilit*"	2
#4	#1 AND value* AND clarif*	10
#5	#1 AND "decision analysis"	25
#6	#1 AND "shared decision" or "medical decision" or "clinical decision" or "treatment decision"	92
#7	#1 AND (treatment and choice) OR (individual and analysis) OR (individual and preference) (in abstract-title)	449
#8	AND NOT "multi-criteria decision" OR "MCDA" OR "analytical hierarchy process" OR "ahp" OR "time trade-off" or "standard gamble" or TTO Exclusion of other stated preference methods	- 28
#9	English language restriction	- 1
Total number of unique articles included from Pubmed		485

Pubmed does not have proximity operators. Therefore, we searched for keywords in search term #7 in the Abstract and Title of articles.

Database Web of Science

Search	Keywords	Number of hits
#1	"conjoint analysis" OR "conjoint measurement" OR "conjoint studies" OR "conjoint choice experiment" OR "discrete choice conjoint experiment" OR "discrete choice experiment" OR "pairwise choices" OR "Best-Worst Scaling" OR "Best Worst Scaling" OR "MaxDiff Scaling" OR "Maximum Difference Scaling" OR "ranking conjoint" OR "rating conjoint" OR "adaptive conjoint analysis" OR "adaptive choice based conjoint" OR "choice based analysis" OR "full profile conjoint" OR "choice based conjoint" OR "choice set" OR "relative preference weight" OR "hypothetical scenario" OR "stated preference"	9567
#2	#1 AND "decision aid"	22
#3	#1 AND "individual* utilit*"	14
#4	#1 AND value* AND clarif*	31
#5	#1 AND "decision analysis"	88
#6	#1 AND individual* near/2 preference (proximity operator used)	275
#7	#1 AND "shared decision" or "medical decision" or "clinical decision" or "treatment decision"	198
#8	#1 AND treatment near/2 choice (proximity operator used)	165
#9	#1 AND individual near/2 analysis (proximity operator used)	34
#10	AND NOT "multi-criteria decision" OR "MCDA" OR "analytical hierarchy process" OR "ahp" OR "time trade-off" or "standard gamble" or TTO Exclusion of other stated preference methods	- 32
#11	English language restriction	- 15
Total number of unique articles included from Web of Science		136

APPENDIX B: EVIDENCE MATRIX

1. General Study Information

Study number	Author, year of publication	Study aim	Disease context	study protocol
1	Fraenkel L, Rabidou N, Wittink D, et al. Improving informed decision-making for patients with knee pain. The Journal of rheumatology. 2007; 34: 1894-8.	Develop an intervention to improve informed decision-making in the clinical practice setting for patients with knee pain, based on ACA.	Rheumatology osteoarthritis	Pilot study (planning study for RCT). Information pamphlet vs ACA
2*	Fraenkel L. Feasibility of Using Modified Adaptive Conjoint Analysis Importance Questions. Patient. 2010; 3: 209-15.*	To develop a modified version of ACA importance question and test their performed in a pilot study of patients with knee pain.	Rheumatology knee pain	pilot-study
2*	Rochon D, Eberth JM, Fraenkel L, et al. Elderly patients' experiences using adaptive conjoint analysis software as a decision aid for osteoarthritis of the knee. Health Expect. 2014; 17: 840-51*	to obtain insight into the experiences of elderly patients who use aca to explore treatment option for their osteoarthritis.	Rheumatology osteoarthritis	implementation Fraenkel et al 2010
3	Pieterse AH, Berkens F, Baas-Thijssen MC, et al. Adaptive Conjoint Analysis as individual preference assessment tool: feasibility through the internet and reliability of preferences. Patient Educ Couns. 2010; 78: 224-33.	We wished to assess the usefulness of ACA at the individual level, to support treatment-related decision making.	Rectal cancer	exploration compare internet vs CAPI ACA
4	de Achaval S, Fraenkel L, Volk RJ, et al. Impact of educational and patient decision aids on decisional conflict associated with total knee arthroplasty. Arthritis care & research. 2012; 64: 229-37.	To examine the impact of a video booklet patient DA supplemented by an interactive VCE on decisional conflict in patients with knee osteoarthritis considering total knee arthroplasty.	Rheumatology knee Osteoarthritis	exploration of ACA RCT (aca vs video vs education forms)

	Author, year of publication	Study aim	Disease context	study protocol
5	Hawley S, Newman L, Griggs J, et al. Evaluating a Decision Aid for Improving Decision Making in Patients with Early-stage Breast Cancer. The Patient - Patient-Centered Outcomes Research. 2015: 1-9.	A pilot-study to develop a web-based decision aid focused on loco regional breast cancer treatment.	Breast cancer	pilot-study
6	Abraham NS, Naik AD, Street RL, Jr., et al. Complex antithrombotic therapy: determinants of patient preference and impact on medication adherence. Patient Prefer Adherence. 2015; 9: 1657-68.	Examine the clinical application of aca to obtain patient preferences and the impact of ACA on medication adherence in complex antithrombotic therapy	Cardiovascular conditions	exploration of ACA
7	Jayadevappa R, Chhatre S, Gallo JJ, et al. Treatment preference and patient centered prostate cancer care: Design and rationale. Contemporary clinical trials. 2015; 45: 296-301.	We analyzed the effect of conjoint task intervention on the association between preferences, treatment and objective and subjective outcomes.	Prostate cancer	clinical trial protocol (no results)

* The study of Fraenkel et al (2010) and Rochon et al. (2012) relate to the same value clarification exercise

2. Selection of Attributes and Levels

S t u d y	method for attribute and level selection	No. of attributes	range of levels	possibility to add extra attributes	type of attributes	graphs or visual pictographs
1	randomized controls trials	5	not completely described (4 ¹ and ?)	not described	process (1) benefits (2) harms (2)	risk information: natural frequencies and pictographs
2	not described	6	3 ⁶	not described	process (2) benefits (2) harms (2)	no
3	literature	4	3 ³ 2 ¹	not described	benefits (2) harms (2)	yes, risk dots graph
4	two clinicians	8	4 ⁸	not described	process (2) benefits (3) harms (3)	yes, risk dots graph (some with smileys)
5	literature, interviews with surgeons and oncologist, and 10 survivor patients, ranking rating exercise	4	2 ⁴	not described	process (2) benefits (1) harms (1)	no
6	systematic literature review + patient focus group	7	3 ⁵ 2 ²	not described	process (1) benefits (2) harms (4)	yes, risk dots graph with smileys
7	one-on-one in-depth interview with 49 patients and three provider focus groups literature review	not described	not described	not described	not described	not described

3. Choice Task Design

Study	CA method	Type of questions / experimental design	No. of choice tasks
1	ACA	no, importance questions followed by paired comparisons	rank exercise: levels of attributes, rate exercise for each attributes (5) paired-comparisons: exact number unknown
2	modified ACA	no, importance questions (adapted) followed by 18 paired comparisons	18
3	ACA	no, importance questions followed by 12 paired comparisons	12, upwards partial (2 attributes for q 1-5, 3 attributes for q 6-10, 4 attributes for q 11-12. Number is selected based on a formula $3 * (N - n - 1) - N$, where N is the number of levels across all attributes and n is the number of attributes
4	Modified ACA	no, importance questions (adapted) followed by paired comparisons	not described
5	ACA	"for the final exercise the patients were presented with 16 scenarios based on combinations of these attributes and levels"	16
6	modified ACA	no, importance questions (adapted) followed by paired comparisons	not described
7	ACA	not described	not described

4. Values Clarification Instrument Design (1 out of 2)

S t u d y	medium used	offline /online	software program used	help, explanation during/befo re task	analysis performed on data
1	computer	offline	sawtooth software	afterwards explanation by research assistant	ACA predicts preference based on the utilities derived from the conjoint questionnaire using least- squares regression analysis
2	computer	offline	sawtooth software	not described	Regression models. Utilities are calculated using a least squares updating algorithm. The final estimate reflect true least squares. Market simulators are used to convert the raw utilites into preferences for specific options.
3	computer	both	sawtooth sotware a feedback module specifically developed	solo activity, no help is given	Ordinary Least Squares
4	computer	offline	sawtooth software	on their own, but there was someone present for questions	program's calculation
5	computer	online (website)	not described	not described	not described
6	computer	offline	sawtooth software	research coordinator provided oral instructions	regression analysis
7	computer	both	sawtooth software	not described	not described

4. Values Clarification Instrument Design (2 out of 2)

Study	Patients received feedback on attribute importance		Patients received a treatment recommendation	
		How?		How?
1	yes	bar graphs indicating the importance of each attribute (x as percentages)	yes	relative treatment preferences: scale showing the relative ranking of the available options. Ranging from 0 (worst choice) to 100 (best choice)
2	yes	bar graphs indicating the importance of each attribute (without exact percentages)	yes	a ruler (0-100) with the postulated best treatment choice for the individual. Ends anchored at best and worst choice and in steps of 10 %.
3	yes	bar graphs indicating the importance of each attribute (with exact percentages) and min/max probability	no	no
4	yes	bar graphs indicating the importance of each attribute (without exact percentages)	no	no
5	not described	not described	yes	text: based on the responses you just gave, the treatment that seems to be the best fit for you is: X
6	yes	bar graphs indicating the importance of each attribute (without exact percentages)	no	no
7	yes	summary of features of prostate cancer treatment that are most important and a graph and list of the top 5 attributes most preferred is generated	not described	not described

5. Evaluation of CA-exercises (1 out of 2)

Study	Evaluation	Outcomes measures	Reported positive outcomes
	1 self-administered paper-pencil questionnaires	Decision Self-Efficacy Scale Preparation for Decision-Making Tool Arthritis Self-Efficacy scale Acceptability of the ACA task (difficulty of the task, recommend to others, bar graphs reflected their values)	greater self-confidence in their abilities to participate in shared decision-making, felt more prepared to participate in decision-making, and had greater arthritis self-efficacy 68% bar graph reflected their values, 27% somewhat, 5% little or not at all. 86% would recommend it to other patients
	2 5 focus groups: content analyses	predicted preferred treatment the software the choice of treatments the treatment attributes cumulative overload	66% agreed with their prediction of their preferred treatment
	3 paper-based closed ended	survey question regarding task difficulty and usefulness of results	3.3/4 scored for clarity 87% not difficult to comprehend - 93% comprehended the relative importance advice 78% would discuss results with clinicians 63% gained insight 62% thought it would be helpful in deciding for treatment
	4 post intervention questionnaire	decision conflict scale (DCS)	Compared to the booklet DCS was statistically significantly reduced with ACA
	5 survey	decision satisfaction, perceived value concordance	The statement: 'treatment decision matched my values' was statistically significant for the tool compared to no decision aid 96% less often unsure or felt the decision was hard to make
	6 qualitative interviews	adherence scores	15% increased adherence score patient activation awareness of treatment preferences increased engagement of clinicians
7	No evaluation	not described	not described

5. Evaluation of CA-exercises (2 out of 2)

Study	Reported negative outcomes	Task acceptability / difficulty
1	not described	difficulty 74% very easy 24% easy 2% very hard
2	non-familiarity with computers Lower education increases difficulty	many problems related to non-familiarity with computers fewer paired choices would be easier
3	15% found the questionnaire annoying, tiring, threatening education level was determinant for comprehension	not described
4	Compared to the video booklet DCS was not statistically significantly reduced with ACA	not described
5	The CA-based VCE was limited by the attributes selected, the use of an adaptive vs. choice-based approach, as well as by the subjective linking of the results of trade-off preferences to actual treatment.	majority viewed website at home
6	logistically prohibitive, infrastructure time and requirements, description of attributes influenced answers (framing issue)	not described
7	not described	not described

7

Best-Worst Scaling for Value Clarification and Decision Support: Structured Approach

Marieke G.M Weernink
Maarten J. IJzerman
Marije van der Straten
Paul J.M. Kil
Janine A. van Til

Submitted

ABSTRACT

Introduction: Best-Worst Scaling (BWS) is a stated preference technique which can be used to clarify treatment values to individual patients by asking them to evaluate the desirability of treatment characteristics. However, its potential in clinical decision making (assisting individual patients in treatment decisions) has not yet been studied.

Objective: The objective of this study is to determine (from a methodological perspective) whether BWS can be seen as a promising in clinical decision support. The article provides information on the procedural steps taken to use BWS in a value clarification exercise (VCE). This will be illustrated by an example of a treatment decision in Parkinson's Disease.

Procedural steps: Three distinct procedural steps were distinguished: (1) decision problem structuring (2) choice task design, and (3) instrument design. First, the clinical problem needs to be clear and the relevant attributes on which the treatments differ have to be selected. Next, decisions have to be made upon the use of full or partial choice tasks, the number of choice tasks (based on experimental design software), and the corresponding best-worst question. Finally, the choice tasks need to be embedded in a larger instrument. Decisions have to be made on the mode of administration, the type of statistical analysis, and the feedback patients receive.

Discussion: Several methodological challenges and uncertainties were encountered in describing the procedural steps. Further research must first determine to which extent the use of D-efficient designs and different statistical analysis techniques affect the reliability of individual preference estimations. From a methodological perspective it is too early to promote widespread use of BWS in VCEs.

INTRODUCTION

In shared decision making, patient decision aids are often used to support patients' understanding of the process of care and the subsequent evidence-based outcomes (1). Values clarification methods are used in decision aids to help patients evaluate the desirability of attributes or options, with the aim that choice of treatment reflects personal preferences and values (2). In the development of value clarification methods, it is difficult to be systematic and internally coherent when not referring to a theory (3, 4). Fagerlin et al. (2) therefore suggests more research is needed regarding value clarification methods based on specific theories.

A value clarification method which has a strong theoretical foundation in random utility theory and Lancaster's theory of value is best-worst scaling (BWS). BWS is a relatively new technique for measuring the importance or preference of multiple items. It can be used to clarify treatment values to individual patients by asking them to evaluate the desirability of treatment characteristics (5). BWS can be grouped under conjoint analysis which is based upon the assumption that "people make decisions about the overall value of a good or treatment based on the value of its multiple, combined (conjoint) features" (6). Three cases of BWS were developed in which the nature and complexity of the items being chosen differs. The 'object case' (BWS case 1) asks patients to select the best and the worst treatment attribute from a selection of treatment attributes. BWS case 1 does not distinguish levels, but only asks individuals for the most and least preferred attribute (7). Alternatively, BWS case 2 (the profile case) has an attribute-level structure and let respondents consider one scenario at a time, and ask individuals to select the best and worst attribute-level within one scenario. BWS case 3 'the multi-profile case' also has an attribute-level structure and mostly resembles a traditional discrete choice experiment: patients are asked to identify the most and least preferred treatment option from sets of three (or more) treatment options (7).

BWS lets patients explicitly and actively think about the relevant trade-offs between treatment characteristics and has the ability to elicit the reasons behind patient choices. Based on its theoretical axioms (8), BWS can present patients with estimates of the relative impact of each attribute on their decision as well as provide an estimate of their preferences for available treatment options (9). Moreover, literature suggests that BWS has a lower cognitive burden compared to other methods with a strong theoretical foundation: discrete choice experiments, time trade-off, and standard gamble (5, 10, 11). Hence, BWS has potential as a decision support tool, but until now it has only been used to support organizational and regulatory decision making (by estimating the preferences of a group of patients).

The aim of this study is to determine (from a methodological perspective) whether BWS case 2 can be seen as a promising in clinical decision support. The article provides information on the procedural steps taken to use BWS in a value clarification exercise (VCE). This will be illustrated by an example of a treatment decision in Parkinson's Disease. The actual implementation of such a tool, together with studying its feasibility from patient's perspective, is beyond the scope of this paper and will be discussed in Chapter 8 of this dissertation. Furthermore, BWS case 2 was specifically selected, because it is more difficult to compare multiple treatment scenarios in a BWS case 3 task compared to evaluating one treatment scenario in a BWS case 2 task (5).

PROCEDURAL STEPS

The majority of conjoint analysis studies is undertaken to elicit preferences in groups of patients and for these situations a general guideline is available (12). For the aim of this study, this guideline was used as a basis, but the steps were adapted to accommodate the estimation of individual preferences. Our structured approach consist of three distinct steps: (1) decision problem structuring, (2) choice task design, and (3) instrument design. Below, the points which are considerably different from the guideline are described in detail. All steps will be illustrated by an example of a treatment decision in Parkinson's Disease.

Step 1: Decision Problem Structuring

First, the clinical decision problem and the range of treatment options need to be clear. Next, the relevant attributes on which the treatments differ have to be selected. Depending on the decision context, attributes can consist of health outcomes (e.g. efficacy, survival), process of care (e.g. waiting time, location) or non-health outcomes (e.g. information provision) (13). The selection of attributes is followed by the definition of levels for each attribute that reflect a realistic range of outcomes associated with the treatment options available. In contrast to other conjoint analysis designs, it is important to select attribute-levels that directly link to treatment options as other levels are valueless in informing the specific treatment decision.

The selection of attributes and levels need to be supported by evidence on the potential range of values and preferences that patients may hold (12). A first step to find this evidence is by conducting a literature review. However, an initial list of attributes can be long and exhaustive, and the human mind is limited in its ability to include many attributes and levels in decision making. Several studies/reports advise the maximum number of attributes to be 5 to 7 (14, 15). Conducting

qualitative research such as interviews and focus groups with physicians and patients can be used to narrow the list of attributes. The selection of attributes and levels is a crucial part of the VCE and cannot be solely based on literature review. The results have to be verified using patient focus-groups and expert opinions.

Illustration in Parkinson's Disease:

In later stages of Parkinson's Disease, oral pharmacological treatment alone is not effective in managing symptoms. Two advanced therapeutic options for these patients are: Levodopa/Carbidopa Intraduodenal Gel Infusion via an external pump (Duodopa pump) or deep brain stimulation (DBS). However, there is little data available comparing the effectiveness of these two interventions (16). Consequently, clinicians are often in a difficult position when attempting to determine a therapeutic option for an individual patient, because the treatments do differ with regard to their risk, complications and process of care (17). A decision aid and value clarification exercise could help patients and clinicians to select the most suitable treatment option based on the patient's preferences.

The aim of this article is to demonstrate the use of BWS in a clinical decision in a simple and straightforward way. For reasons of convenience, it was chosen to select five important attributes on which the two treatment options differ (Table 1). In real practice, as suggested above, thorough literature reviews, patient focus-groups and expert opinions are needed to justify this selection of attributes and levels.

Table 1. Selected attributes and levels

	Treatment option 1: Duodopa pump	Treatment option 2: Deep brain stimulation
A: Type of surgery	Undergoing a minor surgical procedure; a gastrostomy tube is inserted into the duodenum	Undergoing a major surgical procedure; awake brain surgery
B: Device	Wearing an external pump (1.5 pounds) at the height of your hip	Having an internal device (a neurostimulator) under the skin of your collarbone
C: Daily maintenance	Independently change the drug cassette and flush the intestinal tube (each 24hrs)	Daily check of the batteries of the neurostimulator
D: Complications / Risks	Run the risk of problems with mechanical equipment (dislocation, obstruction, breakage)	Run the risk of problems with cognition, behavior, psyche and speech
E: Oral medication	Only taking oral medication at night (instead of throughout the day)	Reduction of the number of pills throughout the day

Step 2: Choice Task Design

When attributes and levels are selected, the next step is to create the hypothetical treatment scenarios. In a full factorial design all possible combinations of attributes and levels are included. For our selection of 5 attributes with two levels the full factorial design results in 32 (2^5) treatment scenarios (Figure 1A). However this number increases rapidly when more attributes and levels are selected ($2^6 = 64$ scenarios, $3^4 = 81$ scenarios etc.). It is not feasible to ask patients to answer best-worst questions for each possible treatment scenario. Moreover, the general guideline for conducting conjoint analysis states that it is good practice to include 8 to 16 choice tasks per respondent (12, 19).

Evidence suggest that it yields reliable interpretable data if one constructs choice tasks using a suitable experimental design (19, 20). As a full factorial design often is not feasible, a subset of choice tasks (a fractional factorial experimental design) is judiciously selected based upon the principles of level balance and orthogonality (21). Level balance is satisfied when each attribute-level appears with equal frequency and orthogonality is satisfied when each pair of attribute-levels appears equally often in the generated design. Moreover, for accurate individual estimates each attribute-level should be shown at least three times to a patient (22). It is time-consuming to construct experimental designs manually, an extensive overview of various software programs can be found in Johnson et al. (2013) (19).

In order to estimate reliable preferences balanced and orthogonal (efficient) experimental designs are preferred (Figure 1B). These, however, are scarce because they are only available for a limited number of attributes with specified levels. If such a design is available, it is likely to be a large set of choice questions, which might result in patients using cognitive heuristics or shortcuts (18). A solution can be the use of D-efficient designs (which approximate full balance and orthogonality) (Figure 1C). However, the effect of using these less efficient designs on the reliability of individual preference estimation is unknown and unexplored.

A second choice, which also influences experimental design decisions, is whether full or partial profile choice tasks are used. In full profile tasks, all selected attributes are included in each choice task (Figure 1A 1B 1C). Full profiles tend to be a more realistic description of the clinical situation, because all relevant factors are included in the choice task. However, they can be (cognitively) more burdensome to patients due to possible information overload. In partial-profile designs, each choice task includes a subset of the selected attributes. However, as Figure 1D shows, this means

Figure 1. Construction of the choice task: Perfect, D-efficient, full profile and partial profile choice task designs.

1a: Full Factorial Design 2⁵

- Attribute balance (16x)
- Orthogonality (8x)
- 32 full profile choice tasks

Attribute ^{levels} → Choice tasks ↓	A ²	B ²	C ²	D ²	E ²
1	A1	B2	C1	D1	E2
2	A2	B2	C1	D1	E2
3	A1	B2	C2	D2	E2
4	A2	B1	C2	D1	E1
5	A2	B2	C1	D2	E1
6	A1	B2	C1	D1	E1
7	A1	B1	C1	D2	E1
8	A1	B1	C1	D2	E2
9	A1	B2	C1	D2	E2
10	A1	B2	C2	D1	E2
11	A2	B1	C2	D2	E1
12	A2	B2	C1	D1	E1
13	A2	B1	C1	D1	E2
14	A2	B2	C1	D2	E2
15	A2	B2	C2	D1	E2
16	A1	B1	C2	D2	E2
17	A1	B1	C1	D1	E2
18	A2	B2	C2	D1	E1
19	A1	B2	C2	D1	E1
20	A2	B1	C2	D1	E2
21	A1	B2	C2	D2	E1
22	A2	B1	C1	D2	E1
23	A1	B1	C2	D1	E1
24	A1	B1	C2	D2	E1
25	A2	B2	C2	D2	E2
26	A1	B1	C1	D1	E1
27	A1	B1	C2	D1	E2
28	A1	B2	C1	D2	E1
29	A2	B2	C2	D2	E1
30	A2	B1	C1	D1	E1
31	A2	B1	C2	D2	E2
32	A2	B1	C1	D2	E2

1c: Fractional Factorial D-efficient Design 2⁵

- Attribute balance (4x)
- No orthogonality (the combination of A1-B2 appears twice, while the combination C1-D1 appears once)
- 8 full profile choice tasks

Attributes ^{levels} → Choice tasks ↓	A ²	B ²	C ²	D ²	E ²
1	A1	B2	C2	D2	E1
2	A1	B2	C2	D1	E2
3	A2	B2	C1	D2	E2
4	A2	B1	C2	D1	E2
5	A1	B1	C1	D2	E2
6	A1	B1	C2	D1	E1
7	A2	B2	C1	D1	E1
8	A2	B1	C1	D2	E1

1b: Fractional Factorial Perfect Design 2⁵

- Attribute balance (8x)
- Orthogonality (4x)
- 16 full profile choice tasks

Attribute ^{levels} → Choice tasks ↓	A ²	B ²	C ²	D ²	E ²
1	A2	B2	C2	D1	E2
2	A2	B1	C2	D2	E1
3	A2	B1	C2	D1	E2
4	A1	B1	C2	D2	E2
5	A1	B1	C1	D1	E2
6	A2	B2	C1	D1	E1
7	A2	B2	C1	D2	E2
8	A2	B1	C1	D2	E2
9	A2	B2	C2	D2	E1
10	A2	B1	C1	D1	E1
11	A1	B1	C1	D2	E1
12	A1	B2	C2	D1	E1
13	A1	B2	C1	D2	E1
14	A1	B2	C1	D1	E2
15	A1	B2	C2	D2	E2
16	A1	B1	C2	D1	E1

**1d: Fractional Factorial D-efficient Design 2⁵
(partial profile)**

- Attribute balance (4x)
- No orthogonality (the combination of A1-B1 appears twice, while the combination A1-B2 appears once)
- 10 partial profile choice tasks

Attributes ^{levels} → Choice tasks ↓	A ²	B ²	C ²	D ²	E ²
1	A1	B2	C2		E2
2	A2		C2	D1	E2
3		B2	C2	D2	E1
4	A1		C1	D2	E2
5	A2	B1	C2		E1
6	A1	B1	C1	D1	
7	A1	B1		D2	E1
8	A2	B1		D1	E2
9		B2	C1	D1	E1
10	A2	B2	C1	D2	

that patients have to answer more choice tasks in total. There are no strict guidelines on when to use full or partial profiles, but Green and Srinivasan (1990) recommend six attributes as the maximum number of attributes for full profiles (23). However, studies have found that patients may evaluate more complex profiles when attributes are concisely written or graphics are used (24, 25).

A third relevant choice is which corresponding best-worst question patients should consider for each choice task. The question must include an underlying subjective dimension, i.e. which aspect of treatment is considered “best” (most important to gain) and “worst” (most important to avoid) (26). However, in clinical situations it is possible that there is nothing to gain, only things that are less important to avoid. The subjective dimension can be adapted to each context (desirable, bothersome etc.) and which best-worst question is best must be agreed upon with clinicians and patients. Lastly, the lay-out of the choice task should be made attractive and comprehensible by visuals or color cues.

Illustration in Parkinson’s Disease: Choice task design

As discussed above, with a 2^5 design it is possible to use a perfect (efficient) experimental design of 16 choice tasks (Figure 1B). This is on the verge of the 8 to 16 choice tasks, which the ISPOR guidelines describe as good practice. Full profiles were used, because the use of five attributes will not lead to information overload. ‘Bothersome’ was chosen as the subjective dimension in the best-worst questions, because it reflects the negativity surrounding the attributes. Figure 2 shows an illustration of one of the choice tasks.

Figure 2. Example of one choice task.

What do you consider the most and least bothersome characteristic of treatment?		
Least bothersome characteristic		Most bothersome characteristic
☐	Minor surgical procedure; a gastrostomy tube is inserted into the duodenum	☐
☐	An internal device (a neurostimulator) under the skin of your collarbone	☒
☐	Independently change the drug cassette and flush the intestinal tube (each 24hrs)	☐
☒	A reduction in the number of pills per day	☐
☐	Run risk at problems with cognition, behavior, psyche and speech	☐

2.3. Instrument Design

The next step is to embed the choice tasks in a value clarification tool. Decisions have to be made on the mode of administration, the type of statistical analysis, and the feedback patients will receive.

VCEs are administered in various ways (paper-questionnaires, computers, interviewer). The perceived advantage of using paper questionnaires might be that older patients are more familiar with them. However, computer-based VCEs have the advantage that they allow for direct presentation of the results to patients. The choice for a specific instrument (computer program or software package) depends, among others, on the chosen methods for statistical analysis. The BWS based VCE can be analysed using counts analysis or Hierarchical Bayes (HB) analysis. Counts analysis examines choice frequencies by subtracting the number of times an attribute-level was chosen as worst from the number of times it was chosen as best (best-minus-worst score). Counts analysis is a simple and rapid analysis, but if the design is not balanced and orthogonal it will produce biased estimates because not each attribute-level had the same opportunity to be chosen as best or worst. HB analysis models the preference data and derives individual part-worth utilities for each treatment characteristic. However, the utilities values are measured on a latent scale, with no minimum or maximum values. The latent scale can be viewed as a representation of the underlying subjective dimension (e.g. degree of importance), but might be difficult to interpret. HB requires targeted software, but is a more robust model for estimating preferences and can present several goodness-of-fit statistics (27). A con is that it needs preference information from other patients to stabilize part-worth utilities for each individual patient and the analysis can be time-consuming. Many packages are able to conduct a simple count-analysis, but HB analysis requires certain computing capabilities and a user license.

Next, using the best-minus-worst scores and/or the part-worth utilities the relative importance of each attribute can be estimated by calculating the difference between the minimum and maximum value within an attribute divided by the total sum of ranges across all attributes. The attribute importance depicts the importance of the difference in the levels of one attribute. In addition, when assuming appropriate task design, the value of treatment options can be determined by adding up the value (part-worth utility or best-minus-worst scores) of its characteristics. Based on the assumption that patients prefer the option which provides them with the highest utility, the treatment option with the highest value can be recommended to the patient. An important element of the VCE is how these results are then presented to patients. The use of graphics or images is encouraged, but the options depend on the software used (28).

Preferably, patients are provided with a printed handout illustrating their preferences, which they subsequently can discuss with their clinician.

Instrument design also means writing a clear introduction to make the patient familiar with the specified attributes and levels (preferably in combination with a validated decision aid). Subsequently, the choice task needs to be clearly explained in order for the patient to understand what exercise to perform and what result to expect. Often a decision aid and value clarification exercise are brought together, in which a patient receives and reads the information on attributes and levels in the decision aid and actively thinks about the relevant trade-offs in the VCE.

Illustration in Parkinson's Disease

Each patient needs to complete 16 choice tasks and Table 2 shows the results derived using best-minus-worst counts analysis and HB analysis. The attribute importance depicts the importance of the difference in the levels of one attribute. For example, the attribute-levels of 'type of surgery' were both negatively valued, but the difference in scores between the two attribute-levels was smaller compared to the difference in attribute-levels of 'daily maintenance' and 'complications and/or risks' and therefore the latter two attributes have a higher attribute importance and have more influence on the treatment decision. Figure 3 shows an example of how feedback on the results can be given to patients. Figure 3A shows the results on relative importance in a horizontal bar graph and Figure 3B uses a vertical bar to display the relative ranking of the two treatment options against the best and worst hypothetical treatment options possible.

DISCUSSION

This article provided information on the procedural steps of how BWS case 2 can be used as a value clarification method to help patients clarify their values in treatment decisions. However, we encountered several methodological challenges in these procedural steps and these will be discussed below.

Step 1: Decision Problem Structuring

One of the main issues is that attributes are selected based upon the results of a literature study and/or the generalization of the results of patient focus groups. However, selected attributes should be supported by evidence on the potential range of preferences and values that the

Table 2. Example of the statistical analysis of the 16 choice tasks for one patient using two methods: counts analysis and Hierarchical Bayes analysis.

		Counts analysis			Hierarchical Bayes analysis		
Attribute	Attribute-level	BW score *	Difference within attribute**	Attribute importance ***	Part-worth utility	Difference within attribute**	Attribute importance ***
A: Type of surgery	Undergoing a minor surgical procedure; a gastrostomy tube is inserted into the duodenum	-2	3	12.5%	-2.65	8.04	13.4%
	Undergoing a major surgical procedure; awake brain surgery	-5			-10.69		
B: Device	Wearing an external pump (1.5 pounds) at the height of your hip	+1	0	0%	9.44	1.35	2.3%
	Having an internal device (a neurostimulator) under the skin of your collarbone	+1			8.09		
C: Daily maintenance	Independently change the drug cassette and flush the intestinal tube (each 24hrs)	-4	8	33.3%	-8.98	19.51	32.5%
	Daily check of the batteries of the neurostimulator	+4			10.53		
D: Complications or risks	Run the risk of problems with mechanical equipment (dislocation, obstruction, breakage)	+5	10	41.7%	12.69	23.38	39.0%
	Run the risk of problems with cognition, behavior, psyche and speech	-5			-10.69		
E: Oral medication	Only taking oral medication at night (instead of throughout the day)	+4	3	12.5%	7.71	7.71	12.9%
	Reduction of the number of pills throughout the day	+1			0.00		
Treatment score¥: Duodopa pump		-2 +1 -4 +5 +4 = +4			-2.65+9.44-8.98+12.69+7.71 = 18.21		
Treatment score¥ Deep Brain Stimulation		5 +1 +4 -5 +1 = -4			-10.69+8.09+10.53-10.69+ 0.00 = -2.76		

***BW score (best minus worst score):** number of times it was chosen as best minus number of times it was chosen as worst.

** **Difference within attribute:** The difference between the minimum and maximum value within an attribute

*****Attribute importance:** The difference between the minimum and maximum value within an attribute divided by the total sum of ranges across all attributes.

¥**Treatment scores:** adding up the 'best minus worst scores' of the levels corresponding to the specific treatment options. The treatment with the highest positive score is the treatment option recommended based on the patient's preference.

Figure 3. *Feedback to patients*

individual patient may hold (12). It may occur that patients attach value to attributes not included in the VCE. For example research of Feldman et al. (2004) showed that 33 different attributes were identified as affecting at least one patient’s decision for prostate cancer treatment (29). In some cases attributes or treatments might not even be an option for patients due to contraindication. Further studies should focus on whether it is possible to make the VCE more adaptive and flexible by for example letting each individual patient first pick their own most important attributes from an extensive list. Subsequently, in theory, linear programming and a suitable solver can be used to develop quickly patient-specific experimental designs. However, this would require assistance from specialized agencies to develop and program software.

Step 2: Choice Task Design

As discussed above, balanced and orthogonal experimental designs are in general preferred, because then each attribute has equivalent statistical power and the attributes are uncorrelated (30). BWS lets patients explicitly and actively think about the relevant trade-offs between attributes and attribute-levels. However, if patients use fast and frugal heuristics of simplifying decision strategies due to the number of choice tasks, then BWS fails to achieve its objectives. In practice, statistical or design efficiency (how well the attributes in the design are balanced and orthogonal) will be balanced against response efficiency (what is cognitively feasible for patients to handle) and the use of D-efficient designs will be no exception (18, 19). When a D-efficient design is used, patients do not have the opportunity to compare each characteristic explicitly to other characteristics (orthogonality), or patients had more often the opportunity to select a certain attribute-level (balance). Currently, the main issue is that the effect of using these less efficient designs on the reliability of individual preferences is unknown and has been unexplored.

Step 3: Instrument Design

In addition to this, when using D-efficient designs, HB analysis produces more reliable estimates compared to counts analysis. Yet HB analysis requires a certain knowledge of statistics and might not be as feasible to implement in clinical settings compared to counts analysis. Moreover, obtaining robust data at the individual level in real time can be difficult, because HB-analysis can be time consuming.

Conclusion

Several methodological challenges and uncertainties were encountered in using BWS to clarify values in individual patient decisions. We conclude that from a methodological perspective it is too early to promote widespread use of BWS in VCEs. Further research must first determine to which extent the use of D-efficient designs and different statistical analysis techniques affect the reliability of individual preference estimations. The actual implementation of this tool was beyond the scope of this paper, but another important theme will be to examine the feasibility and usefulness of such a VCE from the perspective of patients (Chapter 8 of this thesis).

REFERENCES

1. O'Connor AM, Rostom A, Fiset V, Tetroe J, Entwistle V, Llewellyn-Thomas H, et al. Decision aids for patients facing health treatment or screening decisions: Systematic review. *BMJ*. 1999;319(7212):731-4.
2. Fagerlin A, Pignone M, Abhyankar P, Col N, Feldman-Stewart D, Gavaruzzi T, et al. Clarifying values: an updated review. *BMC Med Inform Decis Mak*. 2013;13(2):1-7.
3. Witteman HO, Scherer LD, Gavaruzzi T, Pieterse AH, Fuhrel-Forbis A, Chipenda Dansokho S, et al. Design Features of Explicit Values Clarification Methods: A Systematic Review. *Med Decis Making*. 2016.
4. Pieterse AH, de Vries M, Kunneman M, Stiggelbout AM, Feldman-Stewart D. Theory-informed design of values clarification methods: A cognitive psychological perspective on patient health-related decision making. *Soc Sci Med*. 2013;77:156-63.
5. Louviere JJ, Flynn TN, Marley AAJ. *Best-Worst Scaling Theory, Methods and Applications*. Cambridge: Cambridge University Press; 2015. 360 p.
6. Mele NL. Conjoint analysis - Using a market-based research model for healthcare decision making. *Nurs Res*. 2008;57(3):220-4.
7. Flynn TN. Valuing citizen and patient preferences in health: recent developments in three types of best-worst scaling. *Expert Rev Pharmacoecon Outcomes Res*. 2010;10(3):259-67.
8. Lancaster KJ. A New Approach to Consumer Theory. *J Polit Econ*. 1966;74(2):132-57.
9. Ryan M, Farrar S. Using conjoint analysis to elicit preferences for health care. *BMJ*. 2000;320(7248):1530-3.
10. Potoglou D, Burge P, Flynn T, Netten A, Malley J, Forder J, et al. Best-worst scaling vs. discrete choice experiments: an empirical comparison using social care data. *Soc Sci Med*. 2011;72(10):1717-27.
11. Ali S, Ronaldson S. Ordinal preference elicitation methods in health economics and health services research: using discrete choice experiments and ranking methods. *Br Med Bull*. 2012;103(1):21-44.
12. Bridges JF, Hauber AB, Marshall D, Lloyd A, Prosser LA, Regier DA, et al. Conjoint analysis applications in health--a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. *Value Health*. 2011;14(4):403-13.
13. Ryan M, Skåtun D, Major K. Using Discrete Choice Experiments to Go Beyond Clinical Outcomes when Evaluating Clinical Practice. In: Ryan M, Gerard K, Amaya-Amaya M, editors. *Using Discrete Choice Experiments to Value Health and Health Care. The Economics of Non-Market Goods and Resources*. 11: Springer Netherlands; 2008. p. 101-16.
14. Pearmain D, Swanson J, Kroes E, Bradley M. *Stated Preference Techniques: A guide to practice*. 2 ed. Richmond / Den Haag: Steer Davies Gleave and Hague Consulting Group; 1991.
15. Orme B. *Formulating Attributes and Levels in Conjoint Analysis*. Sequim, WA: Sawtooth Software; 2002.
16. Abbruzzese G, Barone P, Bonuccelli U, Lopiano L, Antonini A. Continuous intestinal infusion of levodopa/carbidopa in advanced Parkinson's disease: efficacy, safety and patient selection. *Funct Neurol*. 2012;27(3):147-54.
17. Clarke CE, Worth P, Grosset D, Stewart D. Systematic review of apomorphine infusion, levodopa infusion and deep brain stimulation in advanced Parkinson's disease. *Parkinsonism Relat Disord*. 2009;15(10):728-41.
18. Flynn TN, Bilger M, Malhotra C, Finkelstein EA. Are Efficient Designs Used in Discrete Choice Experiments Too Difficult for Some Respondents? A Case Study Eliciting Preferences for End-of-Life Care. *Pharmacoeconomics*. 2016;34(3):273-84.
19. Johnson FR, Lancsar E, Marshall D, Kilambi V, Mühlbacher A, Regier DA, et al. Constructing experimental designs for discrete-choice experiments: Report of the ISPOR conjoint analysis experimental design good research practices task force. *Value Health*. 2013;16(1):3-13.
20. Rose JM, Bliemer MCJ. Constructing efficient stated choice experimental designs. *Transport Rev*. 2009;29(5):587-617.
21. Kuhfeld WF. *Marketing research methods in SAS*. 2010 [Available from: http://support.sas.com/resources/papers/tnote/t_note_marketresearch.html].
22. Orme B. *Accuracy of HB Estimation in MaxDiff Experiments*. Sequim, WA: Sawtooth Software; 2005.

23. Green PE, Srinivasan V. Conjoint analysis in marketing research: New developments and directions. *J Marketing*. 1990;54(4):3.
24. Hawley ST, Zikmund-Fisher B, Ubel P, Jancovic A, Lucas T, Fagerlin A. The impact of the format of graphical presentation on health-related knowledge and treatment choices. *Patient Educ Couns*. 2008;73(3):448-55.
25. Sawtooth Software. The CBC System for Choice-Based Conjoint Analysis version 8. Orem, UT: Sawtooth Software; 2013.
26. Auger P, Devinney TM, Louviere JJ. Consumer's social beliefs, an international investigation using Best-Worst scaling methodology. Working paper. Melbourne, Victoria, AU: University of Melbourne, Melbourne Business School,; 2004.
27. Orme B. MaxDiff Analysis: Simple Counting, Individual-Level Logit, and HB. Sequim, WA: Sawtooth Software; 2009.
28. McCaffery KJ, Dixon A, Hayen A, Jansen J, Smith S, Simpson JM. The influence of graphic display format on the interpretations of quantitative risk information among adults with lower education and literacy: A 1y. *Med Decis Mak*. 2012;32(4):532-44.
29. Feldman-Stewart D, Brundage MD, Manen LV, Svenson O. Patient-focussed decision-making in early-stage prostate cancer: insights from a cognitively based decision aid. *Health Expect*. 2004;7(2):126-41.
30. Lusk JL, Norwood FB. Effect of experimental design on choice-based conjoint valuation estimates. *Am J Agr Econ*. 2005;87(3):771-85.

8

Best-Worst Scaling Case 2 for Value Clarification and Decision Support: Patient's Experiences

Marieke G.M Weernink
Maarten J. IJzerman
Paul J.M. Kil
Jeroen P.P. van Vugt
Janine A. van Til

Submitted

ABSTRACT

Introduction: Best-Worst Scaling (BWS) is a stated preference technique which can be used to clarify treatment values to individual patients by asking them to evaluate the desirability of treatment characteristics. However, its potential in clinical decision making (assisting individual patients in treatment decisions) has not yet been studied. The aim of this study was to examine the feasibility and usability of a value clarification exercise based on BWS from the patient's perspective.

Methods: Following the presented framework in chapter 7, two value clarification exercises (VCEs) for treatment decisions in Parkinson's disease and Prostate Cancer were developed. Each VCE was pilot-tested in at least ten patients and evaluated on feasibility, cognitive demand, and usefulness of results.

Results: The VCE choice task design was challenged because of the required balance between response efficiency (what was cognitively feasible for patients to handle) and optimal design methodology (statistical efficiency). Assumptions had to be made to balance clinical and methodological preferences in the VCEs. Pilot-tests have shown that patients found the VCE useful and were willing to spend the extra time and effort. However, there were multiple practical issues encountered, such as recalling the range of possible outcomes on attributes, designating a best option from negative defined attribute-levels and needing assistance during completion.

Discussion: Our study offered useful insights in the use of BWS in VCEs. First applications resulted in positive feedback from patients. They had a positive attitude about the need to actively think about the relevant trade-offs and reported the VCE to be useful and informative, and thought this tool would better prepare them for their consult. However, from the patients point of view, there were also practical issues encountered. Much effort needs to be put on selecting the best-worst question, explanation of the task and the familiarization of patients with the varying attribute-levels.

INTRODUCTION

Following the presented framework in chapter 7 of this dissertation, two value clarification exercises (VCEs) for treatment decisions in Parkinson's disease and Prostate Cancer are developed. The presented framework consists of three procedural steps: decision problem structuring, choice task design and instrument design. In the following sections, the procedural step is shortly summarized (for detailed information the reader is referred to chapter 7), which is followed by its application in the case studies. We will discuss the methodological choices made in the design of these VCEs. Possibilities and limitations are described from both the perspective of researchers and clinicians. Eventually, the VCEs were pilot-tested to examine their feasibility and usefulness from the patient's perspective.

DEVELOPING THE VALUE CLARIFICATION EXERCISES

Decision Problem Structuring

First, the clinical decision problem and the range of treatment options need to be defined (Table 1). Next, the relevant attributes on which the treatments differ have to be selected. The selection of attributes is followed by the definition of levels for each attribute that reflect a realistic range of outcomes associated with the treatment options available. It is especially important to select attribute-levels that directly link to treatment options as other levels are valueless in informing the specific treatment decision.

Table 1. *Decision contexts of two pilot studies*

Parkinson's Disease	Localized Prostate Cancer
In later stages of Parkinson's Disease, oral pharmacological treatment alone is not effective in managing symptoms.	For localized prostate cancer (T1/T2) there is no clear treatment preference based on survival data.
Alternative treatments for patients are: Brain Lesions Deep Brain Stimulation Levodopa/Carbidopa Intraduodenal Gel Infusion Subcutaneous Apomorphine Infusion	Treatment options for patients are: Active Surveillance Radical prostatectomy Internal radiotherapy (brachytherapy) External beam radiotherapy (EBRT)
The treatments differ with regard to efficacy, side-effects and process of care.	The treatments differ with regard to side-effects and process of care.

Localized prostate cancer

For the treatment of localized prostate cancer there is no clear treatment preference based on survival data. For patients with localized prostate cancer treatment options vary between active surveillance strategies and curative treatments (radical prostatectomy, brachytherapy or external beam radiation). To support the decision making process between treatment options for these patients, a decision aid has already been tested and implemented in different hospitals throughout the Netherlands. This DA includes an overview of available treatment options and provides evidence-based information about the difference in treatment options and side effects and possible complications (1). This decision aid included lists of pros and cons of the four treatment alternatives. Selected attributes were derived from this list and had to be broadly supported by clinicians, and had to provide a balanced set of pros and cons of each treatment. Because of the initial large number of attributes (15) and the clinical context, two separate VCEs were designed to reflect two decision moments, namely the choice between active surveillance and curative treatments with 5 attributes (VCE-1), and the choice for a specific curative treatment with 6 attributes (VCE-2). Selected attributes and levels for the two VCEs can be found in Table 2 and Table 3.

Table 2. Attributes and levels of the prostate-cancer VCE-1: Choice between active surveillance and curative treatments.

Attributes	Active Surveillance	Curative Treatment
A Active Treatment	No curative treatment will take place until it is absolutely necessary (1)	Your prostate cancer will be actively treated with the aim to cure you from prostate cancer (2)
B Adverse events	There are no adverse events occurring with the treatment (1)	You may undergo unnecessary treatments which do not prolong your life (2)
C The probability of needing another treatment.	1 in 3 men needs another treatment within 4 years (1)	1 in 4 men needs another treatment within 4 years (2)
D Unnecessary treatments	You are prevented from unnecessary treatments (1)	You may undergo unnecessary treatments which do not prolong your life (2)
E Schedule of tests	There is a tight schedule of PSA measurements and prostate biopsies (1)	There is no tight schedule of check-up and tests (2)

Table 3. Attributes and levels of the prostate-cancer VCE-2: Choice between curative treatments: radical prostatectomy, internal radiotherapy (brachytherapy), and external beam radiotherapy (EBRT).

Attributes		Prostatectomy	Internal radiotherapy	External radiotherapy
A	PSA score	Your PSA score becomes 0 (1)	You will keep a measureable PSA score (2)	You will keep a measureable PSA score (2)
B	Possibility of receiving another treatment	If metastasis occur, another curative treatment is still possible (1)	If metastasis occur, another curative treatment is rarely possible (2)	If metastasis occur, another curative treatment is rarely possible (2)
C	The probability of having urinary problems	10 in 100 men have persistent urinary problems (1)	2 in 100 men have persistent urinary problems (2)	2 in 100 men have persistent urinary problems (2)
D	The probability of having bowel problems	Permanent bowel problems are extremely rare (1)	15 in 100 men have persistent bowel problems (2)	1 in 1000 men have persistent bowel problems (3)
E	The probability of erectile dysfunction	5 to 6 in 10 men experience erectile dysfunction (2)	4 to 9 in 10 men experience erectile dysfunction (2)	4 to 5 in 10 men experience erectile dysfunction (3)
F	Process of treatment	Major surgery where the prostate is removed (3)	Minor surgery, where radioactive seeds are implanted in the prostate (2)	Receive external radiation therapy for the total of 7 weeks, 5 days a week in the hospital (10 minutes per session) (3)

Parkinson's Disease

In later stages of Parkinson's Disease, oral pharmacological treatment alone is not effective in managing symptoms. Four advanced therapeutic options for these patients are: Brain lesions, Levodopa/Carbidopa Intraduodenal Gel Infusion via an external pump (Duodopa pump), Subcutaneous Apomorphine Infusion or deep brain stimulation (DBS). There is little data available comparing the effectiveness of these two interventions (2). Consequently, this often leaves the clinician in a difficult position when attempting to determine a therapeutic option for an individual patient, because the treatments do differ with regard to their risk, complications and process of care (3). A VCE might help to help patients evaluate the desirability of attributes or options. First, an extensive literature research resulted in a first selection of seven attributes. After consultation with the neurologist the formulation of levels were adjusted to improve

comprehension, but no attributes or levels were removed/added. Selected attributes and levels can be found in Table 4.

Table 4. Attributes and levels of the Parkinson's disease VCE: Choice between Brain Lesions, Deep Brain Stimulation, Levodopa/Carbidopa Intraduodenal Gel Infusion and Subcutaneous Apomorphine Infusion.

Attributes	Brain Lesions	Deep Brian Stimulation	Duodopa Gel Infusion	Apomorphine Infusion
A Route of administration	Brain surgery to damage a part of the brain (1)	Brain surgery to implant a neurostimulator (2)	Abdominal surgery and infusion of medication via a pump (3)	Infusion of medication via a pump (4)
B Maintenance	No maintenance needed (1)	Replacing battery requires a surgical procedure every 3-5 years (2)	The drug cassette needs to be changed once or twice a day (3)	Battery must be replaced every 3-4 weeks (4)
C Device	No use of a device (1)	An internal device is implanted (2)	An external device need to be worn (3)	An external device need to be worn (3)
D Dosage adjustment	Dosage adjustment is not needed (1)	Dosage can only be adjusted by a physician (2)	Dosage can be adjusted by patient or caregiver (3)	Dosage can be adjusted by patient or caregiver (3)
E Reduction motor fluctuations	Moderate reduction of motor fluctuations (50% of total daily duration) (1)	Very High reduction of motor fluctuations (80% of total daily duration) (2)	Very High reduction of motor fluctuations (80% of total daily duration) (2)	High reduction of motor fluctuations (70% of total daily duration) (3)
F Oral medication use	Reduction dose in oral medication use (1)	Reduction dose in oral medication use (1)	No oral medication use needed (2)	Reduction dose in oral medication use (1)
G Treatment reversibility	Treatment is not reversible (1)	Treatment is reversible (2)	Treatment is reversible (2)	Treatment is reversible (2)

Choice Task Design

When attributes and levels are selected, the next step is to create the hypothetical treatment scenarios. Decision have to be made upon the use of full or partial profiles, the number of choice tasks, and the corresponding best-worst question.

Localized prostate cancer

For a balanced and orthogonal design, the initial recommended amount of choice tasks was 16 in VCE-1. However, this was criticized by clinicians as they thought this was not cognitively feasible for patients to handle. Eventually, it was decided (at that time) that meeting the requirements of clinicians was more important than maintaining design efficiency, because the aim of this study was to test the feasibility of the VCE. The amount of choice tasks was decreased to six and level balance and orthogonality were at best approximated with the specified number of choice tasks.

For VCE-2 a D-efficient partial profile design was used. Patients had to answer 12 choice tasks, each choice task displayed five out of six attributes. With regards to the corresponding best-worst question our selection of attributes contained many negatively defined attributes (side-effects). Therefore some patients were interviewed to examine which dimension (most/least desirable, greatest advantage/disadvantage) most aligned with the task they were asked to perform. Patients indicated that the dimension most/least bothersome was most appropriate, because the negativity surrounding the attributes was reflected in this dimension. Figure 1 displays one choice task for VCE-1 (top) and VCE-2 (bottom)

Figure 1. Examples of VCE-1 and VCE-2 in the localized prostate cancer study.

What do you consider the most and least bothersome characteristic of treatment?		
Least bothersome characteristic 😊	Choicetask 1	Most bothersome characteristic 😞
☐	No curative treatment will take place until it is absolutely necessary	☐
☐	You run risk at adverse events such as urinary problems, bowel problems and / or erectile dysfunction	☐
☐	1 in 4 men needs another treatment within 4 years.	☐
☐	You are prevented from unnecessary treatments	☐
☐	There is no tight schedule of check-up and tests.	☐

What do you consider the most and least bothersome characteristic of treatment?		
Least bothersome characteristic 😊	Choicetask 1	Most bothersome characteristic 😞
☐	You will keep a measureable PSA score	☐
☐	If metastasis occur, another curative treatment is still possible	☐
☐	1 in 1000 men have persistent bowel problems	☐
☐	4 to 9 in 10 men experience erectile dysfunction	☐
☐	Receive external radiation therapy for the total of 7 weeks, 5 days a week in the hospital (10 minutes per session)	☐

Parkinson's Disease

Also in the Parkinson's Disease case D-efficient designs were used to balance the design efficiency and response efficiency. Furthermore, considering the seven selected attributes, it was discussed whether a full or partial profile choice task should be used. It was decided that pretests needed to determine whether answering 12 full profile tasks was more burdensome compared to answering 16 partial profile tasks (each task displaying four attribute-levels). The next step was to determine a suitable best-worst question. In the Parkinson case, not much time was spend on this matter. In consultation with the neurologist it was chosen to ask patients to select the most and least desirable characteristic in each choice task. Figure 2 displays choice tasks used in the full profile version (top) and partial profile version (bottom) of the VCE.

Figure 2. Choice tasks used in the full profile version (top) and partial profile version (bottom) of the Parkinson's VCE.

We would like you to indicate the most desirable characteristic and the least desirable characteristic by clicking the respective boxes.		
Most desirable characteristic	Choice task 1 of 12	Least desirable characteristic
☐	Infusion of medication via a pump	☐
☐	The drug cassette needs to be changed once or twice a day	☐
☐	An external device need to be worn	☐
☐	Dosage can be adjusted by patient or caregiver	☐
☐	Very High reduction of motor fluctuations (80% of total daily duration)	☐
☐	Reduction dose in oral medication use	☐
☐	Treatment is reversible	☐
Go to Next Question		

For each treatment we would like you to indicate the most desirable characteristic of treatment and the least desirable characteristic of treatment by clicking the respective boxes.		
Most desirable characteristic	Question 1 of 16	Least desirable characteristic
☐	Infusion of medication via a pump	☐
☐	Battery must be replaced every 3-4 weeks	☐
☐	Dosage can only be adjusted by the physician	☒
☒	Very high reduction of motor fluctuations (80% of total daily reduction)	☐
Go to Next Question		

Instrument Design

The next step is to embed the choice tasks into a value clarification tool. Decisions have to be made on the mode of administration, the type of statistical analysis, and the feedback patients will receive.

Localized prostate cancer

In the localized prostate cancer case, the VCE was preceded by an existing Web-based decision aid, therefore a Web-based survey program (Limesurvey (4)) was used to program the choice tasks. Furthermore, we aimed to show patients the results within seconds, but Limesurvey has limited computing capability and therefore a counts analysis was the maximum feasible. Hence, a treatment score was calculated by adding up the best-minus-worst scores of attribute-levels composing the specific treatment. A positive treatment score meant, that the patient chose characteristics of this treatment more often as least bothersome than as most bothersome. Based on the difference between treatment scores a mild/moderate/strong preference message was presented to patients (Figure 3).

Figure 3. Feedback provided to patients in the Prostate cancer study.

Parkinson's Disease

In the Parkinson case, there was no need to use a Web-based program and the VCE was programmed in Microsoft Excel. Using an add-in from XLStat, we were able to conduct instant HB analysis to derive individual utility values for each treatment characteristic (5) Compared to Limesurvey, Microsoft Excel provided us with more (graphical) options to visualize the results. First, a hierarchy of attribute-levels was presented, ranking them from most to least influential on

treatment choice. Next, the relative importance of each attribute was shown in a circle graph. Lastly, a treatment recommendation was presented in a bar graph showing how much one treatment was preferred relatively to the other, because total utility values were difficult to interpret (Figure 4).

Figure 4. Feedback provided to patients in the Parkinson’s Disease study

TESTING THE FEASIBILITY OF THE VALUE CLARIFICATION EXERCISES

Methods

Table 5 shows an overview of all design characteristics of the developed VCEs in the two case studies. Once the VCEs were programmed, think-aloud exercises and cognitive debrief interviews were used to assess the feasibility from the patient's point of view (6). To not interfere with the current decision making process, twelve patients with early-stage Parkinson's disease and ten men diagnosed with localized prostate cancer (who already chose treatment) were recruited. Written informed consent was obtained from participants. Patients used the VCE independently in the presence of the researcher and afterwards the results were discussed with the patient.

Table 5. Overview of all design characteristics of developed values clarification exercises.

	Prostate Cancer VCE-1	Prostate Cancer VCE-2	Parkinson's Disease VCE
Attribute & level selection			
Initial No. of attributes	15	15	7
No. of selected attributes	5	6	7
No. of selected levels	10 (2*5)	15 (3*3 + 2*3)	21 (4*2 + 3*3 + 2*2)
Choice task design			
No. of profiles possible	32 (2 ⁵)	216 (3 ³ * 2 ³)	7776 (4 ² * 3 ³ * 2 ²)
Full or partial profile	Full	Partial	Full/Partial profile
Selected No. of choice tasks	6	9 (5 items per profile)	Full profile: 12 Partial profile: 16 (4 items per profile)
Efficient or D-efficient design	D-efficient	D-efficient	D-efficient
Level-balance criterion	Yes, each level is 3 times shown	Yes, each level is 3 times shown	No, each level is at least 3 times shown
Orthogonality criterion	No	No	No
Selected dimension best-worst question	Most/least Bothersome	Most/least Bothersome	Most/least Desirable
Instrument design			
Statistical analysis	Counts analysis	Counts analysis	Hierarchical Bayes
Mode of administration	Computer	Computer	Computer
Web-based or Desktop	Web-based	Web-based	Desktop software
Software package	Limesurvey	Limesurvey	Microsoft Excel & XLStat
Presentation of results			
Ranking of attribute-levels	No	No	Yes
Attribute importance	No	No	Yes
Treatment recommendation	Yes	Yes	Yes
Strength of preference	Yes	Yes	Yes
Graphical illustrations	No	No	Yes

Results

The main finding across pilot studies was that all patients indicated they were willing to spend the extra time and effort to receive additional information about their preferences. All patients reported the VCE as useful and informative, and thought this tool would better prepare them for the consultation with the clinician. Specific comments on the three steps in the structured approach are discussed below.

Decision problem structuring

Patients were asked whether the VCEs included all relevant aspects of treatment. In the prostate cancer case, patients recognized the attributes from the preceding decision aid and they had no recommendations regarding adding or removing attributes. In the Parkinson study, the attributes were not verified yet by patient interviews or focus groups. Therefore these patients suggested to add one more attribute and one attribute-level. This emphasizes the crucial step of including patients in the process of selecting attributes and levels.

Choice task design

The main question of the pilot tests was whether the number of choice tasks was cognitively feasible for patients to handle. None of the patients complained about the number of choice tasks in the localized prostate cancer case (6 for VCE-1 and 9 for VCE-2). If necessary to receive the results, some patients were even willing to answer more choice tasks. Parkinson patients were asked to complete both the full and partial profile VCE. After six pilot tests, all six patients had indicated they preferred the partial profile version, although they spend less time on the full profile version. The internal validity scores from HB modeling also showed higher inconsistency in responses between tasks in the full profile version. Therefore, it was decided to only continue the pilot tests with the partial profile VCE.

With respect to the difficulty of the choice task itself: patients saw the presence of a researcher as an opportunity to ask questions regarding attribute/task clarification and the hypothetical nature of the choice task. Some patients tried to select one attribute-level as both most and least bothersome, indicating the choice task was not understood correctly based on the written description provided. However, when provided with an oral explanation by the research assistant all patients were able to complete the choice tasks. Yet it was noticed that some patients experienced difficulties with recalling the range of possible outcomes on attributes, because in each choice task only one outcome per attribute was displayed.

Another difference between the pilot studies concerned the best-worst question itself. Three patients with Parkinson found it very strange to choose a negative outcome as 'most desirable' choice. In the subsequent design of the prostate cancer VCE this was tested and discussed beforehand. Eventually, none of the patients in these pilot-tests commented on the requirement of selecting least/most bothersome characteristic.

Instrument design

None of the patients experienced problems using the interface of Limesurvey or Excel. With respect to the usability of results, patients were asked whether they felt the presented results were clear and comprehensible. The patients with prostate cancer only received a treatment recommendation with a mild/moderately strong statement (Figure 2). These patients indicated that they missed the underlying motivations upon which this recommendation was based. In the Parkinson case we were able to present importance scores, which were shown to increase the patient's comprehension. Patients were pleased with the treatment recommendation, and they were asked whether they felt it corresponded with their own expectation or intuition. Most patients could reason why the particular treatment was preferred for them, and were comfortable with the advice given.

DISCUSSION

Following the presented framework in chapter 7 of this dissertation, two value clarification exercises (VCEs) for treatment decisions in Parkinson's disease and Prostate Cancer were developed. Below, a summary is given with respect to this design process and the feasibility of a VCE based on BWS.

Design Process

The choice task design in both pilot studies was challenged because of the required balance between number of choice tasks and optimal design methodology (statistical design efficiency). Although perfect (balanced and orthogonal) design are preferred both case studies had to use D-efficient designs. Since the main focus of this study was on testing the feasibility of the VCE, it was decided (at that time) that reflecting the clinical context was more important than maintaining design efficiency. However, the important remark was made that before actual implementation takes place the reliability of individual preferences estimated with D-efficient design needs to be further investigated.

Feasibility and Usefulness

Although patients were positive about the usefulness of the results and the need to actively think about the relevant trade-offs, there were multiple practical issues encountered. As was previously observed by Whitty et al. (2014) (7), our results also show that patients (and clinicians) found it counter intuitive to indicate something as best when the treatment profile consisted of undesirable health outcomes. However, this can be prevented by improving the framing of the best-worst question (as was done in the prostate cancer pilot study). Furthermore, it is important to familiarize patients with the range of possible outcomes on attributes prior to the task (relative to a 10/100 chance of incontinence, a 2/100 change might be “better”) and it might be beneficial to have an overview of attributes and levels in sight during completion. Some cognitive difficulties with the BWS method did arise, because patients were not familiar with this type of question and the hypothetical nature of the choice task. Yet compared to other trade-off methods such as discrete choice experiments and direct elicitation instruments (time trade-off, standard gamble) literature suggests that BWS has a lower cognitive burden (8-10). If clinicians prefer the advantages of a conjoint analysis method (feedback on attribute importance and/or treatment recommendation), they must bear in mind that this type of VCE might not be suitable for each patient, or can be completed by anyone without assistance.

Conclusion

Our study offered useful insights in the use of BWS in VCEs. First applications resulted in positive feedback from patients. They had a positive attitude about the need to actively think about the relevant trade-offs and reported the VCE to be useful and informative, and thought this tool would better prepare them for their consult. Yet, multiple practical issues were identified. Much effort needs to be put on selection of the best-worst question, explanation of the task and the familiarization of patients with the varying attribute-levels.

REFERENCES

1. Cuypers M, Lamers RED, Kil PJM, van de Poll-Franse LV, de Vries M. Impact of a web-based treatment decision aid for early-stage prostate cancer on shared decision-making and health outcomes: Study protocol for a randomized controlled trial. *Trials*. 2015;16(1).
2. Abbruzzese G, Barone P, Bonuccelli U, Lopiano L, Antonini A. Continuous intestinal infusion of levodopa/carbidopa in advanced Parkinson's disease: efficacy, safety and patient selection. *Funct Neurol*. 2012;27(3):147-54.
3. Clarke CE, Worth P, Grosset D, Stewart D. Systematic review of apomorphine infusion, levodopa infusion and deep brain stimulation in advanced Parkinson's disease. *Parkinsonism Relat Disord*. 2009;15(10):728-41.

4. LimeSurvey Project Team, Schmitz C. LimeSurvey: An Open Source survey tool. Hamburg, Germany: LimeSurvey Project; 2015.
5. Addinsoft. XLSTAT: Hierarchical Bayes add-in in Excel. NY, USA: Addinsoft; 2015.
6. Collins D. Pretesting survey instruments: An overview of cognitive methods. *Qual Life Res.* 2003;12(3):229-38.
7. Whitty JA, Walker R, Golenko X, Ratcliffe J. A think aloud study comparing the validity and acceptability of discrete choice and best worst scaling methods. *PloS one.* 2014;9(4):e90635.
8. Potoglou D, Burge P, Flynn T, Netten A, Malley J, Forder J, et al. Best-worst scaling vs. discrete choice experiments: an empirical comparison using social care data. *Soc Sci Med.* 2011;72(10):1717-27.
9. Ali S, Ronaldson S. Ordinal preference elicitation methods in health economics and health services research: using discrete choice experiments and ranking methods. *Br Med Bull.* 2012;103(1):21-44.
10. Louviere JJ, Flynn TN, Marley AAJ. *Best-Worst Scaling Theory, Methods and Applications.* Cambridge: Cambridge University Press; 2015. 360 p.

9

General Discussion

TREATMENT PREFERENCES IN PARKINSON'S DISEASE

The objective of this dissertation was to investigate the influence of process-utility (which is the value attached to the process and convenience of care without reference to the outcome) on the relative value of treatments. Process-utility is then studied in different chapters with reference to societal health care and (individual) shared clinical decisions. The thesis thereby specifically applied Best-Worst Scaling (BWS) to explore process-utility, a relatively new method for preference elicitation. The following sections will present our main findings and implications for clinical practice.

Impact and Use of Process-Utility in Societal Decision Making

In **Chapter 3** it was concluded that the process of care was the most important driver of patient preferences, because invasive pump infusion and brain surgery have a detrimental effect on value of treatment assuming the health benefits are equal. Furthermore, the occurrence of described side effects (dizziness, drowsiness, and dyskinesia) had less impact on the perceived desirability of treatment than the treatment's effect on motor symptoms (slow movement, posture and balance problems and tremor).

Although different views exist towards including process-utility in societal decision making, the general trend is that many people advocate widening the evaluation space and their arguments gain momentum (1, 2). A growing body of literature exists which measures procedural or process-utility (3-5). However, this does not directly influence its use in economic evaluations and societal decision making (6). Small steps are being made, for instance the National Institute for Health and Clinical Excellence in the United Kingdom still emphasizes health outcomes for its reference case, but now allows consideration of evidence on process characteristics (7).

"If characteristics of healthcare technologies have a value to people independent of any direct effect on health, the nature of these characteristics should be clearly explained and if possible the value of the additional benefit should be quantified. These characteristics may include convenience and the level of information available for patients" (7).

In reimbursement decisions, not taking process-factors into consideration might lead to suboptimal resource allocation, and can result in a mismatch between health care policy and actual

resource use (8). By estimating process-utility, decision makers receive information on whether including process of care supports or conflicts to evaluations based on solely health outcomes (4).

Studies which have focused on elicitation of societal preferences of reimbursement criteria only focused on the most common and currently used criteria such as: disease severity, unmet needs, significant innovation, clinical benefits and cost-effectiveness (9-11). The public preferences towards taking into account a criteria about process of care are unknown and are an topic for future research.

To our best knowledge, the Royal Decree in Belgium is the only Health Technology Assessment (HTA) agency who has selected discomfort of treatment as one of the decision criteria which defines the therapeutic value of a healthcare innovation (12). Since pharmacoeconomic evaluations are focused on health outcomes (morbidity, mortality and health-related quality of life), additional research and analysis is then necessary to measure the applicability and user-friendliness of the product or innovation (13).

Process-Utility in the General Public and in Patients

Because preferences informing societal decisions are commonly derived from the general public, rather than from actual patients [12], we also studied the differences between public and patients.

In **chapter 4**, public preferences were elicited for PD treatment attributes and it was shown that the process of care is also the most important driver of preferences for the public population. However, in comparison to the other attributes, the patients attached higher importance to treatment modality than the public population. Patients also had a more negative perception of the infusion of medication via a pump. Most health outcomes were valued as equally important by patients and public, except for dizziness and slow movement.

For instance, compared to the patient population, public respondents underestimated the effect of improved movement and gait. Consequently, new PD interventions which will improve this motor function will appear cost-effective from the patient perspective, but receive less attractive cost-effectiveness ratios from the public's point of view. As has been extensively discussed in literature, both public and patients' preferences entail information that is potentially relevant for societal decision makers (14, 15). For instance, knowing both perspectives can inform decision makers about the extent to which patients adapt to their condition, by comparing patient and public utility estimates. Moreover, in our study predicted uptake of invasive treatment is lower in

patients compared to the public, which could indicate that desirability of a treatment is overestimated if its value is based on public preferences. These information could inform reimbursement and investment decision making (16). Since, patient and public preferences are not necessarily mutually exclusive, more initiatives are being employed to use patient preferences alongside public preferences in societal decision making (17, 18).

How to Measure Process-Utility?

As described in the general introduction, conjoint analysis methods were introduced as an attempt to go beyond health outcomes and to take account of non-health outcomes and process attributes in the delivery of health care (19, 20). However, in our head-to-head comparison in **chapter 5** the results indicated that the methods BWS case 2, BWS case 3, Time Trade-Off (TTO) and Visual Analogue Scales (VAS) perform equally well with regard to distinguishing treatment alternatives on the basis of aggregated process-utility scores (no rank reversals and highly comparable). The main shortcoming of conjoint analysis (BWS), which limits its use in the HTA context, is that the resulting utilities are measured on a latent scale. Without anchoring and mapping techniques they cannot be used to measure Quality Adjusted Life Years, which are favored in economic analyses (21, 22). However, the results in chapter 5 also showed that individual respondents differentiated less in their stated utility of closely-related treatment profiles using a VAS or TTO (non-significant differences) compared to the estimated utilities of BWS. Therefore, it was concluded that on an individual patient-level, such as in shared decision making, the composed nature of BWS, where treatment attributes rather than treatment as a whole are evaluated, allows for better prioritization of treatment alternatives, especially if treatments share characteristics/attributes.

Incorporating Process-Utility in Value Clarification Experiments

In contrast to its use in societal decision making, shared decision making requires that non-health outcomes or process characteristics are explicitly discussed with patients. What, however, is at issue here is how preferences (process-utility) should be measured. Currently, shared decision making mainly focuses on informing (and educating) the patient with the best evidence-based information possible in decision aids (23). However, a very important issue that might currently be overlooked is the need to better support patients in understanding how different options fit (or do not fit) with what matters to them. It can be questioned whether patients are able to weigh up the possible advantages and disadvantages of the different treatment options on their own, without further

assistance or decisions support tools. In current clinical practice, weighting benefits and harms, even after the use of decision aids, is often performed with the clinician and/or a nurse practitioner. Alternatively, patients are asked to rate and/or rank attributes of care on a single dimensional or multidimensional scale. The interpretation of the effect of these preferences is then left to the clinical practitioner. The main advantage of BWS and other CA-methods is that it offers a way to help patients see how each treatment option aligns or does not align with what matters to them, and how their preferences would influence treatment choice.

In **chapter 6**, a systematical review was conducted to study the use of conjoint analysis to support preference elicitation for clinical decision support. In this review, eight studies were found and all used adaptive conjoint analysis as their preference elicitation method. However, although it was methodologically possible in these studies, it was uncommon to present patients with an explicit treatment recommendation based on their treatment preferences. Presenting patients with an algorithmic treatment recommendation raises a number of issues. First, patients may not want to have a computer tell them what their best option is. Second, physicians might feel hesitant in case the treatment recommended is not the best treatment option according to the physician's experience or the evidence available (24). Third, shared decision making has historically shied away from making treatment recommendations (10, 11). Providing the patient with results on attribute importance leaves more room for interpretation and discussion during shared decision making; however, it is arguable that helping patients to better understand how each treatment option aligns or does not align with what matters to them is a critical—and too often overlooked—step in supporting evidence-informed, values-congruent decisions. However, the question remains whether this should be done by a algorithmic treatment recommendation or by the doctor interpreting the results on attribute importance.

The last two chapters (**chapter 7 and 8**) of this thesis focused on the use of BWS case 2 as a value clarification method to help individual patients clarify their values in treatment decisions. First applications resulted in positive feedback from patients. Patients had a positive attitude about the need to actively think about the relevant trade-offs and were satisfied with the way they received a treatment recommendation. However, from the patients point of view, there were also practical issues encountered with the BWS method. As was previously observed by Whitty et al. (2014) (25), it was reported that patients (and clinicians) found it counter intuitive to indicate something as best when the treatment profile consisted of undesirable health outcomes. However, this can be prevented by improving the framing of the best-worst question. Furthermore, it is important to

familiarize patients with the range of possible outcomes on attributes prior to the task (relative to a 10/100 chance of incontinence, a 2/100 change might be “better”) and it might be beneficial to have an overview of attributes and levels in sight during completion. Besides these practical issues, there were several methodological challenges encountered in the design phase of these VCEs. In general, balanced and orthogonal experimental designs are preferred, but they are often not available or require an impracticable number of choice task (26). BWS lets patients explicitly and actively think about the relevant trade-offs between attributes and attribute-levels, but if patients are forced to use fast and frugal heuristics of simplifying decision strategies due to the number of choice tasks, then BWS fails to achieve its objectives. In practice, statistical or design efficiency (how well the attributes in the design are balanced and orthogonal) will be balanced against response efficiency (what is cognitively feasible for patients to handle) and as a result less efficient designs will often be used (27, 28). When less efficient design are used, patients do not have the opportunity to compare each characteristic explicitly to other characteristics (orthogonality), or patients had more often the opportunity to select a certain attribute-level (balance). One of the challenges here is that it is unknown and unexplored to which extent these methodological compromises affect the reliability of individual preference estimations. Another issue is the difficulty of obtaining robust data at the individual level in real time (conducting e.g. Hierarchical Bayes analysis on the preference data is time-consuming).

Our conclusion is then as follows: the methodological and practical challenges encountered in our research withhold us from supporting the use of BWS case 2 as a value clarification method in clinical decisions. Although practical issues can be overcome (as suggested above), it is important that further research first determines to which extent methodological compromises affect the reliability of individual preference estimations.

Preference Elicitation Using Best-Worst Scaling

Throughout this thesis, BWS case 2 is used within different settings (societal and clinical decision making) and with different stakeholders (patient and general public). In the literature it is suggested that BWS case 2 tasks are generally “easier” than multi-profile ones (discrete choice experiments, BWS case 3) (29, 30). The main perceived advantage of BWS case 2 is that only one scenario is displayed at a time and less information has to be processed by respondents.

In our research (**chapter 3**), we conducted pilot tests to determine which BWS task would be most feasible in the PD patient group. PD patients (already in an early stage) might encounter cognitive

problems, in particular with short term memory and executive function (31). It was confirmed that PD patients had difficulties engaging in a BWS case 3 experiment due to information overload and the complexity of comparing multiple treatment profiles in one choice task. Information overload increased the tendency of patients to ignore one or more attributes, which eventually can lead to biased coefficient estimates (32, 33). In contrast, BWS case 2 choice tasks were well-received and patients indicated that this choice task was easier and faster to complete. On the basis of the existing literature and personal experience during pilot testing it is assumed that the use of BWS-2 might stimulate the comprehension of stated-preference research in vulnerable patient groups.

As age and educational background may impact the conclusions, older (>75) and less educated patients were excluded from the study in **chapter 3** more often due to response errors in their paper-questionnaires (ticked multiple best/worst boxes in one scenario, chose only the most desirable characteristics, or left multiple questions blank). One way to overcome response errors is the use of online surveys, because error messages and extra information can be presented during completion. However, in this age group, computer access and computer literacy might be an issue (34). It is recommended that further research should compare the feasibility of different stated preference methods in vulnerable patient groups in larger sample sizes.

Conclusion

The objective of this dissertation was to investigate the role of process-utility in health care evaluations as well as in shared decision making. This thesis clearly shows that the process of care is an important driver of preferences, in both patients and in the general public. Where incorporation of process-utility in shared decision making is very common, the use of process-utility in societal decision making still is a long way. In contrast to societal decision making, the use of stated preference methods with a strong theoretical foundation is not common in clinical decision making. This thesis has found that BWS case 2 has potential in shared decision making, but methodological issues (e.g. reliability of individual preference estimations) should first be investigated before widespread implementation can take place.

REFERENCES

1. Anscombe GEM. Modern Moral Philosophy. *Philosophy*. 1958;33(124):1 - 19.
2. Mooney G. Beyond health outcomes: the benefits of health care. *Health Care Anal*. 1998;6(2):99-105.
3. Brennan VK, Dixon S. Incorporating process utility into quality adjusted life years: a systematic review of empirical studies. *Pharmacoeconomics*. 2013;31(8):677-91.
4. Liberatore M, Nydick R, Daskalakis C, Kunkel E, Cocroft J, Myers R. Helping men decide about scheduling a prostate cancer screening exam. *Interfaces*. 2009;39(3):209-17.
5. Richman MB, Forman EH, Bayazit Y, Einstein DB, Resnick MI, Stovsky MD. A novel computer based expert decision making model for prostate cancer disease management. *J Urol*. 2005;174(6):2310-8.
6. Liberatore MJ, Myers RE, Nydick RL, Steinberg M, Brown ER, Gay R, et al. Decision counseling for men considering prostate cancer screening. *Comput Oper Res*. 2003;30(10):1421-34.
7. National Institute for Health and Clinical Excellence. Guide to the methods of health technology appraisal. 2013.
8. Higgins A, Barnett J, Meads C, Singh J, Longworth L. Does convenience matter in health care delivery? A systematic review of convenience-based aspects of process utility. *Value Health*. 2014;17(8):877-87.
9. Linley WG, Hughes DA. Decision-makers' preferences for approving new medicines in Wales: a discrete-choice experiment with assessment of external validity. *Pharmacoeconomics*. 2013;31(4):345-55.
10. Frongillo M, Feibelman S, Belkora J, Lee C, Sepucha K. Is there shared decision making when the provider makes a recommendation? *Patient Educ Couns*. 2013;90(1):69-73.
11. Fried TR. Shared Decision Making — Finding the Sweet Spot. *N Engl J Med*. 2016;374(2):104-6.
12. O'Connor AM, Rostom A, Fiset V, Tetroe J, Entwistle V, Llewellyn-Thomas H, et al. Decision aids for patients facing health treatment or screening decisions: Systematic review. *BMJ*. 1999;319(7212):731-4.
13. Cleemput I, Devriese S, Kohn L, Devos C, van Til J, Groothuis-Oudshoorn K, et al. Incorporating societal preferences in reimbursement decisions – Relative importance of decision criteria according to Belgian citizens. Brussels: Belgian Health Care Knowledge Centre (KCE), 2014 Contract No.: KCE Reports 234.
14. Drummond M, Brixner D, Gold M, Kind P, McGuire A, Nord E. Toward a consensus on the QALY. *Value Health*. 2009;12(SUPPL. 1):S31-S5.
15. Brazier J, Akehurst R, Brennan A, Dolan P, Claxton K, McCabe C, et al. Should patients have a greater role in valuing health states? *Appl Health Econ Health Policy*. 2005;4(4):201-8.
16. Dowie J, Kjer Kaltoft M, Salkeld G, Cunich M. Towards generic online multicriteria decision support in patient-centred health care. *Health Expect*. 2015;18(5):689-702.
17. Versteegh MM, Brouwer WBF. Patient and general public preferences for health states: A call to reconsider current guidelines. *Soc Sci Med*. 2016;165:66-74.
18. Abelson J, Wagner F, DeJean D, Boesveld S, Gauvin FP, Bean S, et al. Public and Patient Involvement in Health Technology Assessment: A Framework for Action. *Int J Technol Assess Health Care*. 2016:1-9.
19. Ryan M, Hughes J, Economics UCDCfH, Unit UoAHER. Using Conjoint Analysis to Value Surgical Versus Medical Management of Miscarriage: University College Dublin, Centre for Health Economics; 1995.
20. Ryan M, Bate A, Eastmond CJ, Ludbrook A. Use of discrete choice experiments to elicit preferences. *Qual Health Care*. 2001;10(SUPPL. 1):i55-i60.
21. Rowen D, Brazier J, Van Hout B. A comparison of methods for converting DCE values onto the full health-dead QALY scale. *Med Decis Making*. 2015;35(3):328-40.
22. Brazier J, Rowen D, Yang Y, Tsuchiya A. Comparison of health state utility values derived using time trade-off, rank and discrete choice data anchored on the full health-dead scale. *Eur J Health Econ*. 2012;13(5):575-87.
23. Elwyn G, Connor A, Stacey D, Volk R, Edwards A, Coulter A, et al. Developing a quality criteria framework for patient decision aids: online international Delphi consensus process. *BMJ*. 2006;333(7565):417.

24. Fraenkel L. Incorporating patients' preferences into medical decision making. *Med Care Res.* 2013;70(1 Suppl):80s-93s.
25. Whitty JA, Walker R, Golenko X, Ratcliffe J. A think aloud study comparing the validity and acceptability of discrete choice and best worst scaling methods. *PloS one.* 2014;9(4):e90635.
26. Lusk JL, Norwood FB. Effect of experimental design on choice-based conjoint valuation estimates. *Am J Agr Econ.* 2005;87(3):771-85.
27. Flynn TN, Bilger M, Malhotra C, Finkelstein EA. Are Efficient Designs Used in Discrete Choice Experiments Too Difficult for Some Respondents? A Case Study Eliciting Preferences for End-of-Life Care. *PharmacoEconomics.* 2016;34(3):273-84.
28. Johnson FR, Lancsar E, Marshall D, Kilambi V, Mühlbacher A, Regier DA, et al. Constructing experimental designs for discrete-choice experiments: Report of the ISPOR conjoint analysis experimental design good research practices task force. *Value Health.* 2013;16(1):3-13.
29. Potoglou D, Burge P, Flynn T, Netten A, Malley J, Forder J, et al. Best-worst scaling vs. discrete choice experiments: an empirical comparison using social care data. *Soc Sci Med.* 2011;72(10):1717-27.
30. Louviere JJ, Flynn TN, Marley AAJ. *Best-Worst Scaling Theory, Methods and Applications.* Cambridge: Cambridge University Press; 2015. 360 p.
31. Muslimović D, Post B, Speelman JD, Schmand B. Cognitive profile of patients with newly diagnosed Parkinson disease. *Neurology.* 2005;65(8):1239-45.
32. Lagarde M. Investigating attribute non-attendance and its consequences in choice experiments with latent class models. *Health Econ.* 2013;22(5):554-67.
33. Alemu MH, Mørkbak MR, Olsen SB, Jensen CL. Attending to the Reasons for Attribute Non-attendance in Choice Experiments. *Environ Resour Econ.* 2013;54(3):333-59.
34. Rochon D, Eberth JM, Fraenkel L, Volk RJ, Whitney SN. Elderly patients' experiences using adaptive conjoint analysis software as a decision aid for osteoarthritis of the knee. *Health Expect.* 2014;17(6):840-51.

Summary

The objective of this dissertation was to investigate the influence of process-utility (which is the value attached to the process and convenience of care without reference to the outcome) on the relative value of treatments. Process-utility is then studied in different chapters with reference to societal health care and (individual) shared clinical decisions. The three main objectives for the three different parts of this thesis were:

- Part I: Summarize the available literature on the use of stated preference methods in different healthcare decision contexts.
- Part II: Estimating process-utility for treatments in Parkinson's Disease (PD) using different stated preference methods and for different perspectives.
- Part III: Explore the use of BWS to measure process-utility in clinical decision making.

Part I: Preference Elicitation Methods in Health Care

The first objective of this thesis was to summarize the available literature on the use of stated preference methods in different healthcare decision contexts. In chapter 2 it was found that standard gamble and Time Trade-Off (TTO) were most often used in economic analysis, while in clinical decision making Visual Analogue Scales (VAS), direct choice and rankings were more preferred. In benefit-risk assessment only Discrete Choice Experiments were used to inform market access decisions. Furthermore, it was concluded that valuing the influence of process of care on treatment value appears most relevant in clinical decisions. To a lesser extent, process of care was valued in articles in the context of Health Technology Assessment (HTA) and reimbursement decisions.

Part II: Estimating Process-Utility for Treatments in Parkinson's Disease

The second objective of this dissertation was to investigate the influence of the process and convenience of care on the relative value of treatments in PD. In chapter 3, patient preferences for benefits, side effects and process characteristics of PD treatments were elicited. It was concluded that the process of care was the most important driver of patient preferences in PD. Treatment modalities pump infusion and brain surgery have a detrimental effect on value of treatment assuming the health benefits are equal. Furthermore, the occurrence of described side effects (dizziness, drowsiness, and dyskinesia) had less impact on the perceived desirability of treatment than the treatment's effect on motor symptoms (slow movement, posture and balance problems and tremor). Chapter 4 shows that both the public and patients considered treatment modality as

the most important attribute of PD treatment, although patients attached higher relative importance. Both groups perceived neurosurgery as the most negative treatment modality. Yet patients valued the infusion of medication via a pump more negatively and the oral intake of medication slightly more positively compared to the public population. Most health outcomes were valued as equally important by patients and public, except for dizziness (higher public value) and slow movement (higher patient value).

In these chapters, Best-Worst Scaling (BWS), a relatively new method for preference elicitation, was applied to explore the value of process of care on preference for treatment. BWS measures the importance of attributes by asking the respondents repeatedly to indicate the best and worst item from a subset of items. In chapter 5 it was investigated which preference method (TTO, VAS, BWS case 2 or BWS case 3) is best to estimate process-utility. The results show that all methods perform equally well with regard to distinguishing treatment alternatives on the basis of aggregated process utility scores. However, individual respondents differentiated less in their stated utility of closely-related treatment profiles using a VAS or TTO. Therefore, it was concluded that on an individual patient-level, such as in shared decision making, the composed nature of BWS, where treatment attributes rather than treatment as a whole are evaluated, allows for better prioritization of treatment alternatives (especially if treatments share characteristics or attributes).

Part III Value Clarification in Clinical Practice

Currently, decision aiding in the clinical context mainly focuses on informing (and educating) the patient with the best evidence-based information possible using decision aids. However, a very important issue that might be overlooked is the need to better support patients in understanding how different options fit (or do not fit) with what matters to them. In chapter 6 we systematically reviewed the literature to examine the use of Conjoint Analysis (CA) methods to elicit preferences for clinical decision support on the individual patient level. CA methods have potential in clinical practice because of their ability to estimate the relative value of characteristics and treatment options. It was found that only eight studies have explored the use of CA in clinical practice and all used the adaptive conjoint analysis technique. Most studies resembled each other in design choices made, but patients received different feedback across studies (some patients only received their preference regarding attribute importance, others also were given a treatment recommendation). Overall, patients across included studies reported that the CA-exercises were easy to do, clear and interesting.

No existing applications of BWS was found in the literature review and therefore the last objective of this thesis focused on the use of BWS case 2 as a value clarification method to help individual patients clarify their values in treatment decisions (chapter 7 and 8). First applications in pilot-studies resulted in positive feedback from patients. Patients had a positive attitude about the need to actively think about the relevant trade-offs and were satisfied with the way they received a treatment recommendation. However, from the patients point of view, there were also practical issues encountered with the BWS method. For instance, patients experienced problems with recalling the range of possible outcomes on attributes, selecting a best option from negative defined attribute-levels and needing assistance during completion (to explain the task). Besides these practical issues (which can be overcome), there were several methodological challenges encountered in the design phase of these VCEs. In general, balanced and orthogonal experimental designs are preferred, but they are often not available or require patients to answer an impracticable number of choice task. The aim of BWS is to let patients explicitly and actively think about the relevant trade-offs between attributes and attribute-levels. However, if patients are forced to use fast and frugal heuristics of simplifying decision strategies due to the high number of choice tasks, then BWS fails to achieve its objectives. In practice, statistical or design efficiency (how well the attributes in the design are balanced and orthogonal) will be balanced against response efficiency (what is cognitively feasible for patients to handle) and as a result less efficient designs will often be used. One of the challenges here is that it is unknown and unexplored to which extent these methodological compromises affect the reliability of individual preference estimations.

CONCLUSION

This thesis clearly shows that the process of care is an important driver of preferences, in both patients and in the general public. Where incorporation of process utility in shared decision making is very common, the use of process utility in societal decision making still is a long way. In contrast to societal decision making, the use of stated preference methods with a strong theoretical foundation is not common in clinical decision making. This thesis has found that BWS has potential in shared decision making, but methodological issues (e.g. reliability of individual preference estimations) should first be investigated before widespread implementation can take place.

Samenvatting

Centraal in dit proefschrift staat het meten van de invloed van het 'proces van zorg' op de waarde van behandelingen. Dit wordt ook wel proces-utiliteit genoemd, de waarde die samenhangt met het proces en gemak van zorg, zonder dat hierbij naar de uitkomsten wordt gerefereerd. In dit proefschrift is proces-utiliteit gemeten in verschillende vormen van besluitvorming, onder andere op groepsniveau (maatschappelijk perspectief) en op het individuele niveau (besluitvorming tussen arts en patiënt). Er zijn drie doelen opgesteld voor de drie verschillende delen van dit proefschrift.

- Deel I: Samenvatten van de beschikbare literatuur over het gebruik van verschillende preferentie instrumenten in verschillende contexten van besluitvorming.
- Deel II: Het meten van proces-utiliteit bij de ziekte van Parkinson (ZvP) door middel van verschillende preferentie instrumenten en voor verschillende perspectieven.
- Deel III: Exploreren of BWS gebruikt kan worden om proces-utiliteit te meten in individuele patiënten (voor gedeelde besluitvorming tussen arts en patiënt).

Deel I: Preferentie Instrumenten in de Gezondheidszorg

Dit proefschrift begint met een samenvatting van de beschikbare literatuur over het gebruik van preferentie instrumenten in verschillende contexten van besluitvorming. In hoofdstuk 2 wordt geconcludeerd dat de Standard Gamble en Time Trade-Off (TTO) veelal gebruikt worden in economische analyses, daarentegen wordt in klinische beslissingen de voorkeur gegeven aan Visueel Analoge Schalen (VAS), directe keuzes en rangordeningen. Artikelen over de afweging tussen effectiviteit en risico's van nieuwe geneesmiddelen worden alleen discrete keuze experimenten gebruikt. De kenmerken waarover voorkeuren worden bepaald verschillen per type beslissing. Het meten van het belang van 'proces van zorg' is het meest relevant in klinische beslissingen en werd in mindere mate gevonden in artikelen over vergoedingsbeslissingen en health technology assessments.

Deel II: Het Meten van Proces-Utiliteit voor Behandelingen van de Ziekte van Parkinson

In deel 2 van dit proefschrift wordt de invloed van 'proces van zorg' (proces-utiliteit) op de relatieve waarde van behandelingen voor de ZvP gemeten. In hoofdstuk 3 worden de voorkeuren van patiënten voor kenmerken van behandelingen gemeten: symptomen, bijwerkingen en behandelingsvorm. De verschillende vormen van behandeling zijn: inname van tabletten, een

continue afgifte van de medicatie door een pomp of neurochirurgie. De resultaten van deze studie hebben aangetoond dat de behandelingsvorm de grootste invloed heeft op de wenselijkheid van behandeling. De inname van tabletten heeft een positief effect op de waarde van behandeling, waar de diepe hersen stimulatie en de pomp een negatief effect hebben. Daarnaast is gebleken dat het effect van behandeling op motorische symptomen: tremoren, houdings en balansproblemen, bewegingstraagheid een grotere invloed heeft op de wenselijkheid van een behandeling dan de beschreven bijwerkingen (duizeligheid, slaperigheid, overbeweeglijkheid).

In hoofdstuk 4 worden de voorkeuren van de patiënten vergeleken met de voorkeuren van de algemene bevolking voor behandelingskenmerken voor de ZvP. De resultaten tonen aan dat beide groepen de behandelingsvorm als het meest belangrijke kenmerk van behandeling beschouwen. Door beide groepen wordt de diepe hersen stimulatie als meest negatief beoordeeld. Echter patiënten geven een negatievere beoordeling aan de pomp en een iets positievere beoordeling aan de inname van tabletten in vergelijking met de algemene bevolking. De meeste gezondheidsuitkomsten (symptomen en bijwerkingen) worden door beide groepen als even belangrijk gewaardeerd, met uitzondering van duizeligheid (hogere waarde van de algemene bevolking) en bewegingstraagheid (hogere waarde van patiënten).

Om de invloed van ‘proces van zorg’ (proces-utiliteit) te meten is in voorgaande hoofdstukken gebruik gemaakt van een relatief nieuwe preferentiemethode: Best-Worst Scaling (BWS). Hierbij moeten respondenten kiezen welk kenmerk voor hem/haar het beste is en welk kenmerk het slechtste is uit verschillende subsets van de complete lijst met kenmerken. Het doel van hoofdstuk 5 is onderzoeken welke preferentiemethode het beste proces-utiliteit kan meten: BWS case 2, BWS case 3, TTO of VAS. Op basis van geaggregeerde utiliteiten konden de respondenten de verschillende behandelingsscenario's goed onderscheiden met alle methodes (er waren geen rank reversals en utiliteitscores waren vergelijkbaar). Echter, individuele respondenten hebben meer moeite om met een TTO of VAS een verschil aan te geven in utiliteitscores voor de meer vergelijkbare behandelingen dan met BWS case 2 en BWS case 3. Dus met name voor de waardering van behandelingen die nauw samenhangen, zorgt de BWS methode voor een betere prioritering van behandelingsopties voor individuele patiënten.

Deel III: Waarden en preferenties in de klinische besluitvorming

De hulp die patiënten krijgen bij het maken van beslissingen richt ze nu vooral op het informeren en opleiden van patiënten met de evidence-based informatie in zogenoemde keuzehulpen. Echter,

een zeer belangrijk punt dat nu over het hoofd wordt gezien is de noodzaak om patiënten te laten zien hoe de verschillende behandelopties wel of niet bij hem of haar passen. Conjoint Analysis (CA) technieken (waar BWS ook onder valt) kunnen hierbij een mogelijk hulpmiddel zijn, omdat ze de relatieve waarde van kenmerken, maar ook behandelingsopties kunnen aangeven op basis van de patiënt zijn of haar antwoorden. In de systematische literatuurstudie in hoofdstuk 6 is gekeken naar het huidige gebruik van CA technieken in zogenoemde 'value clarification exercises' (VCE) om patiënten te ondersteunen bij het maken van beslissingen. Acht studies werden geïncludeerd en allen maakten gebruik van de adaptieve CA techniek. De studies waren vergelijkbaar in de gemaakte keuzes met betrekking tot het ontwerp van de VCE, maar patiënten ontvingen wel verschillende resultaten (sommige patiënten ontvingen alleen informatie over hun voorkeuren over behandelingskenmerken, anderen ontvingen ook een behandelingsadvies). Over het algemeen vonden de patiënten de CA-opgaves makkelijk, duidelijk en interessant.

In hoofdstuk 6 werden geen artikelen gevonden waarin BWS werd toegepast in een VCE. In deel 3 van dit proefschrift is daarom onderzocht of BWS case 2 individuele patiënten kan helpen met het verduidelijken van hun waarden en voorkeuren in behandelingsbeslissingen (hoofdstuk 7 en 8). De eerste toepassingen in pilotstudies resulteerden in positieve feedback van patiënten. Patiënten vonden de VCE nuttig en waren bereid om extra tijd en moeite te besteden om informatie te ontvangen over hun voorkeuren. Echter, patiënten ondervonden meerdere praktische problemen. Patiënten vonden het lastig om alle verschillende behandelingskenmerken te herinneren, of vonden het lastig om een beste kenmerk te kiezen uit een lijst met negatief geformuleerde kenmerken. Sommige patiënten hadden meer uitleg nodig bij het maken van de keuzevragen. Naast deze praktische problemen (die kunnen worden ondervangen), waren er verscheidene methodologische problemen in de ontwerpfase van deze VCEs. Over het algemeen, hebben gebalanceerde en orthogonale experimentele designs de voorkeur, maar deze zijn vaak niet beschikbaar of vereisen een onmogelijk aantal keuzevragen. BWS laat patiënten expliciet en actief nadenken over de relevante afwegingen tussen behandelingskenmerken, maar als patiënten gedwongen worden om snel en niet doordachte keuze te maken door de vele keuzetaken, dan wordt de doelstelling van BWS niet bereikt. In de praktijk zal statistische of design efficiëntie (hoe goed de kenmerken zijn gebalanceerd en orthogonaal) moeten worden afgewogen tegen response-efficiëntie (wat is cognitief haalbaar voor patiënten) en zal het gebruik van niet-optimale experimentele designs geen uitzondering zijn. Eén van de grootste uitdagingen is dat het tot op heden onbekend en nog niet onderzocht is hoe het gebruik van niet-optimale designs de betrouwbaarheid van individuele preferenties beïnvloed.

CONCLUSIE

Dit proefschrift toont duidelijk aan dat “de waarde die samenhangt met het proces en gemak van zorg” een belangrijke driver is van preferenties. Dit geldt voor de patiënten, maar ook voor het algemene publiek. Het gebruik van proces-utiliteit in de klinische praktijk is heel gebruikelijk, daarentegen heeft het gebruik van proces-utiliteit in vergoedingsbeslissingen of andere maatschappelijke beslissingen nog een lange weg te gaan. In tegenstelling tot maatschappelijke beslissingen, zijn sterke theoretisch onderbouwde methodes om voorkeuren te meten niet gebruikelijk in de klinische praktijk. Dit proefschrift toont dat BWS case 2 potentie heeft om gebruikt te worden in gedeelde besluitvorming, maar dat vooralsnog methodologische issues moeten worden onderzocht voor het gebruik van BWS in een VCE kan worden aanbevolen.

Dankwoord

Dankwoord

Met enorm veel plezier heb ik de afgelopen jaren gewerkt aan dit proefschrift. Het was een stuk minder plezierig geweest zonder de hulp en afleiding die talloze mensen hebben geboden. Nu is het tijd om jullie te bedanken!

Allereerst wil ik Janine van Til bedanken voor haar begeleiding tijdens het promotietraject. Janine, jij verdient eigenlijk een persoonlijk hoofdstuk als dankwoord. In onze gesprekken heb je mij altijd gemotiveerd en het vertrouwen gegeven dat we er samen iets moois van konden maken. Je bent een zeer betrokken en bevlogen begeleidster en ik kan nog veel van je leren. Daarom ben ik blij dat onze samenwerking niet stopt en wie weet kunnen we in de toekomst samen nog een andere stad dan New York onveilig gaan maken 😊.

Dit proefschrift was er ook niet geweest zonder mijn eerste promotor: Maarten IJzerman. In jou hebben alle promovendi een geweldig voorbeeld van een toegewijde wetenschapper. Je gaat grondig te werk, blijft altijd nieuwsgierig en staat altijd open voor nieuwe projecten en samenwerkingen. Je uitgebreide internationale contacten maakten dat het mogelijk was om de samenwerking buiten de landsgrenzen te zoeken. Bedankt voor alle commentaren, adviezen en kritische vragen, maar bovenal voor je interesse en vertrouwen.

Naast mijn promotoren wil ik graag de leden van mijn promotiecommissie bedanken voor het beoordelen van mijn proefschrift. Bedankt dat u de tijd hebt genomen om mijn proefschrift door te nemen en aanwezig te zijn bij mijn verdediging.

Uiteraard wil ik mijn paranimfen Michelle Kip en Marloes Uiteendewilligen bedanken voor hun hulp en steun bij de promotie zelf. Marloes, 4 jaar geleden mocht ik als bruidsmisje naast jou staan, een hele ervaring! Ik hoop dat jij net zo geniet van het paranimf zijn! Michelle, al die jaren ben je al onafgebroken mijn kamergenoot (die paar maanden niet zullen we maar voor lief nemen). Wat fijn dat jij uit hetzelfde hout gesneden bent, en ook die herkenbare Twentse nuchterheid bezit. Je hebt altijd een gezonde werklust, maar er is (gelukkig) ook tijd genoeg voor de broodnodige afleiding! Je bent een TOP collega!

Een belangrijke coauteur van dit proefschrift die ik ook speciaal wil bedanken is Karin Groothuis. Zonder jou had ik STATA nog steeds niet begrepen. Fijn dat je al tijdens mijn afstuderen de tijd nam om me wegwijs te maken in het programma. Ik heb erg veel respect voor je statistische kennis en zal zeker nog eens een beroep op je doen.

Sarah Janus, het eerste artikel heb ik samen met jou mogen schrijven. Een review schrijven dat was nog niet zo 123 gepiept, en eigenlijk was het een hels karwei. Dan is het toch fijn dat je van elkaar kunt leren en kunt samenwerken. Daarnaast hebben we nog leuke werktripjes naar Berlijn mogen maken die een groot succes waren. Ik wens je veel succes met jou verdediging binnenkort en veel geluk in Groningen.

Voor een deel van de artikelen in dit proefschrift was ik aan een projectgroep verbonden waar artsen en patiënten deel van uitmaakten (het Parkinson Project). Dat vond ik erg bijzonder en leuk om mee te mogen werken! Allen bedankt voor jullie interesse. Kris Movig en Jeroen van Vugt, bedankt voor jullie input op de artikelen, ook al duurt wachten soms lang, uiteindelijk heeft het tot een mooie publicatie geleid.

Holly Witteman and Liana Fraenkel, thank you very much for the opportunity to write an article with you. Through your input and critical questions you have shaped my thinking on using conjoint analysis in clinical decision making.

In het laatste jaar van mijn promotietraject heb ik samengewerkt met urologen in het Elisabeth-TweeSteden Ziekenhuis om patiënten met prostaatkanker te helpen bij hun keuze voor behandeling. Romy Lamers en Paul Kil, jullie enthousiasme werkte aanstekelijk en het is erg leuk om te zien dat jullie zo bezig zijn met de wensen en voorkeuren van de patiënt. Bedankt dat jullie ons de kans hebben gegeven om uit te zoeken of de dingen die wij bedenken achter het bureau ook in de praktijk werken. Stapje voor stapje ontdekken we wat wel werkt en ik ben blij dat onze samenwerking (ook samen met Isabel de Angst) nog even voortduurt.

Bij het doen van mijn onderzoek heb ik hulp gehad van een aantal studenten: Evelien Welhuis, Tina Leung, Keli Chen, Leonie Hulshof en Marije van der Straten. Bedankt voor jullie enthousiasme en interessante bijdrages die ook onderdeel zijn geworden van dit proefschrift.

De meeste tijd van mijn promotie heb ik doorgebracht binnen de vakgroep Health Technology and Services Research. In de afgelopen jaren heb ik daar aardig wat kamers versleten. Ooit begonnen op de 'studentenkamer', toen op naar 5260 (de 'echte' PhD kamer) en nu naar de andere kant van de gang. Het maakt gelukkig niet uit waar je zit, je ontkomt nergens aan de leuke werksfeer: de leuke discussies tijdens de lunch (of het nu wel of niet werk gerelateerd is), de lekkere zelfgemaakte baksels en de tegenwoordig uitbundig gevierde verjaardagen met bijbehorende versiersels en rijmpjes. Alle (oud) collega's en promovendi bedankt voor alle goede adviezen en leuke afleidingen. Gelukkig hoefde ik de promotie-achtbaan niet alleen te doorstaan ☺. Ik wil Magda

Dankwoord

Boere-Boonekamp en Renske van Wijk nog speciaal bedanken: zonder jullie enthousiasme en goede begeleiding tijdens mijn bacheloropdracht had ik nooit 'de weg' naar het promotieonderzoek gevonden. Bedankt voor jullie hulp bij het publiceren van mijn bacheloronderzoek.

Niet te vergeten: onderzoek komt nooit van de grond zonder goede ondersteuning. Joke, Tanja, Maaïke en Annelies bedankt voor jullie hulp en nog veel belangrijker jullie luisterend oor!

Minder op wetenschappelijk gebied, maar des te meer op sociaal gebied wil ik graag mijn vrienden bedanken voor alle leuke dingen die we doen. Paintballen, escaperooms, 30 seconds, sabotageuren of leuk naar de kroeg of een tentfeest. Om op je werk te kunnen presteren moet je daarbuiten ook gelukkig zijn en jullie hebben daar allen aan bijgedragen.

Mijn lieve familie verdient natuurlijk ook een plaats in het dankwoord. Dank dat jullie altijd voor mij klaarstaan, voor de mooie jeugd in Beckum, de mogelijkheid om te studeren, en jullie onvoorwaardelijke steun. Ik ben blij met alle krantenknipsels die jullie uitknippen over Parkinson. Lieve schoonfamilie (sinds kort mag ik jullie ook echt zo noemen 😊) bedankt voor de interesse in mijn onderzoek en de leuke familietripjes.

Lieve Remco, jouw onvoorwaardelijke steun en liefde hebben ervoor gezorgd dat ik trots ben op dit proefschrift. Het jaar 2017 is voor ons allebei een aaneenschakeling van mijlpalen en veranderingen, maar eigenlijk geniet ik het meest van alle momenten die we samen zijn. Elke dag met jou is een feest!

Ik voel me gezegend met zo veel mooie mensen om me heen. Ik kan terugkijken op een leerzame en bijzondere tijd en kijk graag vooruit naar de toekomst!

Veel liefs

Marieke

Curriculum Vitae and Publications

CURRICULUM VITAE

Marieke Weernink werd geboren op 16 September 1988 te Beckum (Hengelo). In 2008 behaalde zij haar VWO-diploma aan het Bataafse Kamp Lyceum in Hengelo. In datzelfde jaar ging zij Gezondheidswetenschappen studeren aan de Universiteit Twente. Tijdens haar bacheloropdracht heeft zij de relatie onderzocht tussen de vitamine D inname van aanstaande moeders en het ontstaan van een schedeldeformatie bij zuigelingen van 2-4 maand oud. De resultaten van dit onderzoek zijn gepubliceerd in het tijdschrift *Maternal & Child Nutrition*. De bachelor Gezondheidswetenschappen werd gevolgd door de master Health Sciences aan de Universiteit Twente. In 2012 sloot ze deze succesvol af met een masterthesis over de rol van eindgebruikers in het ontwerp en ontwikkelingsproces van medical devices. Dit proces werd geïllustreerd aan de hand van de ontwikkeling van een point of care test om de nierfunctie te meten. Na haar studie, begon zij bij de vakgroep Health Technology and Services Research van de Universiteit Twente aan het promotietraject dat leidde tot dit proefschrift. In dit proefschrift onderzocht ze de impact van het proces van zorg op de waarde van behandelingen voor de ziekte van Parkinson. Daarnaast onderzocht ze (in samenwerking met het Elisabeth-TweeSteden Ziekenhuis) hoe individuele patiënten het beste geholpen kunnen worden om hun voorkeuren te achterhalen. Vanaf oktober 2016 is Marieke werkzaam als post-doc onderzoeker bij de vakgroep HTSR. Eén van haar projecten gaat over de implementatie van elektro-encefalogram (EEG) metingen om beter te voorspellen wat de uitkomst van een post-anoxisch coma zal zijn. Eén van de vraagstukken is bijvoorbeeld hoe mensen de kwaliteit van leven na een post-anoxisch coma waarderen.

LIST OF PUBLICATIONS

Journal papers

Weernink, M. G. M., Janus, S. L. M., van Til, J. A., Raisch, D. W., van Manen, J. G., & IJzerman, M. J. (2014). A Systematic Review to Identify the Use of Preference Elicitation Methods in Healthcare Decision Making. *Pharmaceutical Medicine*, 28(4), 175-185. doi:10.1007/s40290-014-0059-1

Weernink, M. G. M., van Til, J. A., van Vugt, J. P. P., Movig, K. L. L., Groothuis-Oudshoorn, C. G. M., & IJzerman, M. J. (2016). Involving Patients in Weighting Benefits and Harms of Treatment in Parkinson's Disease. *PLoS One*, 11(8), e0160771. doi:10.1371/journal.pone.0160771

Weernink, M. G. M., Groothuis-Oudshoorn, C. G. M., IJzerman, M. J., & van Til, J. A. (2016). Valuing Treatments for Parkinson Disease Incorporating Process Utility: Performance of Best-Worst Scaling, Time Trade-Off, and Visual Analogue Scales. *Value in Health*. doi:10.1016/j.jval.2015.11.011

Weernink, M. G. M., van Wijk, R. M., Groothuis-Oudshoorn, C. G. M., Lanting, C. I., Grant, C. C., van Vlimmeren, L. A., & Boere-Boonekamp, M. M. (2016). Insufficient vitamin D supplement use during pregnancy and early childhood: A risk factor for positional skull deformation. *Maternal and Child Nutrition*, 12(1), 177-188. doi:10.1111/mcn.12153

Conference contributions

Janus, S.I.M. and Weernink, M.G.M. and Til, J.A. van and Raisch, D.W. and Manen, J.G. van and IJzerman, M.J. (2014) A Systematic Review To Identify the Use of Preference Elicitation Methods in Health Care Decision Making. Poster presentation. In: ISPOR 17th Annual European Congress, Amsterdam, 10-12 November 2014, Amsterdam.

Weernink, Marieke G.M. and Groothuis-Oudshoorn, Karin G.M. and IJzerman, Maarten J. and Til, Janine A. van (2014) Comparison of treatment alternatives in Parkinson's disease with best-worst scaling, time trade-off and visual analogue scales. Poster presentation In: ISPOR 17th Annual European Congress, Amsterdam, 10-12 November 2014, Amsterdam.

Weernink, M. G. M., Groothuis-Oudshoorn, C. G. M., IJzerman, M. J., & van Til, J. A. (2016). Comparison of the valuation of treatment alternatives in Parkinson's disease with best-worst scaling, time trade-off and visual analogue scales. Presentation. In: Seminar Choice modeling for healthcare:

List of publications

methodological issues in best-worst scaling and discrete choice experiments, August 14, 2014, Enschede.

Weernink, M. G. M., Groothuis-Oudshoorn, C. G. M., van Til, J. A. & IJzerman, M. J. Framing of attribute's levels: influence on the valuation and interpretation of outcomes from a best-worst scaling experiment. Presentation. In: 3rd meeting of the International Academy Health Preference Research, 18 October 2016, Saint Louis.

Weernink, M. G. M., Groothuis-Oudshoorn, C. G. M., van Til, J. A. & IJzerman, M. J. Framing of attribute's levels: influence on the valuation and interpretation of outcomes from a best-worst scaling experiment. Poster Presentation. In: 37th Annual North American Meeting of the Society for Medical Decision Making, 18-21 October 2015, Saint Louis. Nominated for the Lee B. Lusted Student Prize.

Weernink, M. G. M., Groothuis-Oudshoorn, C. G. M., IJzerman, M. J., & van Til, J. A.. Risky decision making in Best-Worst Scaling. Presentation. In: Seminar Using Stated Preference methods in Health and Public Health, 9-10 April 2015, Bilthoven.

Weernink, M. G. M., Groothuis-Oudshoorn, C. G. M., van Til, J. A. & IJzerman, M. J. Framing of attribute's levels: influence on the valuation and interpretation of outcomes from a best-worst scaling experiment. Poster Presentation. In: 16th Biennial European Conference of the Society for Medical Decision Making, 12-14 June 2016, Londen.