Discrete Choice Experiments in Health Care

Theory and Applications

Esther W. de Bekker-Grob

ISBN 978-90-9024498-3

Cover image © iStockphoto.com / gibsonff
Cover design: Anna Bosselaar (anna@zoiets.com)
Lay-out: Esther W. de Bekker-Grob
Printed by Ipskamp Drukkers (www.proefschriften.net)

This thesis was printed with financial support of the J.E. Jurriaanse Stichting, the Department of Public Health Erasmus MC, and Erasmus University Rotterdam.

Discrete Choice Experiments in Health Care: Theory and Applications

Discrete Keuze Experimenten in de Gezondheidszorg: Theorie en Toepassingen

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus
Prof.dr. H.G. Schmidt
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
woensdag 25 november 2009 om 13.30 uur

door

Esther Wilhelmina de Bekker-Grob

geboren te Zevenaar

Promotiecommissie

Promotor:	Prof.dr. E.W. Steyerberg
Overige leden:	Prof.dr. W.B.F. Brouwer
	Prof.dr. M.G.M. Hunink
	Prof.dr. E.J. Kuipers
Co-promotor:	Dr. M.L. Essink-Bot

For those who matter

Contents

1

General introduction

1.1 Measuring benefits in the delivery of health care

Health economics is concerned with issues related to scarcity in the allocation of health care. The basic tasks of any economic evaluation are to identify, measure, value, and compare the costs and benefits of alternatives being considered. Traditional means of measuring benefits in the delivery of health care have concentrated on improvements in health outcomes using clinical outcomes and Quality Adjusted Life Years (QALY). The QALY is a measure of the quantity of life gained weighted by the quality of that life [1]. QALYs are extensively used in economic analyses in health care. They claim to capture the health outcome benefits caused by an intervention [2-4]. However, benefits of a health care intervention or service can be many-sided, e.g. containing non-health outcomes (e.g. amount of information) and process characteristics (e.g. treatment location, route of drug administration, patient experienced burden of testing). For instance, is 'reduction of dying from cervical cancer' the only screening characteristic that is considered by women attending a cervical cancer screening programme? Evidence shows that, within the context of cervical cancer screening, women's preferences for various programmes are also determined by other characteristics than the reduced chance of dying from cervical cancer [5]. Individuals are willing to trade changes in health outcome (change in chance of dying from cervical cancer) with process characteristics (time between smears, time for results, chance of being recalled, chance of abnormality, cost of each smear). This is just one example that illustrats that utility (benefit, satisfaction) of an intervention is derived from both health outcomes and process- and non-health outcomes. Other studies showed that this result is not specific to cervical cancer screening [6-12]. This suggests that, assuming the goal of health interventions or services is to maximise utility, the value of process attributes and non-health outcomes should be considered alongside health outcomes [13]. These might be relevant for individuals' preferences and acceptability for specific health care interventions (i.e. demand-led health care), and for some interventions that do not provide reduction in morbidity or mortality (e.g. cosmetic surgery).

1.2 Discrete choice experiments

The discrete choice experiment (DCE) approach provides opportunities for evaluation of process effects and non-health outcomes additional to traditional QALY analysis. A DCE is a technique for investigating individual preferences. The technique of DCE has its origin in mathematical psychology. The DCE method has been employed by companies to investigate the relative importance of the characteristics of their products influencing consumers' demand [2]. The results are potentially useful to define optimal strategies for improvement of the products and hence to maximize sales. Market researchers first used this technique [14]. Nowadays DCEs have also been used widely in transport economics [15] and environmental economics [16].

In a DCE individuals are offered a series of choice sets. They are asked to choose in each choice set between two or more alternatives. See Table 1.1 for an example of a choice set

Table 1.1 Example of a DCE choice set concerning alternative modes of transport to work.

Attributes	**Car**	**Train**
Duration of transport	20 min	25 min
Walk to/from transportation	0 min	3 min
Price	€ 10	€ 4
My choice	☐	☐

Attribute Alternative Attribute level

concerning alternative modes of transport to work. The technique of DCE is an attribute-based measure of benefit, based on the assumptions that, first, alternatives can be described by their characteristics, known as attributes, and second, an individual's valuation (i.e., benefit, utility, satisfaction or preference) depends upon the levels of these attributes (see Table 1.1). Responses of a DCE are modelled within a benefit (or satisfaction) function which provides information on whether or not the given attributes are important; the relative importance of attributes; the rate at which individuals are willing to trade between attributes; and overall benefit scores for alternatives [2, 17, 18].

The technique of DCE was introduced in health care in the early 1990s [2]. It provides opportunities for evaluation whether a given health, non-health or process attribute of a health care intervention or service is important; the relative importance of these various attributes; and the trade-offs individuals are made between these attributes. In comparison to other stated preferences techniques (e.g. willingness to pay methods to assess the monetary value of alternatives, conjoint analysis methods and the visual analogue scale to rank, rate or scale alternatives, and standard gamble and time trade-off methods to assess the risk-benefit trade-off of alternatives), a DCE presents a reasonably straightforward task and one which more closely resembles a real world decision (i.e., trading off health outcomes, process attributes and/or non-health outcomes) [19].

1.3 Theoretical aspects of discrete choice experiments

The DCE approach combines random utility theory with consumer theory, experimental design theory and econometric analysis. There are key developments taking place in other areas of economics outside health economics, particularly concerning experimental design and methods of analysis [20]. Before introducing some of these technical aspects, a theoretical background of DCE is given first.

1.3.1 Theoretical background

In a DCE in health care individuals are offered a series of choices between two or more health care interventions, services or policies that have different combinations of attribute levels. (See Table 1.2 for an example of a choice set in health care, taken from a DCE study to evaluate preferences for the provision of benign prostatic hyperplasia drug treatment [10]). Analysing the responses allows for evaluation of the relative importance of the attributes for respondents' preferences, and for evaluation of the trade-offs that individuals make between the attributes. The latter are reflected by the marginal rates of substitution (MRS) between these attributes. Thus, the MRS of attribute Y for attribute X is the amount of attribute Y that an individual is willing to exchange for attribute X. For example, looking at Table 1.2, the MRS between the attributes 'sexual side effects from treatment' and 'time to improvement in symptoms' reflects how much longer individuals say to be willing to wait for symptom improvement (in months) to avoid sexual side effects. If cost is included as an attribute the willingness to pay (WTP) can be estimated as well [2]. For example, it is possible to evaluate how much individuals express to be willing to pay (in British Pounds) to avoid sexual side effects based on responses to the choice set presented in Table 1.2.

The discrete choices observed in a DCE are assumed to reveal an underlying (latent) utility function. An individual acting rationally is expected to evaluate the available alternatives in each choice set and to choose the alternative which gives the greatest

Table 1.2 Example of a DCE choice set in health care regarding preferences for the provision of benign prostatic hyperplasia drug treatment (source: [10])

Characteristics of treatment	**Treatment A**	**Treatment B**	**No treatment**
Time to improvement in symptoms	3 months	1 month	No improvement
Treatment reduces the size of prostate	No	Yes	No
Sexual side effects from treatment	Decreased sexual desire	None	None
Non-sexual side effects from treatment	Headache	Dizziness	None
Cost of treatment per month	£30	£50	£0
Chance of having Acute Urinary Retention after 2 years	2%	2%	4%
Chance of requiring surgery after 2 years	4%	6%	4%
Which drug treatment would you prefer?	☐	☐	☐

relative utility, by making trade-offs across the different health outcomes, non-health outcomes and/or process attributes. Thus, an individual will choose alternative A over B, if U (X_A, Z) > U (X_B, Z), where U represents the individual's indirect utility function from certain alternatives, X_A the attributes of alternative A, X_B the attributes of alternative B, and Z socioeconomic characteristics of the individual that influence his/her utility.

Choices observed in DCEs are analysed using *random* utility theory (i.e. an error term is included in the utility function to reflect the unobservable factors in the individual's utility function). Thus, an individual will choose alternative A over B, if V (X_A, Z) + ε_A > V (X_B, Z) + ε_B, where V is the measurable component of utility estimated empirically, and ε_A and ε_B reflect the unobservable factors in the individual's utility function of alternative A and B respectively (X_A, X_B and Z defined as above).

For further details on how to conduct a DCE and detailed explanations of theoretical issues see Louviere et al. [17]; Hensher et al. [21]; Bliemer and Rose [22]; and Ryan et al. [23].

1.3.2 Methodological advances

Since DCEs were introduced in health economics in the early 1990s, they have become increasingly popular in health care settings. This is shown by the increased growth of published DCEs in a various range of journals in recent years [7, 9, 10, 24-29]. However, the most recent overview of published DCEs dates from 2003 [20] and an overview of more recent applications of DCE in health economics is lacking (i.e. which issues are addressed?). Also it is not known to what extent technical practice of DCE in health economics are following developments taking place in other disciplines or areas of economics. Taking stock of the way in which DCE studies build up in the literature is important.

1.3.3 Labelled or unlabelled DCE

A specific question that arises in the application of DCE is whether to present the choice sets in a labelled or unlabelled form. The unlabelled form involves assigning unlabelled alternatives in the choice set, such as 'alternative A', 'alternative B' and so on (see Table 1.2 for example). The labelled form involves assigning labels that communicate information regarding the alternative (see Table 1.1. for example). In marketing applications, labels tend to consist of brand names and logos, which consumers have learnt to associate with different product characteristics and feelings. In the context of health economics, labels tend to consist of generic or brand-name medications, specific screening tests (e.g. colonoscopy, sigmoidoscopy), specific treatments (surgery vs conservative), or other descriptors. An advantage of assigning labels is that alternatives will be more realistic and the choice task less abstract, which may add to the validity of the results (i.e., responses reflect better the real preference structure) and hence, the

results may be better suitable to support decision-making at the policy level. However, by far most commonly applied DCEs in health economics used unlabelled alternatives. In health economics, the differences between labelled and unlabelled DCEs in various aspects of feasibility, trading-behaviour, and convergent validity have not been empirically investigated.

1.4 Objectives of the thesis

The application of DCE to the field of health care seems promising. This thesis addresses theoretical aspects and applications of DCE in health care. Its mains objectives are:

1. To provide an overview of current DCE practice in health economics, and to compare and assess progress in methodological advances.
2. To study the usefulness of DCE for real choice problems in health care for measuring benefits beyond health outcomes.
3. To compare labelled versus unlabelled DCEs in health care in various aspects of feasibility, trading-behaviour, and convergent validity.

1.5 Structure of the thesis

This thesis consists of three parts. Part one (**Chapter 2**) provides the reader an overview of current DCE practice in health economics. These current DCEs (2001-2008) were systematically reviewed and compared with previous DCE practice in health care (1990-2000). Extracted data concentrate on key issues concerning experimental design and choice set construction, estimation procedures, and validity. In doing so this chapter assesses progress in these key areas, and identifies important lessons learnt and to be learnt.

Part two of this thesis (chapters 3-8) describes four applications of DCEs developed for real choice problems in health care to measure benefits beyond health outcomes. These four problems illustrate the potential of DCE to include non-health outcomes and process effects of health care interventions in preference elicitation.

The first DCE focuses on preferences for preventive osteoporosis drug treatment. Various practice guidelines recommend a case-finding approach to identify persons with a high risk of osteoporotic fractures. However, the usefulness of this approach depends on whether the identified persons are willing to take preventive osteoporosis drug treatment, and on what conditions. The DCE was therefore a relevant addition to QALY analysis, because it could include process attributes of drug treatment to elicit the relative weights that patients place on various aspects of preventive drug treatment for osteoporosis (**Chapter 3**), the relative weights that general practitioners place on these aspects and to evaluate the determinants of the preference differences found between patients and GPs (**Chapter 4**).

The second DCE focuses on preferences for brace treatment for idiopathic scoliosis patients. There are two treatments for idiopathic scoliosis patients, namely brace treatment and surgery. Brace treatment can be burdensome, whereas its effectiveness has been unproven so far. This raised the question how effective brace treatment should be for idiopathic scoliosis patients to consider a brace as a reasonable form of treatment. The DCE was therefore a relevant addition to QALY analysis, because it could take into account the 'burden of treatment' (i.e. wearing the brace; and thus process attributes) in the preferences (**Chapter 5**).

The third DCE concerns breast reconstruction modalities. Multiple techniques are available for breast reconstruction. Each technique has its own (dis)advantages. Complementary to the medical analysis performed by the plastic surgeon regarding which breast reconstruction type individual patients could undergo, patients' preferences for the procedure they would opt for are also important determinants. To get insight into patients' benefits for breast reconstruction modalities, QALY-measurement is inappropriate. Namely, all benefits of breast reconstruction modalities are 'non-health' (such as cosmetic result of breast reconstruction) or process (burden of treatment (e.g. more than one surgery), and use of autologous tissue). Therefore the DCE instead of QALY analysis allowed to get insight into benefits for breast reconstruction modalities (**Chapter 6**).

And finally, a fourth DCE concerns on estimating public's preferences for colorectal cancer screening. Several European countries including the Netherlands are currently considering a nation-wide screening programme. The programme will probably be based on a standardized protocol with a single screening test. The uptake of a specific screening test will ultimately determine the population health benefit of a CRC screening programme. A DCE was undertaken to get insight in the determinants of uptake for various screening tests. Again, the DCE was a relevant addition to QALY analysis, because it gave the possibility to take process attributes of colorectal cancer screening tests (e.g., screening interval, preparation, location of screening) into account (**Chapters 7 and 8**).

Part three (**Chapter 9**) describes an empirical comparison between a labelled and an unlabelled DCE in various aspects of feasibility, trading-behaviour, and convergent validity by investigating individuals' preferences for colorectal screening programmes.

This thesis ends with a discussion of the theoretical and practical results and presents a number of conclusions and recommendations (**Chapter 10**).

References

1. National Institute for Health and Clinical Excellence – measuring effects and cost effectiveness: the QALY. http://www.nice.org.uk/newsevents/infocus/infocusarchive/measuringeffectivenessandcosteffectivenesstheqaly.jsp. Retrieved on 14 January 2009.

2. Ryan M, Hughes M (1997). Using conjoint analysis to assess women's preferences for miscarriage management. Health Econ 6(3):261-273.
3. Ryan M (1999). Using conjoint analysis to go beyond health outcomes and take account of patient preferences: an application to in vitro fertilisation. Soc Sci Med 8:535-546.
4. Ryan M, Shackley P (1995). Assessing the benefits of health care: how far should we go? Qual Health Care 4(3):207-213.
5. Wordsworth S, Ryan M (2006). Women's preferences for cervical cancer screening: A study using a discrete choice experiment. Int J Technol Assess Health Care 22(3):344-50.
6. Shackley P, Slack R, Michaels J (2001). Vascular patients' preferences for local treatment: an application of conjoint analysis. J Health Serv Res Policy 6(3):151-157.
7. Ratcliffe J, Van Haselen R, Buxton M, Hardy K, Colehan J, Partridge M (2002). Assessing patients' preferences for characteristics associated with homeopathic and conventional treatment of asthma: a conjoint analysis study. Thorax 57(6):503-508.
8. Hall J, Kenny P, King M, Louviere J, Viney R, Yeoh A (2002). Using stated preference discrete choice modelling to evaluate the introduction of varicella vaccination. Health Econ 11(5):457-465.
9. Sculpher M, Bryan S, Fry P, de Winter P, Payne H, Emberton M (2004). Patients' preferences for the management of non-metastatic prostate cancer: discrete choice experiment. BMJ 328(7436):382.
10. Watson V, Ryan M, Brown CT, Barnett G, Ellis BW, Emberton M (2004). Eliciting preferences for drug treatment of lower urinary tract symptoms associated with benign prostatic hyperplasia. J Urol 172(6):2321-2325.
11. Lloyd A, McIntosh E, Price M (2005). The importance of drug adverse effects compared with seizure control for people with epilepsy: a discrete choice experiment. Pharmacoeconomics 23(11):1167-1181.
12. Lancsar EJ, Hall JP, King M, Kenny P, Louviere JJ, Fiebig DG, Hossain I, Thien FC, Reddel HK, Jenkins CR (2007). Using discrete choice experiments to investigate subject preferences for preventive asthma medication. Respirology 12(1):127-136.
13. Watson V, Ryan M, Watson E (2009). Valuing Experience Factors in the Provision of Chlamydia Screening: An Application to Women Attending the Family Planning Clinic. Value in Health, *in press*.
14. Cattin P, Wittink D (1982). Commercial use of conjoint analysis: a survey. J Marketing 46 (3): 44-53.
15. Hensher DA (1997). Stated preference analysis of travel choices: the state of practice. In: Transport Economics: selected readings, Oum TH, Dodgson JS, Hensher DA et al. (eds.). Amsterdam, Harwood Academic in cooperation with the Korea Research Foundation for the 21st century, 81-109.
16. Adamowicz V, Boxall P (2001). Future directions of stated choice methods for environment valuation. Paper presented at the conference Choice Experiments: A new approach to environmental valuation, London.
17. Louviere JJ, Hensher DA, Swait JD (2000). Stated choice methods: analysis and application. Cambridge University Press, Cambridge
18. Ryan M, Farrar S (2000). Using conjoint analysis to elicit preferences for health care. BMJ 320 (7248):1530–1533
19. Mangham LJ, Hanson K, McPake B (2009). How to do (or not to do) ... Designing a discrete choice experiment for application in a low-income country. Health Policy Plan, *in press*.
20. Ryan M, Gerard K (2003). Using discrete choice experiments to value health care programmes: current practice and future research reflections. Appl Health Econ Health Policy 2(1):55-64.
21. Hensher DA, Greene W (2003). The mixed logit model: the state of practice. Transportation 30: 133-176.
22. Bliemer MCJ, Rose JM (2006). Designing Stated Choice Experiments: State-of-the-art. Paper presented at the 11th International Conference on Travel Behaviour Research.
23. Ryan M, Gerard K, Amaya-Amaya M (2008). Using Discrete Choice Experiments to Value Health and Health Care. (Vol. 11). Dordrecht, The Netherlands: Springer.
24. Ashcroft DM, Seston E, Griffiths CE (2006). Trade-offs between the benefits and risks of drug treatment for psoriasis: a discrete choice experiment with U.K. dermatologists. *Br J Dermatol 155*(6), 1236-1241.

25. Mantovani LG, Monzini MS, Mannucci PM, Scalone L, Villa M, Gringeri A; The Conan Study Group (2005). Differences between patients', physicians' and pharmacists' preferences for treatment products in haemophilia: a discrete choice experiment. Haemophilia 11(6):589-597.

26. Peacock S, Apicella C, Andrews L, Tucker K, Bankier A, Daly MB, Hopper JL (2006). A discrete choice experiment of preferences for genetic counselling among Jewish women seeking cancer genetics services. Br J Cancer, 95(10):1448-1453.

27. Ratcliffe J, Buxton M, McGarry T, Sheldon R, Chancellor J (2004). Patients' preferences for characteristics associated with treatments for osteoarthritis. Rheumatology (Oxford) 43(3):337-345.

28. Roux L, Ubach C, Donaldson C, Ryan M (2004). Valuing the benefits of weight loss programs: an application of the discrete choice experiment. Obes Res 12(8):1342-1351.

29. Salkeld G, Solomon M, Butow P, Short L (2005). Discrete-choice experiment to measure patient preferences for the surgical management of colorectal cancer. Br J Surg 92(6):742-747.

Part I

Review of discrete choice experiments in health care

2

Applying discrete choice experiments to value health and health care: a review of the literature

de Bekker-Grob EW, Ryan M, Gerard K, Steyerberg EW, Essink-Bot ML

Submitted

Summary

Discrete choice experiments (DCEs) have become a commonly used instrument in health economics research. This paper updates a review of published papers between 1990-2000 for the years 2001-2008. Based on this previous review, and a number of other key review papers, consideration is given to 3 issues: experimental design; estimation procedures; and validity of responses. We identified 114 DCEs, covering a range of policy questions much broader than valuing patient experience factors. These included valuing health outcomes, trade-offs between health outcomes and patient experience factors, estimating utility weights within the Quality Adjusted Life Year framework, understanding labour-market choices, developing priority setting frameworks, and doctor's preferences for patient's treatment. Recently published DCEs in health economics made more use of foldover methods and the D-efficient criterion for optimal design. We also note an increase in the use of models with greater flexibility, such as nested logit, latent class logit, and mixed logit. There has been a shift towards statistically more efficient designs and richer econometric models. However, much progress has still to be made towards assessment of external validity, and incorporation of DCE results into a decision-making framework by policy makers.

2.1 Introduction

Discrete choice experiments (DCEs) have become a commonly used technique in health economic research, addressing a wide range of important policy questions [1]. The technique is an attribute-based measure of benefit, based on the assumptions that, health care interventions, services or policies, can be described by their attributes and, second, an individual's valuation depends upon the levels of these attributes. Within a DCE respondents are asked to choose between two or more alternatives. The resulting choices reveal an underlying (latent) utility function. The approach combines random utility theory with consumer theory, experimental design theory and econometric analysis. Details on conducting a DCE and theoretical issues are covered elsewhere [1-5].

Ryan and Gerard [6] identified a number of key methodological issues in their review of the application of DCEs in health between 1990 and 2000. These were echoed by others in the same field [1, 5, 7-10]. Recently, Guttman et al. [11] updated the initial work of Ryan and Gerard [6] by counting the key characteristics of DCEs used. Lancsar and Louviere [12] developed a checklist of issues to consider when developing or reviewing the quality of a DCE. Finally, Louviere and Lancsar [13] surveyed the current state-of-the-art, discussed key issues needing further research and emerging research trends, and suggested ways to move various aspects of DCEs towards best practice. Although these papers have made very useful contributions to the health economics literature, an extended systematic review on current DCE practice is still lacking. We aimed to review current DCEs in health economics, with an in depth focus on the issues 'experimental design'; 'methods of analysis'; and 'validity' [6]. We report on the methods and general results in Section 2, and compare the status of the three key issues between 1990-2000 and 2001-2008 in Section 3. Section 4 provides a discussion, suggestions for future research and some concluding remarks.

2.2 Review of DCEs between 2001 and 2008

2.2.1 Literature search

We conducted a systematic review of current published DCE health economics studies, using Medline to identify English language DCE studies available in print or online between 2001 and 2008. Although this was a narrower search of databases than the baseline study it was expected to identify the large majority of the health related DCE studies published during the period. We used the same search terms as Ryan and Gerard [6] (These were: "discrete choice experiment(s)", "discrete choice modelling", "discrete choice modeling", "stated preference", "part-worth utilities", "functional measurement", "paired comparisons", "pairwise choices", "conjoint analysis", "conjoint measurement", "conjoint studies", and "conjoint choice experiments"), leading to inclusion of studies if choice-based (studies that used Best Worst Scaling were not included (see Flynn et al. [14], for further details), published as a full text article, and applied to a health care topic. Two

of the authors (EWBG & MLEB) extracted data on three methodological key issues: 'experimental design'; 'methods of analysis'; and 'validity'.

2.2.2 Background results

The search generated 682 possible references. After reading abstracts or full articles, 121 references relating to 114 original studies met the inclusion criteria. The Appendix shows the full list of references (2009 references are the published versions of the papers available online at the time of the search between 01-01-2001 and 31-12-2008), and

Table 2.1 Discription of DCE studies

Item	Category	Baseline: 1990-2000		Current: 2001-2008	
		N=34	*(%)*	*N=114*	*(%)*
Country of origin*	UK	20	(59)	55	(48)
	US	7	(21)	14	(12)
	Australia	6	(18)	13	(11)
	Canada	1	(3)	6	(5)
	Denmark	0	(-)	5	(4)
	Netherlands	0	(-)	5	(4)
	Germany	0	(-)	3	(3)
	Other (Switzerland, China, Italy, France, Russia, Japan, Sweden, Spain, Zambia, Ghana, Nepal, Malawi)	0	(-)	13	(11)
Main objective**	Patient and consumer experience' factors (A)	12	(35)	40	(35)
	Valuing health outcomes (B)	3	(9)	8	(7)
	Trade-offs health outcomes and experience factors (C)	14	(41)	38	(33)
	Estimating utility weights within QALY framework (D)	0	(-)	2	(2)
	Job-choices (E)	2	(6)	5	(4)
	Developing priority setting frameworks (F)	2	(6)	6	(5)
	Health professional's preferences for treatment or screening options for patients (G)	1	(3)	17	(15)
	Other (H)	0	(-)	4	(4)

* Percentages do not up to 100% exactly due to rouding error.

** Totals do not up to 100% as each study can have more than one main objective

demonstrates the breadth of applications (there has been a move from the term 'conjoint analysis' to 'Discrete Choice Experiments' (DCEs) and 'Stated Preference Discrete Choice modelling' (SPDCM)). As discussed by Ryan and Gerard [6], DCEs were introduced into health economics in the early 1990s as a method to go beyond the Quality Adjusted Life Years (QALYs), and consider broader aspects of value. More specifically, DCEs were proposed as a technique to value what might be termed patient experience factors [15-16]. A large number of studies in health still addressed issues such as the relative importance of patient experience attributes, trade-offs between attributes, willingness to pay for marginal changes in attributes as well as the overall monetary value of different configurations of services (studies A1-A40). However, within an economic evaluation framework, DCEs are now used to value health outcomes (studies B1-B8); the trade-offs individuals make between health outcomes and patient experience factors (studies C1-C38) and, more recently, to estimate utility weights within the QALY framework (D1 and D2). Applications have also extended beyond economic evaluation, to investigate labour-market choices amongst health care professional (studies E1-E5), priority setting frameworks at the local/national level (studies F1-F6) and preferences regarding clinical decision making (studies G1 and G17).

The UK remained the major contributor to the literature when we compared the years 1990-2000 to 2001-2008, with the US, Australia, and Canada also being major contributors (Table 2.1). However 26 studies (23%) were conducted in countries where

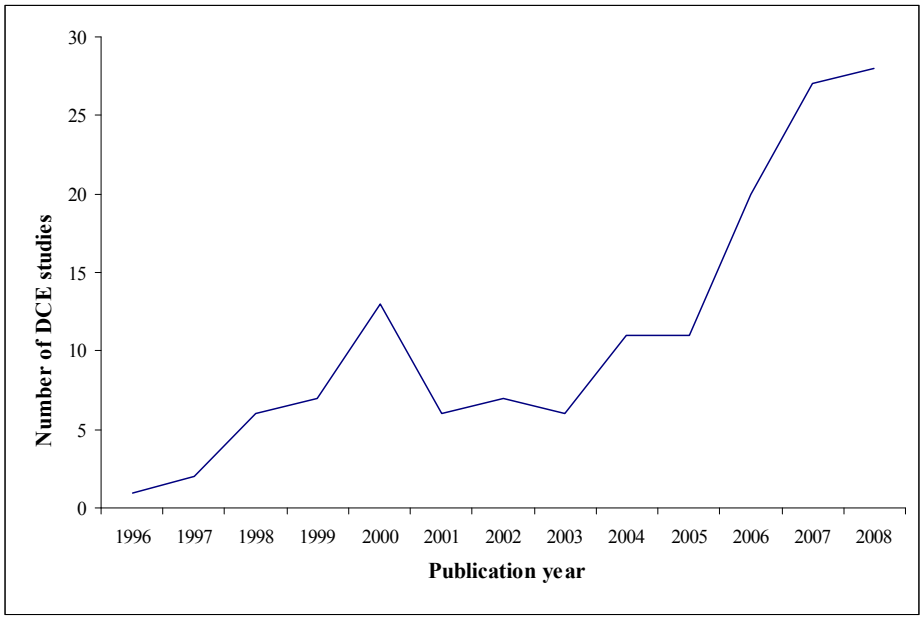


Figure 2.1 Number of DCE studies by year of publication

the technique had not been applied in the baseline review. These included high income, middle income and low income countries. Figure 2.1 shows that the number of applications of DCEs in health is gathering pace – many more studies were published in the period 2006-2008 (73 studies) as in the previous five years (2001-2005; 41 studies).

2.3 Key issues arising from baseline review

2.3.1 Experimental design and choice set construction

A crucial aspect of constructing a DCE is defining the choices in order that parameter estimates are reliable. A full factorial design includes all possible combinations of attributes and levels for making profiles or choice sets. This often results in a large number and experimental design methods are used to create smaller fractional factorial designs. When employing fractional factorial designs the researchers' task is to select an experimental design and construct choices to minimise the variation in relevant parameter estimates [1].

Recent years have seen rapid developments in experimental design methods employed, including orthogonal and D-efficient designs. Orthogonal designs are often based on orthogonal arrays from design catalogues (e.g. Hahn and Shapiro [17]), statistical programs (e.g. SPEED [18]); SPSS, (SPSS Inc., Chicago**,** IL**,** USA)) or web-sites (Sloane [19]). These arrays have the properties of orthogonality (attributes are statistically independent of one another) and level balance (levels of any given attribute appear the same number of times). If a binary choice DCE is employed (a binary choice is where respondents are presented with a number of profiles (one at a time) and asked if they would choose/participate, with possible responses being 'yes' or 'no'), then the profiles generated from the orthogonal design are the choices (e.g., studies C2, C34, F4, G9 and H3).

Where choices present two or more options, choice sets must be created. Whilst initial methods for doing this included random pairing of a set of orthogonal choices (e.g. 16 profiles from an orthogonal array would be randomly allocated into 8 choices) or the use of a constant comparator (which may have been taken from the orthogonal array or have been the current situation), the period since the previous review has seen the development of alternative methods. Louviere et al. [2] propose two methods to move from an orthogonal array to a set of choices: foldover (each of say 16 profiles are paired with their foldover (foldover is a mirror image of the original design [2, 20]; (e.g., in case of two level attributes replace each 0 with 1, and each 1 with 0; in the case of three level attributes replace each 0 with 1, 1 with 2, and 2 with 0) to create 16 choice sets) or foldover with random pairing (here an orthogonal array of say 16 profiles are randomly paired with their foldover). Readily available designs are available using these methods [20-25], as well as help from experts (see software of Street and Burgess [26]).

More recently statistically efficient designs have been developed [20-22, 27]. Statistical efficiency may be measured in a number of ways (see Liang et al. [28] for more details). The D-efficiency criterion (precision of parameter estimates) has been the more popular criterion, primarily because it is computationally less cumbersome. Here the variance–covariance matrix is minimized. Zwerina et al. [29] developed a computer generated statistically D-efficient design using Cook and Nachtsheim's [30] modified Fedorov [31] algorithm. Here the algorithm searches for a choice design which minimizes the D-error or equivalently maximizes the D-efficiency for the nonlinear multinomial logit choice model. This is readily available within SAS software [32].

When deriving efficient designs it has been common to assume, a priori, that parameters are zero. A recent development is to use prior assumptions about parameters [33-41]. It is argued that the statistical efficiency of designs can be improved using informative priors since choices can be derived that rule out dominant alternatives and maximize the information obtained from each choice situation [42]. If there is no prior information a design could be created using zero priors, and given to a small sub-sample (e.g. 10%) of the study's subjects. The estimated parameters from this sample could subsequently be used as priors to create a revised efficient design which could then be given to the remaining sample (e.g., 90%) [42].

2.3.1.1 Experimental design and choice set construction in practice

As with the previous review, main effects fractional factorial designs still dominate (Table 2.2). Forty-two studies (37%) did not report the design source. In studies that reported the design source, software packages remain the most popular design source (59 studies; 52%); 22 studies (20%) reported using SPEED to create an orthogonal array and 14 studies (12%) used SPSS. When using these orthogonal main arrays, choice sets still need to be created. Current practice has changed with respect to this. The less efficient random pairing method is less commonly used, and there has been a move towards foldover methods and D-efficient designs. For example, 12 studies used foldover methods (study A27 used foldover with random pairing; studies A21, A32, A36, C24, C26, C31, C33, D2, F2, F3, and F6 paired the profiles with their foldover) or used specific software to create orthogonal choices (five US studies used Sawtooth software to create orthogonal choices (A39, B4, C21, C29, and H2). This software is used extensively in market research. It might be worth exploring further for relevance to health economics applications given the completeness of the package – however this software is yet not too for conjoint studies only so therefore limited). All 14 studies employing D-efficient designs used SAS software to create choices (studies A20, A25, A30, A34, A35, A38, B5, C27, C38, F1, G10, and G15) or gave no further details (studies A5 and A37). There had been no application of designs incorporating a priori assumptions for parameter estimates and a further 32 studies (28%) which did not report sufficient detail of the design method used.

Table 2.2 Experimental design and construction of choice sets

Item		Baseline: 1990-2000		Current: 2001-2008	
		N=34	*(%)*	*N=114*	*(%)*
Design type	Full factorial	4	(12)	0	(-)
	Fractional factorial	25	(74)	114	(100)
	Not clearly reported	5	(15)	0	(-)
Design plan	Main effects only	25	(74)	100	(89)
	Main effects, 2-way interactions	2	(6)	6	(5)
	Not applicable	4	(12)	0	(-)
	Not clearly reported	3	(9)	8	(7)
Design source	Software package	19	(56)	59	(52)
	SPEED	13	(38)	22	(19)
	SPSS	2	(6)	14	(12)
	SAS	0	(-)	14	(12)
	SAWTOOTH	2	(6)	5	(4)
	Other	2	(6)	0	(-)
	No further details	0	(-)	4	(4)
	Catalogue	2	(6)	6	(5)
	Website	0	(-)	3	(3)
	Expert	4	(12)	4	(4)
	Not clearly reported	9	(26)	42	(37)
Design method*	Orthogonal rays				
	One profile per choice set (e.g. binary choices)	3	(9)	12	(11)
	Random pairing	18	(53)	19	(17)
	Pairing with constant comparator	6	(18)	23	(20)
	Foldover - random pairing	0	(-)	1	(1)
	Foldover	0	(-)	11	(10)
	D-efficiency (SAS)	0	(-)	14	(12)
	Other (pragmatically chosen)	4	(12)	2	(2)
	Not clearly reported	3	(9)	32	(28)

Notes: * Percentages do not up to 100% exactly due to rounding error

2.3.2 Estimation procedure

Responses from the choices are modelled within a random utility theory (RUT) framework [43]. Here it is assumed that whilst the individual knows the nature of her utility function, the researcher does not. Therefore, as shown in Equation 1, the latent utility of an alternative i in a choice set C_n (as perceived by individual n) is decomposable into two additively separable parts: (i) a systematic (explainable) component specified as a function

of the attributes of the alternatives $V(X_{in}, \beta)$; and (ii) a random (unexplainable) component ε_{in} representing unmeasured variation in preferences.

$$U_{in} = V(X_{in}, \beta) + \varepsilon_{in} \quad \text{(Eq. 1)}$$

It is assumed that individual n will choose alternative i if that alternative maximizes her utility amongst all alternatives in the choice set C_n.

The starting point of model selection is the choice of the error distribution (and the distribution of the differences in errors of the alternatives). This determines whether it is appropriate to use probit, logit, multinomial logit or more advanced models. When the choice faced by respondents in a DCE is binary or if the choice set includes only two 'forced' alternatives (i.e. would you choose alternative A or B), binary probit or logit models are appropriate. Early studies in health predominantly focused on binary choice or 'forced' choice, using these (random effects; i.e. using panel specification) logit or probit models to analyse the response data [6]. As more studies began to collect multinomial rather than binary choice data (e.g., by including neither or opt-out options to allow for realism), the binary logit/probit models were inappropriate, and the workhorse to analyse multinomial choice has been the McFadden's multinomial logit (MNL) [43]. The MNL has four important assumptions: (i) identically distributed errors (i.e., constant error variance or homoscedasticity); (ii) independent errors (i.e., independence of irrelevant alternatives (IIA); thus assuming that all options are equal substitutes); (iii) no panel data (i.e., no correlation allowed for within responses); and (iv) no taste variation (i.e. homogenous preferences across respondents). However, the assumptions of MNL may be restrictive in describing human behaviour and in turn may restrict the realism of the policy analysis that follows. Therefore, much research effort has been (and continues to be) devoted to increasing the behavioural realism of choice models.

Figure 2.2 shows three alternative families of models which were developed to relax the restrictions of the McFadden's MNL model (see Ryan et al. [1] for more details): (i) the heteroscedastic models, which relax the assumption of identically distributed errors; (ii) the generalised extreme value (GEV) models, which relax the IIA assumption partially [44, 45] (e.g. the nested logit model (NL)); and (iii) the flexible models, which relax the assumptions of independent errors and allow for random taste variation (MNL, GEV and heteroscedastic models allow for heterogeneous preferences using sub-group analysis; the key contribution of the flexible models is that sub-groups do not have to be identified in advance, the data does this) and for multiple observations as well (e.g. mixed logit model (MXL) and latent class model (LCM)).

More recently it has been argued that the normal mixing distribution commonly used in the MXL is mis-specified [46-50]. More specifically, it is argued that much of the heterogeneity in attribute weights is accounted for by a scale effect, with the scale varying

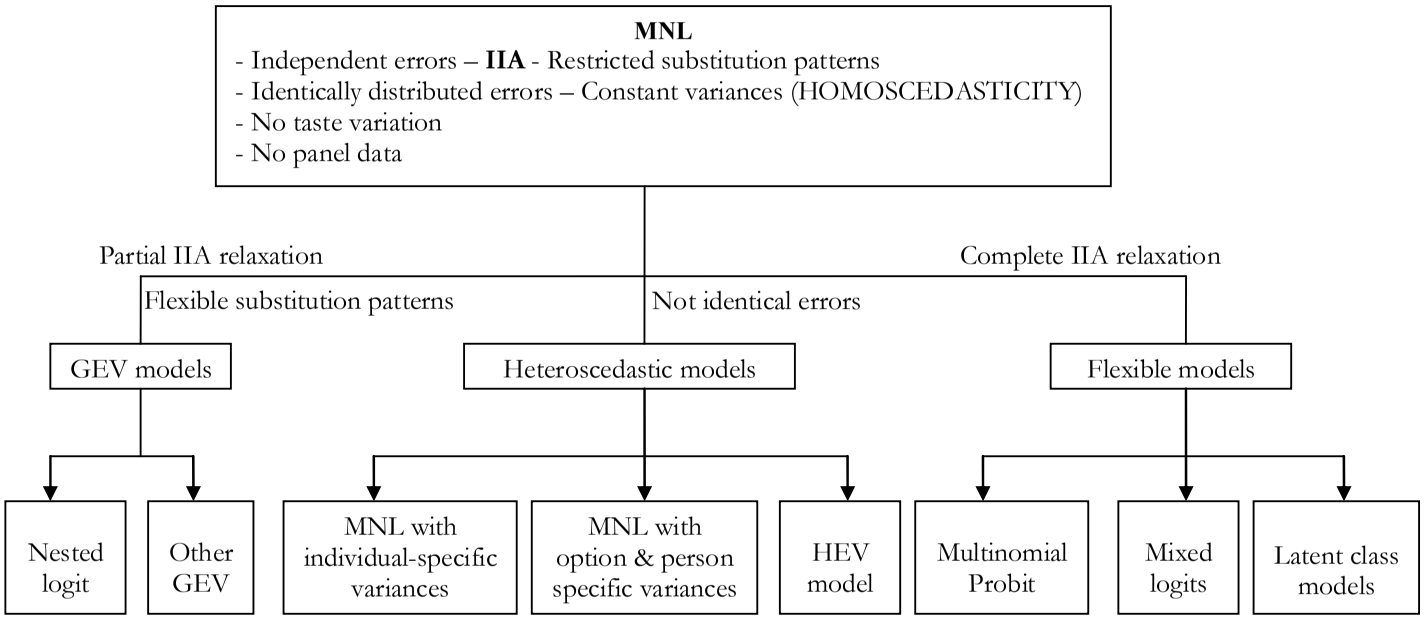


Figure 2.2 State of the art in choice modelling (source: Ryan et al. [1])

across individuals. However, the "scale" is not identified in discrete choice data – a problem that is typically resolved by normalizing it to a constant. Fiebig et al. [50] found that models that account for scale heterogeneity (i.e., 'generalised'-MNL or 'scale heterogeneity'-MNL) are preferred to the MXL.

2.3.2.1 Methods of analysis used in practice

Five studies in the current review period used the NL model (A18, C17, C20, C36, and C37) compared to none in the baseline period (Table 2.3). In these studies it was assumed that the alternatives of the given health service were closer substitutes than 'opting out' (i.e. no health service). Use of the nested logit model also allowed testing whether the levels of attributes within an alternative affected the decision to attend a health care intervention/service. For example, Study C17 showed that the attributes offered by the cervical screening programs did not influence a woman's decision to participate, and therefore altering the screening attributes will have no impact on uptake.

The LCM was applied in one study in the current review (Study A34), giving insight into preference distribution for the attributes of a general practitioner appointment. Significant preference heterogeneity for the attributes of a general practitioner appointment included in the experiment was found, and the LCM led to a significant improvement in fit compared to the logit model. In this LCM model the class membership probability was a function of constants only, implying that the probability of belonging to each class was

Table 2.3 Estimation procedures

Item	Category	Baseline: 1990-2000		Current: 2001-2008	
		N=34	*(%)*	*N=114*	*(%)*
Estimation procedure*	Probit	6	(18)	8	(7)
	Random effects probit	18	(53)	47	(41)
	Logit	1	(3)	13	(11)
	Random effects logit	1	(3)	6	(5)
	MNL	6	(18)	25	(22)
	Nested logit (NL)	0	(-)	5	(4)
	Mixed logit (MXL)	1	(3)	6	(5)
	Latent class (LCM)	0	(-)	1	(1)
	Other**	1	(3)	4	(4)
	Not clearly reported	2	(6)	4	(4)

Notes: * Totals do not up to 100% as each study can use more than one estimation procedure
** OLS; Generalised estimating equations; Monte Carlo Markov Chain Algorithms; Cox proportional hazard model

constant across individuals. This assumption can be relaxed by including socio-demographic characteristics in the class membership model. Investigating the probability of belonging to a given class is an advantage of LCM compared with MXL.

Six studies in the current review period employed the MXL model (A30, A34, A38, C16, C19, and C25), compared to one in the baseline review. All six studies found evidence of preference heterogeneity. Studies A30, A34, A38, and C16 reported that the MXL resulted in an improvement in goodness of fit compared to the binary logit or MNL. For example, study A30 found that the attribute 'group meetings' for cardiac rehabilitation activity was insignificant when analysed using a logit model, but highly significant with a large standard deviation using MXL; two-thirds of the variation showed a positive preference for the attribute and one-third a negative preference. The MXL model provided therefore additional policy relevant information as well as a better fit to the data than a simple model.

The MXL has a number of disadvantages such as what parameters to randomised, what distributions to impose, and the amount of choice data. These were borne out in the applications. Study C19 faced a relatively small sample size (n=57); therefore only the intercepts were assumed random. Study A30 assumed all parameters to be normally distributed, thus allowing positive or negative preferences, and noted that other distributions such as lognormal or some bounded distribution would be preferable. For example, Study C16 assumed that the price proxy (cost attribute) and the waiting time for results of genetic carrier testing was normal distributed when positive values would not be expected. Issues are also raised when estimating WTP using the MXL model. Study A34 showed that the MXL, where the cost coefficient was allowed to vary, had a markedly better fit than the MXL in which the cost coefficient was fixed. However, for estimating WTP a fixed cost coefficient seems to be better, since this: i) ensures that the coefficient has the correct sign; and ii) implies WTP is normally distributed if the remaining coefficients are normally distributed [51].

2.3.3 Validity

Given DCEs rely on responses to hypothetical data, investigating validity of responses is crucial. Table 2.4 presents a summary of tests of validity. The best test demonstrates external validity: i.e. whether or not behaviour stated in a hypothetical context is carried through into the choices made in the real world. External validity is perhaps the single most important research question for DCE in health economics. Yet very little research has been undertaken. This probably reflects the difficulty of investigating this issue in publicly provided health care systems where respondents have limited choice and usually do not pay at the point of consumption (although there are exceptions to this rule). In the only study identified which attempted to test external validity, Telser and Zweifel [52] aimed to study the external validity of preferences for hip protectors. Their study related

WTP values estimated in a DCE (the stated choice) to the same respondents' later willingness to participate in a wearing trial of two months' duration (the actual choice). Whilst this study provides some insight into validity, a limitation was that we do not know if respondents actual WTP was as stated in the stated preference experiment. Hence, this study was not comparing the performance of stated preference to actual behaviour.

Evidence of internal validity of DCEs is more common, including tests of theoretical validity (whether the parameters move in the expected direction) and tests of the standard assumptions on individual preferences (non-satiation, transitivity, Sen's expansion and contraction properties [53]; and compensatory decision-making). Tests of theoretical validity are easily performed within a study, and the results continue to be encouraging, with coefficients moving in the expected directions. Non-satiation continues to be the most common internal test, despite concerns of being too easy to pass [54] and not being essential for rationality [5]. However, transitivity, a fundamental test of rationality when using the preference based approach to preference elicitation [5], does not seem to be growing in use, perhaps reflecting the difficulty of applying this test within a DCE. Two studies had employed rationality tests from choice theory, based on Sen's contraction consistency principle (studies D2 and H1). Tests of compensatory decision making (whether respondents consider all the attributes when making choices) continue to be investigated by investigating whether respondents always chose according to the best level of a given attribute. Such a decision-making pattern is taken as evidence of non-

Table 2.4 Validity

Item	Category	Baseline: 1990-2000		Current: 2001-2008	
		N=34	*(%)*	*N=114*	*(%)*
Validity*	External validity tests	0	(-)	0	(-)
	Internal validity tests				
	Theoretical	22	(65)	64	(56)
	Non-satiation	15	(44)	56	(49)
	Transitivity	3	(9)	5	(4)
	Sen's expansion and contraction	0	(-)	2	(2)
	Compensatory decision making	12	(35)	36	(32)
Use of qualitative methods*	Increasing face validity				
	Attribute selection	6	(18)	79	(69)
	Level selection	6	(18)	38	(33)
	Pre-testing questionnaire	16	(47)	36	(32)
	Strengthen understanding responses				
	Debriefing choices	0	(-)	5	(4)

Note: * Totals do not add to 100% as each study can use validity tests and/or qualitative methods for more than one category

compensatory decision-making. However, Ryan and Gerard [6], noted that such tests are limited since compensatory behaviour may also result in such a choice pattern. Further, even if respondents are not trading, and marginal rates of substitution therefore cannot be estimated, these preferences are still important from a policy perspective [5].

Once respondents have been identified who 'fail' the axioms of economic theory, the question is raised of what to do with them. Testing for internal validity should not automatically lead to deleting responses based on 'irrational' preferences or 'not-compensatory' decision making. However, 30 out of 74 studies that tested for such internal validity deleted responses a priori to the analyses (e.g., studies A7, A22, A25, B2, C5, C30, E3, and G5), which may result in the removal of valid preferences, induce sample selection bias, and reduce the statistical efficiency and power of the estimated choice models [5]. A number of studies reported separate analyses to investigate the implication of retaining these subjects with dominant preferences, and reported that the findings were not (dramatically) different whether or not these subjects were retained (see A23, A26, B6, B7, C3, C18, and C21).

Qualitative work has been proposed to develop the attributes and levels within a DCE (increasing the face validity, Coast and Horrocks [55]), as well as to better understand responses [54, 56]. Regarding the former, focus groups (e.g., studies A6, A11, A12, A20, B2, C6, C15, and G1) and interviews (e.g., A8, A26, C1-C3, C7 and G6) have been increasingly used to enhance attribute and level selection as well as the lay-out, comprehension and design of the DCE questionnaire.

In the current review, five studies used qualitative approaches to debrief respondents in order to strengthen understanding of why people state particular preferences and, in turn, how results are to be interpreted (studies C4, C20, C24, C32 and G8). Studies C4 and C20 collected data about the participant's understanding of the questionnaire via a follow-up telephone interview; a couple of patients in Study C24 completed the survey in a 'think aloud' exercise, followed by a cognitive debrief interview to identify areas of misunderstanding; and studies C32 and G8 examined the free-text comments provided by the respondents in order to examine further their experiences of choices.

2.4 Discussion

Whilst DCEs were introduced into health economics to value patient experience factors, applications today are much broader. The technique has been used to value health outcomes, investigate trade-offs between health outcomes and patient experience factors, and more recently, to estimate utility weights within the QALY framework. Applications have also extended beyond economic evaluation, to investigate labour-market choices amongst health care professional, priority setting frameworks at the local/national level and preferences regarding clinical decision making. The number of applications has grown rapidly and has spread to many different health care systems that had previously not applied the technique.

Our review of the technical practice of DCE in health economics suggests a number of important developments. Although Guttman et al. [11] concluded that the methods and design employed are quite similar during the past two decades, we found that the importance of the experimental design component of a DCE is being increasingly recognised in health economics. The less efficient random pairing method was less commonly used, with a move towards D-efficient designs. As a result, the precision of the parameter estimates may be increased whereas the biases of the results may be decreased. Almost all studies that used statistically efficient designs were conducted in the most current period 2006-2008, which is an encouraging development. In these studies SAS software was frequently used, probably because it creates choice sets directly and is flexible (e.g. possibility of restricted designs and designs taking into account alternative specific parameters). Street and Burgess [24] found that design methods available in SAS often produce efficient designs, but the effects are not fully independent, resulting in some attribute effects with shared covariances. Recent software has been designed specifically for stated choice experimental designs (Ngene, http://www.choice-metrics.com/), allowing for the generation of various stated choice experimental designs. We expect this software to gain ground in health economics for designs using prior assumptions about parameters. Louviere and Lancsar [13] note that "more research is needed to identify general conditions where a priori knowledge of signs and/or true parameters will significantly improve design properties." This is likely to be an important area for future research. We would also hope to find less studies published without details of the experimental design methods and source.

The use of richer econometric models for estimation is in its infancy in health economics, although there is some evidence exists of a move towards models with greater flexibility. Some current studies used NL and LCM models compared to none in the baseline period 1990-2000. Also the MXL model was used more often. As a result, particularly the adjustment for preference heterogeneity increased. Wansbeek et al. [57] has singled heterogeneity out as 'the most salient feature of consumer demand at the micro level'. Increasing computational power and increasingly richer data sets, allows models such as MXL and LCM, or even heteroscedastic models, to be empirically testable. There seems to be policy relevance to move towards richer econometric models, because these models can give additional policy relevant information, and may avoid wrong policy conclusions due to incorrect assumptions in the choice model. However, there are also problems in applying these more advanced models, such as that the researcher needs to make additional decisions regarding which coefficients to vary with which distribution, and the number of latent classes. The limitations of these advanced modelling methods should be recognised by practitioners. Following on from this, we expect MXL models will lose ground in health economics, and that in the near future 'new' models (such as 'generalised'-MNL or 'scale heterogeneity'-MNL) will be introduced that account for scale heterogeneity.

Thirdly, external validity is still under-researched. The challenge was posed by Ryan and Gerard [6] to find imaginative solutions to test this fundamental research question. Hereto,

stated and actual behaviour have to be compared. At this point we may learn from other areas. Studies from other areas (Environmental and Agricultural economics) investigated the external validity of DCEs by using controlled laboratory experiments [58-62] (a controlled laboratory experiment takes place in a laboratory with subjects knowing they are part of an experiment), framed field experiments [63] (a framed field experiment takes place in the field context with subjects knowing they are part of an experiment), or natural field experiments [62] (a natural field experiment takes place in the field without subjects knowing they are part of an experiment), which is the strongest external validity test. In a laboratory or framed field experiment, individual's desire to "do the right thing", or make the "moral" choice [64] may influence behaviour. Laboratory experiments may however provide a starting point to test for external validity of DCEs in health or health care. It is hoped that imaginative solutions emerge to test the external validity of DCEs in health economics. Internal validity tests, on the other hand, are largely incorporated in the DCE practice in health economics, which should be continued. Further research is needed to find out if transitivity tests or Sen's contraction property tests are more appropriate to test the internal validity of DCE responses than non-satiation tests. Mixed methods may be useful to get more insight into the internal validity of DCEs. Qualitative techniques, such as the think aloud technique, may show that seemingly 'irrational' choice behaviour may not be so irrational after all [56]; and thus may provide clues whether deleting 'irrational' preferences or 'non-compensatory' simple tests are valid. Further, even if respondents are not trading, and marginal rates of substitution therefore cannot be estimated, these preferences are still important from a policy perspective [5].

A further issue raised at the time of the 1990-2000 survey was generalisability of results. The recent growth of DCE in health economics means that there is now a stock of over 150 published studies (and growing). It follows that the point has been reached where it is viable and timely to consider theoretical, methodological and practical application of credible DCE transfers in health economics. It has also been argued that DCE methodology may be particularly well suited for transfers because of its conceptually rich foundation [65]. Progress in this area would benefit policy makers, such as at the NHS, by informing decision-makers and commissioners of research how to best deal with situations when it may be more efficient to make use of existing information rather than pay for new studies or when information is needed quickly, or new empirical studies cannot be afforded.

Following on from this, one notable feature of our review is the absence of the use of DCE output within an economic evaluation decision-making framework. This is clearly an important area for future research. Whilst QALYs continue to be the recommended measure of value when making recommendations regarding provision of alternative health care interventions by bodies such as NICE [66], the importance of the patient experience is also being increasingly recognised [67, 68]. Future work should explore the use of DCEs within an economic evaluation modelling framework.

One significant methodological limitation of our study is that we may not be accurately representing the state of DCE practice due to publication lags. It could well be that more

progress has been attained that we are able to give credit for (i.e., the 'picture' is probably better than this review implied). This can only come to light at a later time. For example, the 11 studies that were found to embrace richer estimations procedures (i.e., studies A16, A30, A34, A38, C16, C17, C19, C20, C25, C36, and C37) were also more recently published (i.e. 2005-2008).

In summary, the current systematic review shows that the applications of DCEs today are much broader than valuing patient experience factors. Regarding methodological issues, recently published DCEs in health economics made more use of foldover methods and D-efficient criterion for designing DCEs compared to 1990-2000. Also the use of models with greater flexibility such as NL, LCM and MXL slightly increased in time. There has been a shift towards use of more statistically efficient designs and richer econometric models. However, much progress has still to be made towards external validity. Also consideration needs to be given to incorporation of the results of a DCE into a decision-making framework by policy makers.

References

1. Ryan M, Gerard K, Amaya-Amaya M (2008). Using Discrete Choice Experiments to Value Health and Health Care. (Vol. 11). Dordrecht, The Netherlands: Springer.
2. Louviere J, Hensher DA, Swait JD (2000). Stated choice methods: analysis and application. Cambridge: Cambridge University Press.
3. Hensher DA, Rose JM, Greene WH (2005). Applied choice analysis: a primer. Cambridge: Cambridge University Press.
4. Bliemer MCJ, Rose JM (2006). Designing Stated Choice Experiments: State-of-the-art. Paper presented at the 11th International Conference on Travel Behaviour Research.
5. Lancsar E, Louviere J (2006). Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? Health Econ 15: 797-811.
6. Ryan M, Gerard K (2003). Using discrete choice experiments to value health care programmes: current practice and future research reflections. Applied Health Economics Health Policy 2: 55-64.
7. Hanley N, Ryan M, Wright R (2003). Estimating the monetary value of health care: lessons from environmental economics. Health Economics 12: 3-16.
8. Fiebig DG, Louviere J, Waldman D (2005). Contemporary issues in modelling discrete choice experimental data in health economics. Working paper, University of New South Wales [Electronic Version] from http://wwwdocs.fce.unsw.edu.au/economics/staff/DFIEBIG/ContemporaryHEv120Apr05.pdf.
9. Viney R, Savage E, Louviere J (2005). Empirical investigation of experimental design properties of discrete choice experiments in health care. Health Econ 14: 349-362.
10. Belkar R, Fiebig DG, Haas M, Viney R (2006). Why worry about awareness in choice problems? Econometric analysis of screening for cervical cancer. Health Economics 15: 33-47.

11. Guttman R, Castle R, Fiebig D (2009). Use of discrete choice experiments in health economics: an update of the literature. CHERE Working Paper 2009/2, UTS, Sydney.
12. Lancsar E, Louviere J (2008). Conducting discrete choice experiments to inform healthcare decision making. A user's guide. Pharmacoeconomics 26 (8) 661-677.
13. Louviere J, Lancsar E (2009). Choice experiments in health; the good, the bad, and the ugly and toward a brighter future. Health Economics, Policy, and Law. In press.
14. Flynn TN, Louviere J, Peters TJ, Coast J (2007). Best--worst scaling: What it can do for health care research and how to do it. Journal of Health Economics 26: 171-189.
15. Ryan M (1999). Using conjoint analysis to go beyond health outcomes and take account of patient preferences: an application to in vitro fertilisation. Social Science and Medicine 48: 535-546.
16. Ryan M, Hughes M (1997). Using conjoint analysis to assess women's preferences for miscarriage management. Health Economics 6: 261-273.
17. Hahn GJ, Shapiro SS (1966). A catalog and computer program for the design and analysis of orthogonal symmetric and asymmetric fractional factorial experiments. Schenectady, NY, USA: General Electric Research and Development Center.
18. Bradley M (1991). User's manual for the speed version 2.1 stated preference version 2.1 stated preference experiment editor and designer. Hague Consulting Group, Hague.
19. Sloane NJA (2009). A library of orthogonal arrays. http://www.research.att.com/~ njas/oadir/. [Accessed 2009 April 09]
20. Street DJ, Burgess L, Louviere J (2005). Quick and easy choice sets: constructing optimal and nearly optimal stated choice experiments. Intern J of Research in Marketing 22: 459-470.
21. Burgess LB, Street D (2003). Optimal designs for 2(k) choice experiments. Communications in Statistics: Theory and Methods 32 (11) 2185-2206.
22. Burgess LB, Street D (2005). Optimal designs for choice experiments with asymmetric attributes. Journal of Statistical Planning and Inference 134 (1) 288-301.
23. Street DJ, Burgess L (2004). Optimal and near-optimal pairs for the estimation of effects in 2-level choice experiments Journal of Stat Plan & Inference 118: 185-199.
24. Street DJ, Burgess L (2007). The Construction of Optimal Stated Choice Experiments: Theory and Methods. Hoboken, New Jersey, Wiley.
25. Street DJ, Burgess L, Viney R, Louviere J (2008). Designing Discrete Choice Experiments for Health Care. Ryan M, Gerard K, and Amaya-Amaya M (eds) In Using Discrete Choice Experiments to Value Health and Health Care (pp. 47-72). Dordrecht, The Netherlands: Springer.
26. Street DJ, Burgess L (2007). Discrete choice experiments [computer software]. Sydney: University of Technology, [online]. Available from URL: http: //crsu.science.uts.edu.au/choice/ [Accessed 31 March 2009].
27. Zwerina K, Huber J, Kuhfeld A (2005). A general method for constructing efficient choice designs (SAS technical notes TS772e): The SAS Institute 2005.
28. Liang L, Anderson-Cook CM, Robinson TJ (2005). Cost penalized estimation and prediction evaluation for split-plot designs. http://www.stat.org.vt.edu/dept/web-e/tech_reports/TechReport05-1.pdf [Accesses 23 February 2009].
29. Zwerina K, Huber J, Kuhfled W (1996). A general method for constructing efficient choice designs. SAS working paper. http://citeseerx.ist.psu.edu/viewdoc/download?doi =10.1.1.31.9438&rep=rep1&type=pdf. [Accessed 12 May 2009].

30. Cook RD, Nachtsheim CJ (1989). Computer-aided blocking of factorial and response-surface designs. Technometrics 31: 339-346.
31. Fedorov VV (1972). Theory of optimal experiments, translated and edited by WJ Studden and EM Klimko, New York: Academic Press.
32. Kuhfeld W (2000). Marketing Research Methods in the SAS System, Version 8 Edition. SAS Institute.
33. Sandor Z, Wedel M (2001). Designing conjoint choice experiment using mangers' prior beliefs. Journal of Marketing Research 38: 430-443.
34. Sandor Z, Wedel M (2002). Profile construction in experimental choice designs for mixed logit models. Marketing Science 21: 455-475.
35. Sandor Z, Wedel M (2005). Heterogeneous conjoint choice designs. Journal of Marketing Research 42: 210-218.
36. Rose JM, Bliemer MCJ (2005). Sample optimality in the design of stated choice experiments. Working paper ITLS-WP-05-13.
37. Rose JM, Bliemer MCJ (2008a). Stated preference experimental design strategies, in Hensher DA and Button KJ (eds). Handbook of Transport Modelling, Elsevier, Oxford, Ch 8, 151-180.
38. Ferrini S, Scarpa R (2007). Designs with a priori information for nonmarket valuation with choice experiments: A Monte Carlo study. Journal of Environmental Economics and Management 53 (3) 342-363.
39. Kessels, R., Goos, P. and Vandebroek, M (2006). A Comparison of Criteria to Design Efficient Choice Experiments. Journal of Marketing Research 43 (8) 409-419.
40. Rose JM, Bliemer MCJ, Hensher DA, Collins AT (2008). Designing efficient stated choice experiments in the presence of reference alternatives. Transportation Research Part B: Methodological 42 (4) 395-406.
41. Bliemer MCJ, Rose JM, Hensher DA (2009). Efficient stated choice experiments allowing for estimating nested logit models. Transportation Research Part B 43(1) 19-35.
42. Rose JM, Bliemer MCJ (2008b). Discrete Choice Analysis and Choice Experiment Design. Executive Course, November 26-29, 2008, Rotterdam, the Netherlands.
43. McFadden, D (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), Frontiers in Econometrics (pp. 105-142). New York: Academic Press.
44. McFadden D (1979). Quantitative methods for analysing travel behaviour of individuals: some recent developments. In: Behavioural Travel Modelling. Hensher D and Sotpher P (eds). London: Croom Heml, pp. 279-318.
45. McFadden D (1981). Ecenometric models of probabilistic choice. In: Structural Analysis of Discrete Data with Econometric Applications. Maski C and Macfadden D (eds). Cambridge: MIT Press pp. 198-272.
46. Louviere, JJ, Carson RT, Ainslie A, Cameron TA, DeShazo JR, Hensher D, Kohn R, Marley T, Street DJ (2002). Dissecting the random component of utility. Marketing Letters 13: 177-193.
47. Louviere JJ, Eagle T (2006). Confound It! That Pesky Little Scale Constant Messes Up Our Convenient Assumptions. Proceedings, 2006 Sawtooth Software Conference, 211-228, Sawtooth Software, Sequem, Washington, USA.
48. Louviere JJ, Street D, Burgess L, Wasi N, Islam T, Marley AAJ (2008). Modeling the choices of individuals decision makers by combining efficient choice experiment designs with extra preference information. Journal of Choice Modeling 1 (1) 128-163.
49. Meyer RJ, Louviere JJ (2007). Formal Choice Models of Informal Choices: What Choice Modelling Research Can (and Can't) Learn from Behavioral Theory. Review of Marketing Research 4: 3-32.

50. Fiebig DG, Keane MP, Louviere JJ, Wasi N (2009). The Generalized Multinomial Logit Model: Accounting for Scale and Coefficient Heterogeneity. Working paper. http://www.business.uts.edu.au/censoc/papers/wp09002.pdf [accessed 31 March 2009].
51. Revelt D, Train T (1998). Mixed logit with repeated choices: households' choice of appliance efficiency level. Review of Economics and Statistics 80: 647-657.
52. Telser H, Zweifel P (2007). Validity of discrete-choice experiments evidence for health risk reduction. Applied Economics 39 (1) 69-78.
53. Sen A (1993). Internal consistency of choice. Econometrica 61: 495-521.
54. San Miguel F, Ryan M, Amaya-Amaya M (2004). Irrational stated preferences: a quantitative and qualitative investigation. Health Economics 14: 307-322.
55. Coast J, Horrocks S (2007). Developing attributes and levels for discrete choice experiments using qualitative methods. Journal of Health Services Research Policy 12: 25-30.
56. Ryan M, Watson V, Entwistle V (2009). Rationalising the 'irrational': a think aloud study of discrete choice experiment responses. Health Economics 18 (3) 321-336.
57. Wansbeek T, Meijer E, Wedel M (2001). Comment on microeconomics. Journal of Econometrics 100/101: 89-91.
58. Carlsson F, Martinsson P (2001). Do Hypothetical and Actual Marginal Willingness to Pay Differ in Choice Experiments? Application to the Valuation of the Environment. Journal of Environmental Economics and Management 41 (2) 179-192.
59. Chang JB, Lusk JL, Norwood FB (2008). External Validity of Hypothetical Surveys and Laboratory Experiments. Selected Poster prepared for presentation at the American Agricultural Economics Association Annual Meeting, Orlando, FL, July 27-29, 2008.
60. Cummings RG, Harrison GW, Rutström EE (1995). Homegrown values and hypothetical surveys: is the dichotomous choice approach incentive compatible? The American Economic Review 85 (1) 260-266.
61. Lusk JL, Schroeder TC (2004). Are Choice Experiments Incentive Compatible? A Test with Quality Differentiated Beef Steaks. American Journal of Agricultural Economics 86 (2) 467 – 482.
62. Shogren JF, Fox JA, Hayes DT, Roosen J (1999). Observed choices for food safety in retail, survey, and auction markets. American Journal of Agricultural Economics 81 (5) 1192-1199.
63. List JA, Shogren JF (1998). Calibration of the difference between actual and hypothetical valuations in a field experiment. Journal of Economic Behaviour and Organization 37: 193-205.
64. Levitt SD, List JA (2006). What do laboratory experiments tell us about the real world. http://pricetheory.uchicago.edu/levitt/Papers/jep%20revision%20Levitt%20& %20List.pdf. [Accessed April 6th, 2009]
65. Bateman IJ, Carson RT, Day B et al (2002). Economic Evaluation with Stated Preference Techniques, A Manual. Edward Elgar Publishing Limited: Cheltenham.
66. National Institute for Health and Clinical Excellence (2008). Guide to the Methods of Technology Appraisal. Available at (www.nice.org.uk).
67. Department of Health (2005). Now I feel tall: what a patient led NHS feels like. London: Department of Health.
68. Coulter A. (2005). What do patients and the public want from primary care: British Medical Journal 331 (7526) 1199-1201.

Appendix 2.A

Studies in current review are grouped by study objective.

A. Patient or consumer experience factors

Study **A1**
van der Pol M, Cairns J. Estimating time preferences for health using discrete choice experiments. Soc Sci Med. 2001 May;52(9):1459-70

Study **A2**
Longworth L, Ratcliffe J, Boulton M. Investigating women's preferences for intrapartum care: home versus hospital births. Health Soc Care Community. 2001 Nov;9(6):404-13.

Ratcliffe J, Longworth L. Investigating the structural reliability of a discrete choice experiment within health technology assessment. International Journal of Technology Assessment in Health Care. 2002; 18 (1) 139-144.

Study **A3**
Hundley V, Ryan M, Graham W. Assessing women's preferences for intrapartum care. Birth. 2001 Dec;28(4):254-63.

Hundley V, Ryan M. Are women's expectations and preferences for intrapartum care affected by the model of care on offer? BJOG. 2004 Jun;111(6):550-60.

Study **A4**
Gyrd-Hansen D, Slothuus U. The citizen's preferences for financing public health care: a Danish survey. Int J Health Care Finance Econ. 2002 Mar;2(1):25-36

Study **A5**
Moayyedi P, Wardman M, Toner J, Ryan M, Duffett S. Establishing patient preferences for gastroenterology clinic reorganization using conjoint analysis. Eur J Gastroenterol Hepatol. 2002 Apr;14(4):429-33.

Study **A6**
Phillips KA, Maddala T, Johnson FR: Measuring preferences for health care interventions using conjoint analysis: an application to HIV testing. Health Serv Res 2002 Dec, 37(6):1681-1705.

Study **A7**
Scott A, Watson MS, Ross S. Eliciting preferences of the community for out of hours care provided by general practitioners: a stated preference discrete choice experiment. Soc Sci Med. 2003 Feb;56(4):803-14.

Study **A8**
Taylor S, Armour C: Consumer preference for dinoprostone vaginal gel using stated preference discrete choice modelling. Pharmacoeconomics 2003, 21(10):721-735.

Study **A9**
Burge P, Devlin N, Appleby J, Rohr C, Grant J. Do patients always prefer quicker treatment? : a discrete choice analysis of patients' stated preferences in the London patient choice project. Appl Health Econ Health Policy. 2004;3(4):183-94.

Study **A10***
Bishop AJ, Marteau TM, Armstrong D, Chitty LS, Longworth L, Buxton MJ, Berlin C: Women and health care professionals' preferences for Down's Syndrome screening tests: a conjoint analysis study. Bjog 2004 Aug, 111(8):775-779.

Study **A11**
Dwight-Johnson M, Lagomasino IT, Aisenberg E, Hay J: Using conjoint analysis to assess depression treatment preferences among low-income Latinos. Psychiatr Serv 2004 Aug, 55(8):934-936.

Study **A12**
Roux L, Ubach C, Donaldson C, Ryan M: Valuing the benefits of weight loss programs: an application of the discrete choice experiment. Obes Res 2004 Aug, 12(8):1342-1351.

Study **A13**
Gerard K, Lattimer V, Turnbull J, Smith H, George S, Brailsford S, Maslin-Prothero S. Reviewing emergency care systems 2: measuring patient preferences using a discrete choice experiment. Emerg Med J. 2004 Nov;21(6):692-7

Gerard K, Lattimer V. Preferences of patients for emergency services available during usual GP surgery hours: a discrete choice experiment. Fam Pract. 2005 Feb;22(1):28-36.

Study **A14**
Salkeld G, Solomon M, Butow P, Short L. Discrete-choice experiment to measure patient preferences for the surgical management of colorectal cancer. Br J Surg. 2005 Jun;92(6):742-7.

Study **A15**
Hanson K, McPake B, Nakamba P, Archard L. Preferences for hospital quality in Zambia: results from a discrete choice experiment. Health Econ. 2005 Jul;14(7):687-701

Study **A16**
Ryan M, Major K, Skåtun D. Using discrete choice experiments to go beyond clinical outcomes when evaluating clinical practice. J Eval Clin Pract. 2005 Aug;11(4):328-38.

Study **A17***
Mantovani LG, Monzini MS, Mannucci PM, Scalone L, Villa M, Gringeri A: Differences between patients', physicians' and pharmacists' preferences for treatment products in haemophilia: a discrete choice experiment. Haemophilia 2005 Nov, 11(6):589-597.

Study **A18**
Ryan M, Diack J, Watson V, Smith N: Rapid prenatal diagnostic testing for Down syndrome only or longer wait for full karyotype: the views of pregnant women. Prenat Diagn 2005 Dec, 25(13):1206-1211.

Study **A19**
Longo MF, Cohen DR, Hood K, Edwards A, Robling M, Elwyn G, Russell IT. Involving patients in primary care consultations: assessing preferences using discrete choice experiments. Br J Gen Pract. 2006 Jan;56(522):35-42

Study **A20**
Kjaer T, Gyrd-Hansen D, Willaing I: Investigating patients' preferences for cardiac rehabilitation in Denmark. Int J Technol Assess Health Care 2006 Spring, 22(2):211-218.

Study **A21**
Akkazieva B, Gulacsi L, Brandtmuller A, Péntek M, Bridges JF. Patients' preferences for healthcare system reforms in Hungary: a conjoint analysis. Appl Health Econ Health Policy. 2006;5(3):189-98.

Study **A22***
Lewis SM, Cullinane FM, Carlin JB, Halliday JL: Women's and health professionals' preferences for prenatal testing for Down syndrome in Australia. Aust N Z J Obstet Gynaecol 2006 Jun, 46(3):205-211

Study **A23**
Rubin G, Bate A, George A, Shackley P, Hall N. Preferences for access to the GP: a discrete choice experiment. Br J Gen Pract. 2006 Oct;56(531):743-8

Study **A24**
Peacock S, Apicella C, Andrews L, Tucker K, Bankier A, Daly MB, Hopper JL. A discrete choice experiment of preferences for genetic counselling among Jewish women seeking cancer genetics services. Br J Cancer. 2006 Nov 20;95(10):1448-53.

Study **A25**
Porteous T, Ryan M, Bond CM, Hannaford P. Preferences for self-care or professional advice for minor illness: a discrete choice experiment. Br J Gen Pract. 2006 Dec;56(533):911-7

Study **A26**
Seston EM, Elliott RA, Noyce PR, Payne K. Women's preferences for the provision of emergency hormonal contraception services. Pharm World Sci. 2007 Feb;29(3):183-9.

Study **A27**
Hjelmgren J, Anell A. Population preferences and choice of primary care models: a discrete choice experiment in Sweden. Health Policy. 2007 Oct;83(2-3):314-22.

Study **A28**
Schwappach DL, Strasmann TJ. Does location matter? A study of the public's preferences for surgical care provision. J Eval Clin Pract. 2007 Apr;13(2):259-64.

Study **A29**
Turner D, Tarrant C, Windridge K, Bryan S, Boulton M, Freeman G, Baker R. Do patients value continuity of care in general practice? An investigation using stated preference discrete choice experiments. J Health Serv Res Policy. 2007 Jul;12(3):132-7.

Study **A30**
Kjaer T, Gyrd-Hansen D. Preference heterogeneity and choice of cardiac rehabilitation program: results from a discrete choice experiment. Health Policy. 2008 Jan;85(1):124-32.

Study **A31***
Gidman W, Elliott R, Payne K, Meakin GH, Moore J. A comparison of parents and pediatric anesthesiologists' preferences for attributes of child daycase surgery: a discrete choice experiment. Paediatr Anaesth. 2007 Nov;17(11):1043-52.

Study **A32**
Pitchforth E, Watson V, Tucker J, Ryan M, van Teijlingen E, Farmer J, Ireland J, Thomson E, Kiger A, Bryers H. Models of intrapartum care and women's trade-offs in remote and rural Scotland: a mixed-methods study. BJOG. 2008 Apr;115(5):560-9.

Study **A33**
Fitzpatrick E, Coyle DE, Durieux-Smith A, Graham ID, Angus DE, Gaboury I. Parents' preferences for services for children with hearing loss: a conjoint analysis study. Ear Hear. 2007 Dec;28(6):842-9.

Study **A34**
Hole AR. Modelling heterogeneity in patients' preferences for the attributes of a general practitioner appointment. J Health Econ. 2008 Jul;27(4):1078-94.

Study **A35**
Cheraghi-Sohi S, Hole AR, Mead N, McDonald R, Whalley D, Bower P, Roland M. What patients want from primary care consultations: a discrete choice experiment to identify patients' priorities. Ann Fam Med. 2008 Mar-Apr;6(2):107-15.

Study **A36**
Grutters JP, Joore MA, Kessels AG, Davis AC, Anteunis LJ. Patient preferences for direct hearing aid provision by a private dispenser. A discrete choice experiment. Ear Hear. 2008 Aug;29(4):557-64.

Study **A37**
Gerard K, Salisbury C, Street D, Pope C, Baxter H. Is fast access to general practice all that should matter? A discrete choice experiment of patients' preferences. J Health Serv Res Policy. 2008 Apr;13 Suppl 2:3-10.

Study **A38**
Negrín MA, Pinilla J, León CJ. Willingness to pay for alternative policies for patients with Alzheimer's Disease. Health Econ Policy Law. 2008 Jul;3(Pt 3):257-75.

Study **A39**
Wellman GS, Vidican C. Pilot study of a hierarchical Bayes method for utility estimation in a choice-based conjoint analysis of prescription benefit plans including medication therapy management services. Res Social Adm Pharm. 2008 Sep;4(3):218-30.

Study **A40**
Clark M, Moro D, Szczepura A. Balancing patient preferences and clinical needs: Community versus hospital based care for patients with suspected DVT. Health Policy. 2009 May;90(2-3):313-9. [Epub 2008 Dec 6].

B. Valuing health outcomes

Study **B1**
Osman LM, McKenzie L, Cairns J, Friend JA, Godden DJ, Legge JS, Douglas JG: Patient weighting of importance of asthma symptoms. Thorax 2001 Feb, 56(2):138-142.

McKenzie L, Cairns J, Osman L: Symptom-based outcome measures for asthma: the use of discrete choice methods to assess patient preferences. Health Policy 2001 Sep, 57(3):193-204.

Study **B2**
Ratcliffe J, Buxton M, McGarry T, Sheldon R, Chancellor J: Patients' preferences for characteristics associated with treatments for osteoarthritis. Rheumatology (Oxford) 2004 Mar, 43(3):337-345.

Study **B3***
Lee A, Gin T, Lau AS, Ng FF: A comparison of patients' and health care professionals' preferences for symptoms during immediate postoperative recovery and the management of postoperative nausea and vomiting. Anesth Analg 2005 Jan, 100(1):87-93.

Study **B4**
Haughney J, Partridge MR, Vogelmeier C, Larsson T, Kessler R, Stahl E, Brice R, Lofdahl CG: Exacerbations of COPD: quantifying the patient's perspective using discrete choice modelling. Eur Respir J 2005 Oct, 26(4):623-629.

Study **B5**
Osoba D, Hsu MA, Copley-Merriman C, Coombs J, Johnson FR, Hauber B, Manjunath R, Pyles A: Stated preferences of patients with cancer for health-related quality-of-life (HRQOL) domains during treatment. Qual Life Res 2006 Mar, 15(2):273-283.

Study **B6**
Johnson FR, Ozdemir S, Manjunath R, Hauber AB, Burch SP, Thompson TR. Factors that affect adherence to bipolar disorder treatments: a stated-preference approach. Med Care. 2007 Jun;45(6):545-52.

Study **B7**
Seston EM, Ashcroft DM, Griffiths CE. Balancing the benefits and risks of drug treatment: a stated-preference, discrete choice experiment with patients with psoriasis. Arch Dermatol. 2007 Sep;143(9):1175-9.

Study **B8**
Aspinall PA, Johnson ZK, Azuara-Blanco A, Montarzino A, Brice R, Vickers A. Evaluation of quality of life and priorities of patients with glaucoma. Invest Ophthalmol Vis Sci. 2008 May;49(5):1907-15.

C. Investigating trade-offs between health outcomes and patient or consumer experience factors

Study **C1**
Shackley P, Slack R, Michaels J. Vascular patients' preferences for local treatment: an application of conjoint analysis. J Health Serv Res Policy. 2001 Jul;6(3):151-7.

Study **C2**
Telser H, Zweifel P: Measuring willingness-to-pay for risk reduction: an application of conjoint analysis. Health Econ 2002 Mar, 11(2):129-139.

Study **C3**
Ratcliffe J, Van Haselen R, Buxton M, Hardy K, Colehan J, Partridge M: Assessing patients' preferences for characteristics associated with homeopathic and conventional treatment of asthma: a conjoint analysis study. Thorax 2002 Jun, 57(6):503-508.

Study **C4**
Hall J, Kenny P, King M, Louviere J, Viney R, Yeoh A: Using stated preference discrete choice modelling to evaluate the introduction of varicella vaccination. Health Econ 2002 Jul, 11(5):457-465.

Study **C5**
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):775-784.

Study **C6**
Salkeld G, Solomon M, Short L, Ryan M, Ward JE: Evidence-based consumer choice: a case study in colorectal cancer screening. Aust N Z J Public Health 2003, 27(4):449-455.

Study **C7**
Sculpher M, Bryan S, Fry P, de Winter P, Payne H, Emberton M: Patients' preferences for the management of non-metastatic prostate cancer: discrete choice experiment. BMJ 2004 Feb, 328(7436):382.

Study **C8**
Aristides M, Weston AR, FitzGerald P, Le Reun C, Maniadakis N: Patient preference and willingness-to-pay for Humalog Mix25 relative to Humulin 30/70: a multicountry application of a discrete choice experiment. Value Health 2004 Jul-Aug, 7(4):442-454.

Study **C9**
Watson V, Ryan M, Brown CT, Barnett G, Ellis BW, Emberton M: Eliciting preferences for drug treatment of lower urinary tract symptoms associated with benign prostatic hyperplasia. J Urol 2004 Dec, 172(6 Pt 1):2321-2325.

Study **C10**
Weston A, Fitzgerald P: Discrete choice experiment to derive willingness to pay for methyl aminolevulinate photodynamic therapy versus simple excision surgery in basal cell carcinoma. Pharmacoeconomics 2004, 22(18):1195-1208.

Study **C11***
Lee A, Gin T, Lau AS, Ng FF: A comparison of patients' and health care professionals' preferences for symptoms during immediate postoperative recovery and the management of postoperative nausea and vomiting. Anesth Analg 2005 Jan, 100(1):87-93.

Study **C12**
Fraenkel L, Constantinescu F, Oberto-Medina M, Wittink DR: Women's preferences for prevention of bone loss. J Rheumatol 2005 Jun, 32(6):1086-1092.

Study **C13**
Lloyd A, McIntosh E, Price M: The importance of drug adverse effects compared with seizure control for people with epilepsy: a discrete choice experiment. Pharmacoeconomics 2005, 23(11):1167-1181.

Study **C14**
Gerard K, Lattimer V, Surridge H, George S, Turnbull J, Burgess A, Lathlean J, Smith H. The introduction of integrated out-of-hours arrangements in England: a discrete choice experiment of public preferences for alternative models of care. Health Expect. 2006 Mar;9(1):60-9

Study **C15**
Mahadevia P, Shah S, Mannix S, Brewster-Jordan J, Kleinman L, Liebman C, O'Dowd L: Willingness to pay for sensory attributes of intranasal corticosteroids among patients with allergic rhinitis. J Manag Care Pharm 2006 Mar, 12(2):143-151.

Study **C16**
Hall J, Fiebig DG, King MT, Hossain I, Louviere JJ: What influences participation in genetic carrier testing? Results from a discrete choice experiment. J Health Econ 2006 May, 25(3):520-537.

Study **C17**
Wordsworth S, Ryan M, Skatun D, Waugh N: Women's preferences for cervical cancer screening: a study using a discrete choice experiment. Int J Technol Assess Health Care 2006 Summer, 22(3):344-350.

Study **C18**
Byrne MM, Souchek J, Richardson M, Suarez-Almazor M: Racial/ethnic differences in preferences for total knee replacement surgery. J Clin Epidemiol 2006 Oct, 59(10):1078-1086.

Study **C19**
Lancsar EJ, Hall JP, King M, Kenny P, Louviere JJ, Fiebig DG, Hossain I, Thien FC, Reddel HK, Jenkins CR: Using discrete choice experiments to investigate subject preferences for preventive asthma medication. Respirology 2007 Jan, 12(1):127-136.

King MT, Hall J, Lancsar E, Fiebig D, Hossain I, Louviere J, Reddel HK, Jenkins CR: Patient preferences for managing asthma: results from a discrete choice experiment. Health Econ 2007, 16(7):703-717.

Study **C20**
Caldow J, Bond C, Ryan M, Campbell NC, Miguel FS, Kiger A, Lee A. Treatment of minor illness in primary care: a national survey of patient satisfaction, attitudes and preferences regarding a wider nursing role. Health Expect. 2007;10(1):30-45

Study **C21**
Bishai D, Brice R, Girod I, Saleh A, Ehreth J: Conjoint analysis of French and German parents' willingness to pay for meningococcal vaccine. Pharmacoeconomics 2007, 25(2):143-154.

Study **C22**
Ossa DF, Briggs A, McIntosh E, Cowell W, Littlewood T, Sculpher M. Recombinant erythropoietin for chemotherapy-related anaemia: Economic value and health-related quality-of-life assessment using direct utility elicitation and discrete choice methods. PharmacoEconomics 2007; 25 (3) 223-237.

Study **C23**
Lloyd A, Penson D, Dewilde S, Kleinman L. Eliciting patient preferences for hormonal therapy options in the treatment of metastatic prostate cancer. Prostate Cancer Prostatic Dis. 2008;11(2):153-9.

Study **C24**
Lloyd A, McIntosh E, Rabe KF, Williams A. Patient preferences for asthma therapy: a discrete choice experiment. Prim Care Respir J. 2007 Aug;16(4):241-8.

Study **C25**
Goto R, Nishimura S, Ida T. Discrete choice experiment of smoking cessation behaviour in Japan. Tob Control. 2007 Oct;16(5):336-43.

Study **C26**
Lloyd A, Doyle S, Dewilde S, Turk F. Preferences and utilities for the symptoms of moderate to severe allergic asthma. Eur J Health Econ. 2008 Aug;9(3):275-84.

Study **C27**
Haughney J, Fletcher M, Wolfe S, Ratcliffe J, Brice R, Partridge MR. Features of asthma management: quantifying the patient perspective. BMC Pulm Med. 2007 Dec 6;7:16.

Study **C28**
Walzer S, Zweifel P. Willingness-to-pay for caregivers of children with asthma or wheezing conditions. Ther Clin Risk Manag. 2007 Mar;3(1):157-65.

Walzer S. What do parents want from their child's asthma treatment? Ther Clin Risk Manag. 2007 Mar;3(1):167-75.

Study **C29**
Langenhoff BS, Krabbe PF, Ruers TJ. Computer-based decision making in medicine: A model for surgery of colorectal liver metastases. Eur J Surg Oncol. 2007 Dec;33 Suppl 2:S111-7

Study **C30**
Snoek GJ, van Til JA, Krabbe PF, Ijzerman MJ. Decision for reconstructive interventions of the upper limb in individuals with tetraplegia: the effect of treatment characteristics. Spinal Cord. 2008 Mar;46(3):228-33.

Study **C31**
de Bekker-Grob EW, Essink-Bot ML, Meerding WJ, Pols HA, Koes BW, Steyerberg EW. Patients' preferences for osteoporosis drug treatment: a discrete choice experiment. Osteoporos Int. 2008 Jul;19(7):1029-1037.

Study **C32**
Mortimer D, Segal L. Is the value of a life or life-year saved context specific? Further evidence from a discrete choice experiment. Cost Eff Resour Alloc. 2008 May 20;6:8.

Study **C33**
Richardson G, Bojke C, Kennedy A, Reeves D, Bower P, Lee V, Middleton E, Gardner C, Gately C, Rogers A. What Outcomes Are Important to Patients with Long Term Conditions? A Discrete Choice Experiment. Value Health. 2009 12 (2) 331-339 [Epub 2008 Jul 18]

Study **C34**
Watson V, Ryan M, Watson E. Valuing Experience Factors in the Provision of Chlamydia Screening: An Application to Women Attending the Family Planning Clinic. Value Health. [Epub 2008 Sep 9]

Study **C35**
Petrou S, McIntosh E. Women's Preferences for Attributes of First-Trimester Miscarriage Management: A Stated Preference Discrete-Choice Experiment. Value Health. [Epub 2008 Sep 16]

Study **C36**
Chuck A, Adamowicz W, Jacobs P, Ohinmaa A, Dick B, Rashiq S.The Willingness to Pay for Reducing Pain and Pain-Related Disability. Value Health. [Epub 2008 Sep 16]

Study **C37**
McTaggart-Cowan HM, Shi P, Fitzgerald JM, Anis AH, Kopec JA, Bai TR, Soon JA, Lynd LD. An evaluation of patients' willingness to trade symptom-free days for asthma-related treatment risks: a discrete choice experiment. J Asthma. 2008 Oct;45(8):630-8.

Study **C38**
Herbild L, Bech M, Gyrd-Hansen D. Estimating the Danish Populations' Preferences for Pharmacogenetic Testing Using a Discrete Choice Experiment. The Case of Treating Depression. Value Health [Epub 2008 Oct 9]

D. Estimating utility weights within the QALY framework

Study D1
Ryan M, Netten A, Skåtun D, Smith P. Using discrete choice experiments to estimate a preference-based measure of outcome--an application to social care for older people. J Health Econ. 2006 Sep;25(5):927-44.

Study **D2**
Burr JM, Kilonzo M, Vale L, Ryan M. Developing a preference-based Glaucoma Utility Index using a discrete choice experiment. Optom Vis Sci. 2007 Aug;84(8):797-808.

E. Job-choices

Study **E1**
Scott A. Eliciting GPs' preferences for pecuniary and non-pecuniary job characteristics. J Health Econ. 2001 May;20(3):329-47.

Study **E2**
Ubach C, Scott A, French F, Awramenko M, Needham G. What do hospital consultants value about their jobs? A discrete choice experiment. BMJ. 2003 Jun 28;326(7404):1432

Study **E3**
Wordsworth S, Skåtun D, Scott A, French F. Preferences for general practice jobs: a survey of principals and sessional GPs. Br J Gen Pract. 2004 Oct;54(507):740-6

Study **E4**
Scott A, Bond C, Inch J, Grant A. Preferences of community pharmacists for extended roles in primary care: a survey and discrete choice experiment. Pharmacoeconomics. 2007;25(9):783-92.

Study **E5**
Mangham LJ, Hanson K. Employment preferences of public sector nurses in Malawi: results from a discrete choice experiment. Trop Med Int Health. 2008 Dec;13(12):1433-41.

F. Developing priority setting frameworks

Study **F1**
Johnson FR, Backhouse M. Eliciting stated preferences for health-technology adoption criteria using paired comparisons and recommendation judgments. Value in Health. 2006; 9 (5) 303-311.

Study **F2**
Baltussen R, Stolk E, Chisholm D, Aikins M. Towards a multi-criteria approach for priority setting: an application to Ghana. Health Econ. 2006 Jul;15(7):689-96.

Study **F3**
Baltussen R, ten Asbroek AH, Koolman X, Shrestha N, Bhattarai P, Niessen LW. Priority setting using multiple criteria: should a lung health programme be implemented in Nepal? Health Policy Plan. 2007 May;22(3):178-85.

Study **F4**
Tappenden P, Brazier J, Ratcliffe J, Chilcott J. A stated preference binary choice experiment to explore NICE decision making. Pharmacoeconomics. 2007;25(8):685-93.

Study **F5**
Ratcliffe J, Bekker HL, Dolan P, Edlin R. Examining the attitudes and preferences of health care decision-makers in relation to access, equity and cost-effectiveness: a discrete choice experiment. Health Policy. 2009 Apr;90(1):45-57. [Epub 2008 Oct 19]

Study **F6**
Green C, Gerard K. Exploring the social value of health-care interventions: a stated preference discrete choice experiment. Health Econ. [Epub 2008 Nov 25]

G. Health professional's preferences for treatment or screening options for patients

Study **G1**
Mark TL, Swait J: Using stated preference modeling to forecast the effect of medication attributes on prescriptions of alcoholism medications. Value Health 2003 Jul-Aug, 6(4):474-482.

Study **G2***
Bishop AJ, Marteau TM, Armstrong D, Chitty LS, Longworth L, Buxton MJ, Berlin C: Women and health care professionals' preferences for Down's Syndrome screening tests: a conjoint analysis study. BJOG 2004 Aug, 111(8):775-779.

Study **G3***
Lee A, Gin T, Lau AS, Ng FF: A comparison of patients' and health care professionals' preferences for symptoms during immediate postoperative recovery and the management of postoperative nausea and vomiting. Anesth Analg 2005 Jan, 100(1):87-93.

Study **G4***
Mantovani LG, Monzini MS, Mannucci PM, Scalone L, Villa M, Gringeri A: Differences between patients', physicians' and pharmacists' preferences for treatment products in haemophilia: a discrete choice experiment. Haemophilia 2005 Nov, 11(6):589-597.

Study **G5**
Lewis SM, Cullinane FN, Bishop AJ, Chitty LS, Marteau TM, Halliday JL: A comparison of Australian and UK obstetricians' and midwives' preferences for screening tests for Down syndrome. Prenat Diagn 2006 Jan, 26(1):60-66.

Study **G6**
Berchi C, Dupuis JM, Launoy G: The reasons of general practitioners for promoting colorectal cancer mass screening in France. Eur J Health Econ 2006 Jun, 7(2):91-98.

Study **G7***
Lewis SM, Cullinane FM, Carlin JB, Halliday JL: Women's and health professionals' preferences for prenatal testing for Down syndrome in Australia. Aust N Z J Obstet Gynaecol 2006 Jun, 46(3):205-211.

Study **G8**
Ashcroft DM, Seston E, Griffiths CE: Trade-offs between the benefits and risks of drug treatment for psoriasis: a discrete choice experiment with U.K. dermatologists. Br J Dermatol 2006 Dec, 155(6):1236-1241.

Study **G9**
Caldon LJ, Walters SJ, Ratcliffe J, Reed MW. What influences clinicians' operative preferences for women with breast cancer? An application of the discrete choice experiment. Eur J Cancer. 2007 Jul;43(11):1662-9.

Study **G10**
McGregor JC, Harris AD, Furuno JP, Bradham DD, Perencevich EN. Relative influence of antibiotic therapy attributes on physician choice in treating acute uncomplicated pyelonephritis. Med Decis Making. 2007 Jul-Aug;27(4):387-94.

Study **G11***
Gidman W, Elliott R, Payne K, Meakin GH, Moore J. A comparison of parents and pediatric anesthesiologists' preferences for attributes of child daycase surgery: a discrete choice experiment. Paediatr Anaesth. 2007 Nov;17(11):1043-52.

Study **G12**
Papanikolaou P, Lyne P, Ratcliffe J. Using the discrete choice experimental design to investigate decision-making about pressure ulcer prevention by community nurses. Health Soc Care Community. 2007 Nov;15(6):588-98.

Study **G13**
Danishevski K, McKee M, Sassi F, Maltcev V. The decision to perform Caesarean section in Russia. Int J Qual Health Care. 2008 Apr;20(2):88-94.

Study **G14**
Hitchcock W, Mellon M, Memran M, Parasuraman B, Ramachandran S, Walzer S. Caregiver preferences for pediatric asthma treatment delivery systems. Adv Ther. 2007 Nov-Dec;24(6):1240-53.

Study **G15**
Lee WC, Joshi AV, Woolford S, Sumner M, Brown M, Hadker N, Pashos CL. Physicians' preferences towards coagulation factor concentrates in the treatment of Haemophilia with inhibitors: a discrete choice experiment. Haemophilia. 2008 May;14(3):454-65.

Study **G16**
Szeinbach SL, Harpe SE, Williams PB, Elhefni H. Testing for allergic disease: parameters considered and test value. BMC Fam Pract. 2008 Aug 26;9:47.

Study **G17**
Witt J, Scott A, Osborne RH. Designing choice experiments with many attributes. An application to setting priorities for orthopaedic waiting lists. Health Econ. [Epub 2008 Sep 4]

H. Other

Study **H1**
Bech M. Politicians' and hospital managers' trade-offs in the choice of reimbursement scheme: a discrete choice experiment. Health Policy. 2003 Dec;66(3):261-75.

Bech M. County council politicians' choice of hospital payment scheme: a discrete choice study. Applied Health economics and Health Policy. 2003; 2 (4)225-32.

Study **H2**
Huis in't Veld MH, van Til JA, Ijzerman MJ, Vollenbroek-Hutten MM. Preferences of general practitioners regarding an application running on a personal digital assistant in acute stroke care. J Telemed Telecare. 2005;11 Suppl 1:37-9.

Study **H3**
Thompson CA, Foster A, Cole I, Dowding DW (2005). Using social judgement theory to model nurses' use of clinical information in critical care education. Nurse Education Today. 2005; 25 68-77.

Study **H4**
Arana JE, Leon CJ, Quevedo JL (2006). The effect of medical experience on the economic evaluation of health policies. A discrete choice experiment. Social Science and Medicine. 2006; 63 512-524.

Note: * studies that have more than one main objective.

Part II

Applications of discrete choice experiments in health care

3

Patients' preferences for osteoporosis drug treatment: a discrete choice experiment

de Bekker-Grob EW, Essink-Bot ML, Meerding WJ, Pols HAP, Koes BW, Steyerberg EW

Osteoporosis International 2008; 19(7):1029-1037

Abstract

Introduction Active case finding for osteoporosis is used to identify patients who may benefit from preventive drugs. We aimed to elicit the relative weight that patients place on various aspects of preventive drug treatment for osteoporosis.

Methods We designed a discrete choice experiment, in which women had to choose between drug profiles that differed in five treatment attributes: effectiveness, side effects (nausea), total treatment duration, route of drug administration, and out-of-pocket costs. We included 120 women aged 60 years and older, identified by osteoporosis case finding in 34 general practices in the Netherlands. A conditional logit regression model was used to analyse the relative importance of treatment attributes, the trade-offs that women were willing to make between attributes, and their willingness to pay.

Results All treatment attributes proved to be important for women's choices. A reduction of the relative 10-year risk of hip fracture by 40% or more by the drug was considered to compensate for nausea as a side effect. Women were prepared to pay an out-of-pocket contribution for the currently available drug treatment (bisphosphonate) if the fracture risk reduction was at least 12%.

Conclusions Women identified by active osteoporosis case finding stated to be prepared to take preventive drugs, even if side effects were expected and some out-of-pocket contribution was required.

3.1 Introduction

Osteoporosis constitutes a major public health problem. In the USA, approximately 1.5 million fractures annually are attributable to osteoporosis, including 700,000 vertebral fractures, 250,000 distal forearm (Colles') fractures, 250,000 hip fractures, and 300,000 fractures of other limb sites [1]. Osteoporotic fractures have a major economic impact on society and on the quality of life of patients [2,3]. Preventive drug treatments, such as bisphosphonates (alendronate, etidronate, and risedronate) reduce the risk of osteoporotic fractures in women with postmenopausal osteoporosis by stabilising or increasing the bone density [4–8]. Various practice guidelines recommend a case-finding approach to identify persons with a high risk of osteoporotic fractures [9, 10]. Once identified, women are commonly prescribed preventive drug treatments. However, insight into the relative importance of attributes of preventive drug treatment (such as a bisphosphonate) is limited. Ideally, patients with high fracture risks make an informed decision on taking preventive medication, based on deliberative trade-offs between the burden of medication (e.g., duration, side effects), individual fracture risk, and treatment efficacy.

This study investigated patients' preferences for preventive drug treatment for osteoporosis by means of a discrete choice experiment (DCE), where attributes of hypothetical drug treatments were systematically varied. DCEs have increasingly been used in health care as an approach to elicit patient preferences [11–15]. The DCE was used to determine the trade-offs that community-dwelling elderly women make between the different treatment attributes, and their willingness to pay for each attribute. We also investigated whether high-risk patients (i.e., 10-year risk of a hip fracture greater than 6%) had different preferences than low-risk patients.

3.2 Materials and methods

3.2.1 Study sample and elicitation mode

We recruited community-dwelling elderly women (aged over 60 years) from 34 general practices (in the area of Rotterdam, the Netherlands), who participated in a study on osteoporosis case finding. This latter study used a simple risk score to identify women at high risk of osteoporotic fractures, based on Dutch guidelines [9, 10]. After completion of the risk score, women were asked if they were willing to participate in the current study. Women were informed about their lifetime fracture risk (low or high). We aimed to administer the DCE questionnaire to 120 women with an overrepresentation of women with a high fracture risk (n=60). Earlier studies have shown that this number of respondents is sufficiently large for reliable statistical analyses [15–19]. The DCE questionnaire was sent by post and a trained medical student collected the answers from the respondent by telephone a week later.

3.2.2 DCE

DCEs assume that a given healthcare intervention or treatment can be described by its characteristics (attributes) and that any woman's preferences for an intervention or treatment are determined by the levels of the attributes [15]. Attributes should be identified beforehand as potentially important for the choice of an intervention or treatment [20]. The relative importance of attributes and the trade-offs that women make between them can be assessed when women are offered a series of choices between treatment alternatives that have different combinations of attribute levels [21].

3.2.3 Attributes and attribute levels

The choice of attributes and the attribute levels was based on a literature review focusing on bisphosphonates, expert interviews (n=5; three GPs, the director of the Dutch Osteoporosis Foundation, and a specialist in internal medicine), and personal interviews with 15 community-dwelling women aged over 60 years (i.e., the target group) with and without osteoporosis (n=10 and n=5, respectively).

Table 3.1 Attributes and levels for osteoporosis drug treatment

Attributes and levels	**Beta coefficients in regression analysis**
Route of drug administration:	
Tablet once a month (TABLETmonthly)	
Tablet once a week (TABLETweekly)	β_1
Injection by GP every four months (INJECTIONfourmonths)	β_2
Injection by GP every month (INJECTIONmonthly)	β_3
10-year risk reduction of a hip fracture (%) (EFFECTIVENESS):	β_4
5	
10	
25	
50	
Nausea (during two hours after intake) (NAUSEA):	β_5
No (0)	
Yes (1)	
Total treatment duration (years) (TIME)	β_6
1	
2	
5	
10	
Total cost to you (€) (COST)	β_7
0	
120	
240	
720	

We asked experts and women in the interviews to comment on and complete the list of treatment attributes that was created from literature review. We also asked women to rank the attributes from most important to less important with respect to their preferences for osteoporosis treatment. The number of attributes in a DCE is limited (due to impact on the random component variability) [22], and the ranking results allowed us to make an a priori selection of the most relevant attributes. These were: effectiveness of treatment, side effect of treatment (nausea), total treatment duration, route of drug administration, and costs (Table 3.1). Most of the attribute levels of preventive drug treatment for osteoporosis in our experiment were directly related to bisphosphonates. We also included some hypothetical levels. By including hypothetical levels we can extend the assessment of preferences beyond the currently available treatments for osteoporosis to treatments that are not yet traded in real markets, but may become available in the future. The results are potentially useful to guide the development of new drugs for osteoporosis, because we identified what is important for such a drug to be accepted by the target group. The interviews helped us to determine the hypothetical attribute levels. For example, we determined the levels for the cost-attribute by asking women directly their willingness to pay for preventive drug treatment for osteoporosis.

3.2.4 Study design and questionnaire

The combination of attributes and attribute levels (four attributes with four levels, and one attribute with two levels) resulted in 512 hypothetical drug treatment profiles ($4^4 * 2^1$). For obvious practical reasons, not all of these could be used in a questionnaire. Therefore, we generated a sample of hypothetical drug treatment profiles from all these 512 drug profiles for the questionnaire (i.e., we used a fractional factorial design) [23].

	Treatment A	**Treatment B**	**No treatment**
Route of drug administration	Tablet once a week	Injection by GP every 4 months	Not applicable
10-year risk reduction of a hip fracture	10%	25%	0%
Nausea (during 2 hours after use)	Yes	No	No
Total treatment duration	2 years	5 years	0 years
Total cost to you (thus per month)	€ 0 (€ 0)	€ 120 (€ 2)	€ 0 (€ 0)
Which treatment do you prefer?	□ **A**	□ **B**	□ **None**

Figure 3.1 Example of a choice set as presented in the questionnaire

This sample must be large enough to estimate at least all main effects in a regression analysis. In our case, a sample of 16 hypothetical drug treatment profiles was sufficient [22]. Based on these 16 drug treatment profiles, choice sets were created. Each choice set consisted of two drug treatment profiles and a 'no drug' treatment option; see Figure 3.1 for an example.
The first drug treatment profile (i.e., Treatment A) of each choice set was always one of the 16 hypothetical drug treatment profiles selected for the fractional factorial design. We created the second drug treatment profile (i.e., Treatment B) of each choice set by means of a specific technique ('fold-over') to ensure minimal overlap of attribute levels (i.e., Treatment A and Treatment B always had different attribute levels in each choice set). Too much overlap would reduce the information obtained on trade-offs between attribute levels. Our questionnaire contained 16 choice sets (see Appendix 3.A). We included a dominant choice set in the questionnaire to test for rationality (i.e., a choice set including one drug treatment profile characterized by logically preferable levels on all attributes). The questionnaire started with a detailed written description of each attribute and its levels (the complete questionnaire is available from the authors on request). The questionnaire was pilot tested (n=10) to check for any problems in interpretation and face validity.

3.2.5 Analyses

The DCE was analysed by taking each choice among the three options (two drug treatment profiles, and a 'no drug' treatment option) as an observation. Data from respondents who failed the dominant question were excluded from further analyses. The remaining observations were analysed by a conditional logit regression model. Assuming that all attributes have an independent influence on a woman's preference, the following model was estimated [22]:

$$V = \beta 0 + \beta 1\text{TABLETweekly} + \beta 2\text{INJECTIONfourmonths} + \beta 3\text{INJECTIONmonthly} + \beta 4\text{EFFECTIVENESS} + \beta 5\text{NAUSEA} + \beta 6\text{TIME} + \beta 7\text{COST}$$

where

- V represents the utility derived for preventive osteoporosis drug treatment.
- β0 is a constant reflecting the respondents' preference for receiving osteoporosis drug treatment relative to no osteoporosis drug treatment.

– β1 to β7 are coefficients that indicate the relative importance of each attribute (Table 3.1). β1 to β3 are dummy variables of the attribute 'route of drug administration', with tablet once a month as the base level.

The absolute value of V has no direct interpretation [22]. The sign of a coefficient reflects whether the attribute has a positive or negative effect on utility. The value of a coefficient indicates the relative importance of the corresponding attribute. A statistically significant coefficient was interpreted as indicating that the respondent considered the attribute important. A priori we expected all attributes to be important, and that only the attribute 'effectiveness' would have a positive effect. The utility from the 'no drug' treatment option was normalized to zero. The trade-offs that the respondents were willing to make between the attributes were estimated by the ratios of the coefficients. For example, β5/β6 represents an estimate of how much longer the respondent is willing to take osteoporosis drug treatment (in years) to avoid nausea. The value of coefficient β7 is used to estimate the willingness to pay (WTP). For example, the WTP to avoid nausea was estimated as β5/−β7, where β7 represents the importance of a 100 euro change in price.

We conducted a subgroup analysis by using interaction terms in the conditional logit regression model to assess whether high-risk patients (i.e., 10-year risk of a hip fracture higher than 6%) had different preferences than low-risk patients.

3.3 Results

3.3.1 Respondents

Of the 181 women (76 low and 105 high fracture risk patients) invited, 120 responded (overall response rate 120/181=66%; 60/76 (79%) and 60/105 (57%) for low and high fracture risk patients, respectively). Low and high fracture risk patients did not differ in educational level ($\chi 2$ test, p=0.22), but the high fracture risk patients were older, and more frequently lived without a partner (Table 3.2).

3.3.2 DCE results

Most women indicated that they found the DCE questions (very) clear and had no difficulty in completing the questionnaire. In total, 117 of 120 women (98%) passed the dominant question. All coefficients were significant (Table 3.3).

All signs were consistent with a priori expectations. The positive constant term suggests that respondents preferred drug treatment over 'no drug' treatment if all other attributes were set to zero. The positive sign of the coefficient 'effectiveness' indicates that women preferred a drug treatment with a higher risk reduction of 10-year risk of a hip fracture over a drug treatment with a lower risk reduction. The negative signs for the other coefficients indicate that women preferred a cheaper and shorter drug treatment without nausea. A monthly tablet was preferred to other routes of drug administration (i.e., weekly tablet or injection).

Table 3.2 Respondent characteristics

	All patients (%)		**Low fracture risk patients** (%)		**High fracture risk patients** (%)		**P-value**[a]
Group	120	(100.0)	60	(50.0)	60	(50.0)	
Age (years)							<0.001 *
60-64	28	(23.3)	21	(35.0)	7	(11.7)	
64-69	18	(15.0)	11	(18.3)	7	(11.7)	
70-74	20	(16.7)	17	(28.3)	3	(5.0)	
75-79	28	(23.3)	10	(16.7)	28	(30.0)	
80 and older	26	(21.1)	1	(1.7)	25	(41.7)	
Household							<0.001 *
single	53	(44.2)	17	(28.3)	36	(60.0)	
with partner	67	(55.8)	43	(71.7)	24	(40.0)	
Eductional level							0.215
Low	63	(52.5)	31	(51.7)	32	(53.3)	
Intermediate	46	(38.3)	26	(43.3)	20	(33.3)	
High	11	(9.2)	3	(5.0)	8	(13.3)	

[a] significant difference between low and high fracture risk patient groups
* significant at the 5% level

Table 3.3 Women's preferences for preventive osteoporosis drug treatment

Attribute	**Beta coefficient**	**p Value**	**95% CI**	
Constant (drug treatment)	1.23	<0.001*	0.81	1.66
Drug administration (base level tablet once a month):				
tablet once a week	-0,31	<0.001*	-0.45	-0.17
injection every four months	-0.21	0.027*	-0.41	-0.02
injection once a month	-0.44	<0.001*	-0.64	-0.25
Effectiveness (10% risk reduction)	0.28	<0.001*	0.23	0.34
Side effect nausea	-1.10	<0.001*	-1.30	-0.89
Treatment duration (1 year)	-0.04	<0.001*	-0.06	-0.02
Cost (€100)	-0.15	<0.001*	-0.18	-0.11

* significant at the 5% level

Number of observations 5,589 (117 respondents x 16 choices x 3 options per choice, minus 27 missing values), Pseudo R^2 = 0.1847, log pseudolikelihood = -1668.7

The magnitude of the attribute coefficients corresponds with the relative importance of the attributes. For a correct interpretation of the comparison of the coefficients of the attributes, we need to pay attention to the different units of measurement. For example, the coefficient of 0.28 of 'effectiveness' implies the increase in utility per 10% of risk

reduction. A risk reduction of 40% is four times larger than 10% risk reduction. Thus, a risk reduction of 40% implies a utility of 1.10 for the attribute 'effectiveness' (i.e., four multiplied with the coefficient of 0.28 of 'effectiveness'). A superficial comparison of the coefficients in Table 3.3 may easily lead us to the wrong conclusion that nausea as a side effect had a larger influence on women's choices for preventive drug treatment than the attribute 'effectiveness'. As shown in the computation above, a risk reduction of more than 40% contributes more to the utility of a preventive drug treatment for osteoporosis than absence of nausea. A 40% risk reduction (i.e., utility 1.10) compensates for the disutility of minus 1.10 of the attribute 'nausea'.

A positive utility value of a specific drug profile indicates a preference for that treatment to no treatment. The utility of a currently most frequently used preventive osteoporosis drug treatment (bisphosphonate taken as a weekly tablet, approximately 30% fracture risk reduction, nausea as a possible side effect, total treatment duration of five years, and no out-of-pocket payment) equals

V = 1.23 – 0.31(TABLETweekly) + 3.0 * 0.28(EFFECTIVENESS) – 1.10(NAUSEA) – 5 * 0.04(TIME) - 0 * 0.15(COST) = 0.46.

This outcome has a positive sign. Thus, the women in our sample preferred this drug treatment over no treatment.

3.3.3 Trade-offs

Based on the expressed preferences, women were prepared to adhere to drug treatment an estimated 5.7 years longer to change from an injection every 4 months to a tablet once a month, if all other attributes remained constant (Table 3.4). For every 10% additional fracture risk reduction, they were prepared to adhere to drug treatment 7.5 years longer. Respondents were willing to pay an extra total amount of €752 to avoid nausea as a side effect, or €26 for every 1-year decrease in total drug treatment duration.

For bisphosphonates, we estimated that respondents were willing to pay up to an estimated 338 euro out-of-pocket payment to receive treatment compared with no treatment (WTP = 847(constant) − 212(weekly tablet) + 3.0*195(risk reduction) – 752(side effect nausea) – 5*26(treatment duration)). They would thus be willing to pay for this treatment if the fracture risk reduction was at least 12 % (see Appendix 3.B).

3.3.4 High versus low fracture risk patients

The results of conditional logit regression modelling of data from both risk groups are presented in Table 3.5. Only the interaction between the effectiveness of treatment and risk group was significant (p=0.05). Lower levels of effectiveness of the preventive drug treatment for osteoporosis were more acceptable to high-risk patients than to low-risk patients. High-risk patients therefore accepted a less effective drug to reduce their fracture risk.

Table 3.4 Women's time and monetary trade-offs for preventive osteoporosis drug treatment

Attribute	Willingness to adhere to the drug treatment longer (years)	WTP for total treatment (€)	Interpretation note
Constant (no drug treatment)	32.7	847	For drug treatment vs no drug treatment
Drug administration (base level tablet once a month):			
tablet once a week	8.2	212	For change from tablet once a week to tablet once a month
injection every four months	5.7	147	For change from injection every 4 months to tablet once a month
injection once a month	11.7	304	For change from injection once a month to tablet once a month
Effectiveness	7.5	195	For 10% reduction in 10-year risk of a hip fracture
Side effect nausea	29.0	752	For change from side effect to no side effect
Treatment duration		26	For 1-year decrease in total drug treatment duration
Cost	3.9		For €100 decrease in drug treatment cost

Table 3.5 Differences between low and high-risk patients' preferences for preventive osteoporosis drug treatment

Attribute	Beta coefficient of low risk patients	Beta coefficient of high risk patients	p-Value
Constant (drug treatment)	1.178	1.316	0.748
Drug administration (base level tablet once a month):			
tablet once a week	-0.360	-0.255	0.464
injection every 4 months	-0.125	-0.317	0.323
injection once a month	-0.445	-0.454	0.966
Effectiveness (1% risk reduction)	0.023	0.034	0.050 *
Side effect nausea	-1.046	-1.161	0.582
Treatment duration (1 year)	-0.033	-0.045	0.583
Cost (€1)	-0.002	-0.001	0.435

*significant at the 5% level

Number of observations 5,589 (117 patients (i.e. 58 low-risk + 59 high-risk patients) x 16 choices x 3 options per choice, minus 27 missing values), Pseudo R^2 = 0.1895, log pseudolikelihood = -1658.8

3.4 Discussion

Women identified by active case finding for osteoporosis said in this DCE that they were prepared to adhere to preventive drug treatment. Treatment effectiveness (hip fracture risk reduction), side effects (nausea), total treatment duration, route of drug administration, and out-of-pocket costs were all relevant to women's preferences for drug treatment. Nausea as a side effect had a large influence on women's choices, though a risk reduction of 40% or more was sufficient to make nausea as a side effect acceptable. Patients with high fracture risk were more prepared to take a less effective drug treatment than low fracture risk patients.

This study illustrated the feasibility of DCE to elicit elderly women's preferences for osteoporosis drug treatment. An acceptable fraction of potential respondents agreed to participate in the experiment, and only 3 of 120 failed the dominant question. This study therefore adds to the available literature on the usefulness of DCE to investigate preferences for drug treatment [15, 24–26].

In a previous DCE investigating women's preferences for osteoporosis treatment, early postmenopausal women were willing to use a tablet once a week if the drug reduced the absolute lifetime risk of fracture by at least 10% [27]. In our study, we considered relative reductions in 10-year risk of a hip fracture, which limits the comparability between these two DCEs. According to Fraenkel et al., women willing to consider a drug treatment strongly preferred taking a tablet once a week rather than having an injection in a doctor's office. We also found that women preferred a tablet once a week to injection once a

month, but that an injection in a doctor's office every 4 months was preferred to having to take a tablet once a week. This was in line with findings of another study which showed that many women preferred annual injections to weekly oral medication [28].

The women in our study showed a very positive attitude towards preventive drug treatment for osteoporosis, and said they were prepared to take preventive drug treatment even when the effectiveness of the treatment was zero. This may reflect a kind of placebo effect for drug treatment. In practice, a substantial proportion of women discontinue treatment with bisphosphonates [29, 30]. Various patient characteristics (e.g., being retired) were associated with a high compliance with treatment for osteoporosis [31]. Further research is needed to investigate why patients, who are prepared to start with bisphosphonates, discontinue treatment.

The effectiveness of the preventive osteoporosis drug treatment was less important for high-risk patients than for low-risk patients. All patients had knowledge of lifetime fracture risk (low or high). Patients at high risk were probably more aware of the consequences of their fracture risk than low-risk patients, and were therefore more prepared to take a less effective drug treatment.

In our study we used a postal questionnaire and the answers of the respondent were collected later by telephone. The use of an interviewer can be regarded as a strength in the design, because this procedure led to data completeness as well as a check of a respondent's understanding. However, this design also had some limitations. First, although we included nausea as the most relevant gastrointestinal side effect, other gastrointestinal side effects that are also common (e.g., stomach pain, heartburn) were not included in the DCE. More in general, we selected the most relevant attributes in our DCE using interviews, but this careful procedure does not guarantee that attributes that we did not include are irrelevant to women's preferences for osteoporosis treatment. Second, we studied main effects only, since these generally account for 70% to 90% of explained variance in a DCE [22]; specific combinations of attribute levels may have specific effects that remained unidentified. Third, the current results could be validated by comparison with actual behaviour of women in drug treatment for osteoporosis. Fourth, the respondents were women selected by active case finding, which precludes generalization of the findings to all women. On the other hand, women identified by active osteoporosis case finding are those who will have to decide about treatment and therefore constitute the most relevant study group.

Patient-centred and demand-led care is becomingly increasingly important in current medical practice. Understanding which and how drug treatment attributes influence women's preferences for osteoporosis drug treatment is important to optimize the treatment design that patients will follow. The present study showed that the target group may well accept the currently available bisphosphonates with sufficient margin (satisfactory effectiveness, side effects, and so on); this is an important result for policy decision-making on the introduction of active case finding on a large scale, in addition to considerations related to cost-effectiveness.

In conclusion, this DCE showed that women identified by active osteoporosis case finding showed a positive attitude to preventive drug treatment, even if side effects (such as nausea) were expected and some out-of-pocket contribution was required.

References

1. Riggs BL, Melton LJ 3rd (1995). The worldwide problem of osteoporosis: insights afforded by epidemiology. Bone 17 (5 Suppl):505S–511S.
2. Silverman SL (2005). Quality-of-life issues in osteoporosis. Curr Rheumatol Rep 7(1):39–45
3. Atik OS, Gunal I, Korkusuz F (2006). Burden of osteoporosis. Clin Orthop Relat Res 443:19 24.
4. Ensrud KE, Black DM, Palermo L, Bauer DC, Barrett-Connor E, Quandt SA, Thompson DE, Karpf DB (1997). Treatment with alendronate prevents fractures in women at highest risk: results from the Fracture Intervention Trial. Arch Intern Med 157(22):2617–2624.
5. Pols HA, Felsenberg D, Hanley DA, Stepan J, Munoz-Torres M, Wilkin TJ, Qin-sheng G, Galich AM, Vandormael K, Yates AJ et al (1999). Multinational, placebo-controlled, randomized trial of the effects of alendronate on bone density and fracture risk in postmenopausal women with low bone mass: results of the FOSIT study. Fosamax International Trial Study Group. Osteoporos Int 9 (5):461–468.
6. Black DM, Cummings SR, Karpf DB, Cauley JA, Thompson DE, Nevitt MC, Bauer DC, Genant HK, Haskell WL, Marcus R et al (1996). Randomised trial of effect of alendronate on risk of fracture in women with existing vertebral fractures. Fracture Intervention Trial Research Group. Lancet 348(9041):1535– 1541.
7. Meunier PJ (1999). Evidence-based medicine and osteoporosis: a comparison of fracture risk reduction data from osteoporosis randomised clinical trials. Int J Clin Pract 53(2):122–129.
8. Harris ST, Watts NB, Genant HK, McKeever CD, Hangartner T, Keller M, Chesnut CH 3rd, Brown J, Eriksen EF, Hoseyni MS et al (1999). Effects of risedronate treatment on vertebral and nonvertebral fractures in women with postmenopausal osteoporosis: a randomized controlled trial. Vertebral Efficacy with Risedronate Therapy (VERT) Study Group. JAMA 282(14):1344–1352.
9. Guideline on osteoporosis. Utrecht: The Dutch Institute for Healthcare Improvement (CBO); 2002.
10. Kanis JA, Delmas P, Burckhardt P, Cooper C, Torgerson D (1997). Guidelines for diagnosis and management of osteoporosis. The European Foundation for Osteoporosis and Bone Disease. Osteoporos Int 7(4):390–406.
11. Gyrd-Hansen D, Sogaard J (2001). Analysing public preferences for cancer screening programmes. Health Econ 10(7):617–634.
12. Ryan M, Hughes J (1997). Using conjoint analysis to assess women's preferences for miscarriage management. Health Econ 6 (3):261–273.
13. Ryan M, Farrar S (2000). Using conjoint analysis to elicit preferences for health care. BMJ 320(7248):1530–1533.
14. Sculpher M, Bryan S, Fry P, de Winter P, Payne H, Emberton M (2004). Patients' preferences for the management of non-metastatic prostate cancer: discrete choice experiment. BMJ 328(7436):382.
15. Watson V, Ryan M, Brown CT, Barnett G, Ellis BW, Emberton M (2004). Eliciting preferences for drug treatment of lower urinary tract symptoms associated with benign prostatic hyperplasia. J Urol 172(6 Pt 1):2321–2325.
16. Aristides M, Weston AR, FitzGerald P, Le Reun C, Maniadakis N (2004). Patient preference and willingness-to-pay for Humalog Mix25 relative to Humulin 30/70: a multicountry application of a discrete choice experiment. Value Health 7(4):442–454.

17. Weston A, Fitzgerald P (2004). Discrete choice experiment to derive willingness to pay for methyl aminolevulinate photodynamic therapy versus simple excision surgery in basal cell carcinoma. Pharmacoeconomics 22(18):1195–1208.
18. Lee A, Gin T, Lau AS, Ng FF (2005). A comparison of patients' and health care professionals' preferences for symptoms during immediate postoperative recovery and the management of postoperative nausea and vomiting. Anesth Analg 100(1):87–93.
19. Bishai D, Brice R, Girod I, Saleh A, Ehreth J (2007). Conjoint analysis of French and German parents' willingness to pay for meningococcal vaccine. Pharmacoeconomics 25(2):143–154.
20. Farrar S, Ryan M, Ross D, Ludbrook A (2000). Using discrete choice modelling in priority setting: an application to clinical service developments. Soc Sci Med 50(1):63–75.
21. Ryan M, Scott DA, Reeves C, Bate A, van Teijlingen ER, Russell EM, Napper M, Robb CM (2001). Eliciting public preferences for healthcare: a systematic review of techniques. Health Technol Assess 5(5):1–186.
22. Louviere JJ, Hensher DA, Swait JD (2000). Stated choice methods: analysis and application. Cambridge University Press, Cambridge.
23. Hahn GJ, Shapiro SS (1966). A catalogue and computer program for the design and analysis of orthogonal symmetric and asymmetric fractional factorial experiments. General Electric Research and Development Centre, Schenectady, NY, USA.
24. Lloyd A, McIntosh E, Price M (2005). The importance of drug adverse effects compared with seizure control for people with epilepsy: a discrete choice experiment. Pharmacoeconomics 23 (11):1167–1181.
25. Mahadevia P, Shah S, Mannix S, Brewster-Jordan J, Kleinman L, Liebman C, O'Dowd L (2006). Willingness to pay for sensory attributes of intranasal corticosteroids among patients with allergic rhinitis. J Manag Care Pharm 12(2):143–151.
26. Lancsar EJ, Hall JP, King M, Kenny P, Louviere JJ, Fiebig DG, Hossain I, Thien FC, Reddel HK, Jenkins CR (2007). Using discrete choice experiments to investigate subject preferences for preventive asthma medication. Respirology 12(1):127–136.
27. Fraenkel L, Constantinescu F, Oberto-Medina M, Wittink DR (2005). Women's preferences for prevention of bone loss. J Rheumatol 32(6):1086–1092.
28. Fraenkel L, Gulanski B, Wittink D (2006). Patient treatment preferences for osteoporosis. Arthritis Rheum 55(5):729–735.
29. Bartl R, Gotte S, Hadji P, Hammerschmidt T (2006). [Adherence with daily and weekly administration of oral bisphosphonates for osteoporosis treatment] Adharenz mit taglichen und wochentlichen oralen Bisphosphonaten in der Osteoporosetherapie. Dtsch Med Wochenschr 131(22):1257–1262.
30. Caro JJ, Ishak KJ, Huybrechts KF, Raggio G, Naujoks C (2004). The impact of compliance with osteoporosis therapy on fracture rates in actual practice. Osteoporos Int 15(12):1003–1008.
31. Blotman F, Cortet B, Hilliquin P, Avouac B, Allaert FA, Pouchain D, Gaudin AF, Cotte FE, El Hasnaoui A (2007). Characterisation of patients with postmenopausal osteoporosis in French primaryhealthcare. Drugs Aging 24(7):603–614.

Appendix 3.A

Overview of the 16 choice sets used in our questionnaire.

	Treatment A					Treatment B				
	Attributes					Attributes				
Choice set	I	II	III	IV	V	I	II	III	IV	V
1	b	a	a	d	b	c	b	b	a	c
2	d	a	b	c	d	a	b	a	d	a
3	a	c	b	d	c	b	d	a	a	d
4	c	c	a	c	a	d	d	b	d	b
5	a	b	b	c	b	b	c	a	d	c
6	a	d	a	b	d	b	a	b	c	a
7	c	a	b	b	c	d	b	a	c	d
8	d	c	a	b	b	a	d	b	c	c
9	b	d	a	c	c	c	a	b	d	d
10	d	b	a	a	c	a	c	b	b	d
11	a	a	a	a	a	b	b	b	b	b
12	b	c	b	a	d	c	d	a	b	a
13	b	b	b	b	a	c	c	a	c	b
14	d	d	b	d	a	a	a	a	a	b
15	c	d	b	a	b	d	a	a	b	c
16	c	b	a	d	d	d	c	b	a	a

Code book:
I Route of drug administration
- a tablet once a month
- b tablet once a week
- c injection by GP every 4 months
- d injection by GP every month

II Effectiveness
- a 5% risk reduction of a hip fracture
- b 10% risk reduction of a hip fracture
- c 25% risk reduction of a hip fracture
- d 50% risk reduction of a hip fracture

III Side effects (nausea)
- a no nausea
- b nausea for up to two hours after intake

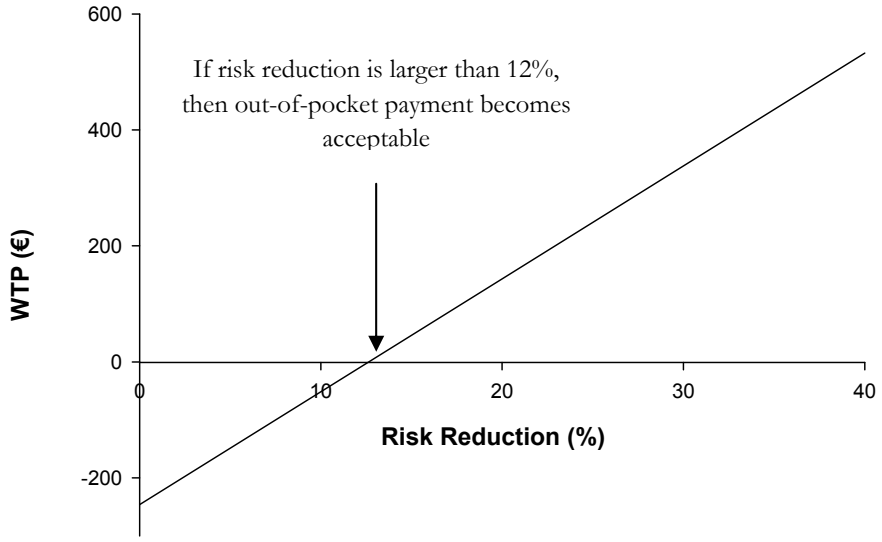
IV Total treatment duration
- a 1 year
- b 2 years
- c 5 years
- d 10 years

V Out-of-pocket costs
- a € 0
- b € 120
- c € 240
- d € 720

Treatment A of choice set 1 (i.e., code 'baadb') represents a hypothetical preventive drug treatment for osteoporosis with the following characteristics: tablet once a week as a route of drug administration, 5% risk reduction of a hip fracture as treatment effectiveness, no side effects (nausea), total treatment duration of 10 years, and 120 euro out-of-pocket costs for the entire treatment. In the questionnaire this treatment was presented with Treatment B of choice set 1 (i.e., code 'cbbac') with the following characteristics: injection by GP every 4 months as a route of drug administration, 10% risk reduction of a hip fracture as treatment effectiveness, nausea for up to two hours after intake as a side effect, total treatment duration of 1 year, and 240 euro out-of-pocket costs for the entire treatment. Respondents had to choose between this Treatment A, Treatment B, or no treatment.

Appendix 3.B

Regression line showing the relationship between risk reduction (%) by means of bisphosphonate (tablet once a week, side effect nausea, and treatment duration five years) and WTP (€).



4

Preferences of GPs and patients for preventive osteoporosis drug treatment: a discrete choice experiment

de Bekker-Grob EW, Essink-Bot ML, Meerding WJ, Koes BW, Steyerberg EW

Pharmacoeconomics 2009; 27(3):211-219

Abstract

Background Osteoporotic fractures have a serious economic impact on society and on the quality of life of patients. Differences in opinions on the desirability of preventive treatment initiation may hamper the process and outcome of shared decision making between physician and patient.

Objective To evaluate and compare preferences of GPs and patients for preventive osteoporosis drug treatment.

Methods Discrete-choice experiment (DCE) involving 34 general practices in the area of Rotterdam, the Netherlands. Participants included 40 GPs and 120 women aged >60 years who participated in a study on osteoporosis case finding. We included any woman aged >60 years, with an over-representation of women with a high fracture risk (n=60).

Outcomes (i) The relative weights that GPs and patients place on five treatment attributes of preventive osteoporosis drug treatment: effectiveness, nausea as an adverse effect, total treatment duration, route of drug administration and out-of-pocket costs; and (ii) the determinants of any differences in preferences between GPs and patients.

Results The response rate was 40/59 (68%) for GPs and 120/181 (66%) for patients. All treatment attributes proved to be important for preferences of GPs and patients. GPs had a significantly less favourable attitude towards preventive osteoporosis drug treatment than patients; they placed significantly higher values on effectiveness of preventive drug treatment and short total preventive treatment duration than patients.

Conclusions GPs and patients showed different preferences towards preventive osteoporosis drug treatment. Addressing each of these differences may have a positive effect on the process and outcomes of shared decision making regarding treatment initiation.

4.1 Background

Health professionals and patients may or may not share the same preferences related to the patient's treatment. If preferences of health professionals and patients are polarized, decision making on patients' treatment may be more difficult to achieve in a model of shared medical decision making [1]. Differences in opinions on the desirability of treatment may hamper the process and outcome of shared decision making between physician and patient.

Osteoporosis is a disease characterized by low bone mass, microarchitectural deterioration of bone tissue leading to enhanced bone fragility and a consequent increase in fracture risk [2]. Osteoporosis is a major public health problem. In the USA, approximately 1.5 million fractures annually (including vertebral, distal forearm, hip and other limb sites) are attributable to osteoporosis [3]. In the Netherlands, the situation is similar, with about 75 000 fractures attributable to osteoporosis annually [4]. Osteoporotic fractures have a serious economic impact on society and on the quality of life of patients [5, 6]. Preventive osteoporosis drug treatment, such as bisphosphonates, may reduce the risk of osteoporotic fractures [7-10].

The aim of this study was to evaluate and compare preferences of GPs and patients for preventive osteoporosis drug treatment by means of discrete-choice experiments (DCEs), and to investigate determinants of the differences found. DCEs are increasingly used in healthcare to elicit preferences [11-14]; however, only a few DCEs have directly compared preferences of health professionals and patients [15-18]. DCEs assume that a given healthcare intervention or treatment can be described by its characteristics (attributes) and that any respondent's preferences for an intervention or treatment are determined by the levels of the attributes [14]. Attributes are identified beforehand as potentially important for the choice of an intervention or treatment [21]. The relative importance of attributes and the trade-offs that respondents make between them can be assessed when respondents are offered a series of choices between treatment alternatives that have different combinations of attribute levels [22].

4.2 Methods

4.2.1 Study population

We included GPs and women aged >60 years from 34 general practices (in the area of Rotterdam) who participated in a study on osteoporosis case finding. An over-representation of women with a high fracture risk (n=60), identified using a simple risk score, based on (inter)national guidelines [4, 19], was included in the study sample. More details of the recruitment of women for the original study were described previously [20].

4.2.2 Study format

We followed the recommended five stages of undertaking a DCE [12]. First, we identified the attributes (characteristics) of osteoporosis drug treatment by means of a literature review, expert interviews (Dutch Osteoporosis Foundation, a specialist in internal medicine, and three GPs), patient interviews (n=15) and a study in community-dwelling women aged ≥60 years (i.e. target group) with and without osteoporosis (n=10 and n=5, respectively). Five treatment attributes were selected: effectiveness of treatment (reduction of risk of hip fracture), nausea as an adverse effect of treatment, total treatment duration, route of drug administration and costs.
Second, we assigned levels to these attributes, using the same sources of information (Table 4.1). In our experiment, most of the attribute levels of preventive drug-treatment for osteoporosis were directly related to bisphosphonates. Additional inclusion of hypothetical attribute levels enables extension to the assessment of preferences beyond the currently available treatments for osteoporosis to treatments that are not yet traded in real markets, although may become available in the future.

Table 4.1 Attributes and levels for preventive osteoporosis drug treatment

Attributes and levels	**Beta coefficients in regression analysis**
Route of drug administration:	
Tablet once a month (TABLETmonthly)	
Tablet once a week (TABLETweekly)	β_1
Injection by GP every four months (INJECTIONfourmonths)	β_2
Injection by GP every month (INJECTIONmonthly)	β_3
10-year risk reduction of a hip fracture (%) (EFFECTIVENESS):	β_4
5	
10	
25	
50	
Nausea (during two hours after intake) (NAUSEA):	β_5
No (0)	
Yes (1)	
Total treatment duration (years) (TIME)	β_6
1	
2	
5	
10	
Total cost to you (€) (COST)	β_7
0	
120	
240	
720	

Third, hypothetical drug-treatment profiles were selected for the questionnaire. The combination of attributes and attribute levels (four attributes with four levels, and one attribute with two levels) resulted in 512 hypothetical drug-treatment profiles ($4^4 * 2^1$). For obvious practical reasons, not all of these could be used in a questionnaire. Therefore, we used a fractional factorial design (i.e. we generated a sample of hypothetical drug treatment profiles from all these 512 drug profiles for the questionnaire) [23]. This sample must be large enough to estimate at least all main effects in a regression analysis. Our fractional factorial design included a sample of 16 hypothetical drug treatment profiles, to allow for estimation of all main effects [24, 25]. These drug-treatment profiles fulfilled the demands of level balance (i.e. all attribute levels occurred with equal frequency) and orthogonality (i.e. levels were independent of each other).
Fourth, choice sets were created. Each choice set consisted of two drug-treatment profiles and a 'no drug' treatment option (see Figure 4.1). The first drug-treatment profile (i.e. treatment A) of each choice set was always one of the 16 hypothetical drug-treatment profiles of the fractional factorial design. We created the second drug-treatment profile (i.e. treatment B) of each choice set by means of cycle 'fold-over' technique to ensure minimal overlap of attribute levels (i.e. treatment A and treatment B always had different attribute levels in each choice set). Too much overlap would reduce the information obtained on trade-offs between attribute levels. According to the results of an analysis using the software of Street and Burgess (Department of Mathematical Sciences, University of Technology, Sydney, NSW, Australia)[30] our choice set design had an efficiency of 95% compared with an optimal choice set design. All main effects were uncorrelated.

	Treatment A	**Treatment B**	**No treatment**
Route of drug administration	Tablet once a week	Injection by GP every 4 months	Not applicable
10-year risk reduction of a hip fracture	10%	25%	0%
Nausea (during 2 hours after use)	Yes	No	No
Total treatment duration	2 years	5 years	0 years
Total cost to you (thus per month)	€ 0 (€ 0)	€ 120 (€ 2)	€ 0 (€ 0)
Which treatment do you prefer?	□ **A**	□ **B**	□ **None**

Figure 4.1 Example of a choice set as presented in the questionnaire

Our design used a series of choice sets fulfilling the demands of level balance, orthogonality and minimal overlap (i.e. repeated occurrence of an attribute level in a choice set was kept at minimum). We included a dominant choice set in the questionnaire to test for rationality (i.e. a choice set including one drug-treatment profile characterized by logically preferable levels on all attributes). The dominant choice set was an extra choice set (i.e. 17 choice sets were created).
The questionnaire started with a detailed written description of each attribute and its levels. In particular, the attribute 'effectiveness' was explained with examples, because it is known from the literature that respondents may have difficulties understanding numerical risks. The questionnaire was pilot tested on GPs and patients (n=2 and n=8, respectively) to check for any problems in interpretation and face validity. None of the respondents raised any problems with understanding the questionnaire, so the pilot test did not result in any changes to the questionnaire. The questionnaire was sent to all respondents by post. GPs and patients received the same choice sets. Patients were asked to express their preferences as if it was their own treatment; GPs expressed in each choice set which treatment option they preferred for a female patient aged >60 years with a high fracture risk (10-year risk of a hip fracture higher than 6%) from their viewpoint as a physician.
A trained medical student collected the answers from the patient by telephone 1 week later. The GPs returned the self-completed questionnaire by post.

4.2.3 Analyses

We analysed the DCE by taking each choice (excluding the dominant choice set) among the three possibilities (two drug-treatment profiles and a 'no drug' treatment option) as an observation. Data from respondents who failed the dominant question were excluded from further analyses. Conditional logit regression models (i.e. multinomial models) were used to investigate the effects of the treatment attributes on GPs' and patients' acceptance of preventive osteoporosis drug treatment. These models were implemented in STATA software (Version 8.0, College Station, TX, USA). Assuming that all attributes have an independent influence on a GP's or woman's preference, the following model was estimated (equation 1):

$$V = \beta_0 + \beta_1 \text{TABLETweekly} + \beta_2 \text{INJECTIONfour months} + \beta_3 \text{INJECTIONmonthly} + \beta_4 \text{EFFECTIVENESS} + \beta_5 \text{NAUSEA} + \beta_6 \text{TIME} + \beta_7 \text{COST} \quad \text{(Eq. 1)}$$

where V represents the utility derived for preventive osteoporosis drug treatment, as derived from the observed choice behaviour; β_0 is a constant reflecting the respondents' preference for receiving osteoporosis drug treatment relative to no osteoporosis drug treatment; β_1 to β_7 are coefficients that indicate the relative importance of each attribute

(Table 4.1). β_1 to β_3 are variables of the attribute 'route of drug administration', with 'Tablet once a month' as the base level (Table 4.1).
The sign of a coefficient reflects whether the attribute has a positive or negative effect on utility. The absolute value of these utility scores has a relative interpretation: the higher the utility score, the stronger the respondent's preference for a particular preventive osteoporosis drug-treatment profile. The value of a coefficient indicates the relative importance of the corresponding attribute. Preferences (sign, significance of coefficient) towards a single attribute need to be interpreted under the condition of 'ceteris paribus', i.e. by keeping everything else equal. The results of the patients' preferences were described previously [20].
The differences between GPs' and patients' preferences were estimated by including statistical interaction terms in the model.
Using the coefficients, we generated the ranks and total utility scores assigned by GPs and patients to the 16 hypothetical drug-treatment profiles for preventive osteoporosis treatment.

4.3 Results

Between October 2006 and June 2007, we invited 59 GPs and 181 patients to take part in the study. The response rate was 40/59 (68%) for GPs and 120/181 (66%) for patients. GPs and patients differed in age, sex and educational level ($p \leq 0.001$; Table 4.2).

Table 4.2 Respondents characteristics

Characteristics	**GPs** (n = 40)	**Patients** (n = 120)	**p-value**[a]
Age, yr (mean ± sd)	49.3 ± 7.9	71.8 ± 7.9	<0.001
Sex (%)			<0.001
male	80	0	
female	20	100	
Household			
single	NA	53	
with partner	NA	67	
Eductional level (%)			<0.001
low	0	53	
intermediate	0	38	
high	100	9	

[a] difference between general practitioners and patients
NA = not available

4.3.1 Discrete choice experiment results

Of 40 GPs, 39 (98%) passed the dominant question. All treatment attribute coefficients had p-values <0.05 (Table 4.3), i.e. all attributes were significant determinants of GPs' choices. However, with respect to the route of drug administration, the attribute levels 'tablet once a week' and 'injection once every 4 months' did not differ from the attribute level 'tablet once a month'. Thus, GPs valued the routes of drug administration equally, with the exception of 'an injection once a month', which was less preferred than a 'tablet once a month'. The negative constant term suggests that GPs preferred 'no drug' treatment for the patient over drug treatment if all other attributes were set to zero (i.e. 10-year risk reduction of a hip fracture of 0%, tablet once a month, no nausea as an adverse effect, very short treatment duration and no out-of-pocket payment). According to the positive sign of the coefficient for the attribute 'effectiveness', GPs preferred a more effective treatment than a less effective treatment as would be expected. According to the negative signs of the coefficients, GPs preferred a cheaper and shorter drug treatment without nausea as an adverse effect over a more expensive and longer drug treatment with nausea as an adverse effect.

Of 120 patients, 117 (98%) passed the dominant question. All coefficients were significant at the 5% level (Table 4.3). Patients preferred drug treatment over 'no drug' treatment if all other attributes were set to zero (positive sign of the constant term). The positive sign of the coefficient 'effectiveness' indicates that, as expected, patients preferred a drug treatment with a higher risk reduction of 10-year risk of a hip fracture over a drug treatment with a lower risk reduction. The negative signs for the other coefficients indicate that patients preferred a cheaper and shorter drug treatment without nausea over a more expensive and longer drug treatment with nausea as an adverse effect. A monthly tablet was preferred to other routes of drug administration (i.e. weekly tablet or injection).

The opposite signs of the constant terms indicate that GPs and patients had opposite preferences for preventive osteoporosis drug treatment. The distinction between the more reserved (rather negative) attitude towards preventive osteoporosis drug treatment of GPs and the more positive attitude of patients is further corroborated by the results shown in Table 4.3. GPs had a higher coefficient value for 'effectiveness' of drug treatment and a more negative value for 'treatment duration', suggesting that GPs demand a higher effectiveness and a shorter treatment duration than patients for a treatment to become acceptable. GPs did not prefer a monthly tablet to a weekly tablet as a route of drug administration, but patients did.

Ranking of the 16 hypothetical preventive treatment scenarios by their relative utilities as derived from the choice experiment showed that GPs preferred only five of them to 'no drug' treatment (i.e. positive utility score; Table 4.4), whereas patients preferred 12 of 16 hypothetical treatments over 'no drug' treatment.

Table 4.3 Preferences of GPs and patients for preventieve osteoporosis drug treatment

Attribute	Preferences of GPs				Preferences of patients				GPs vs patients' preferences
	coefficient[a]	p-value	95% CI		coefficient[b]	p-value	95% CI		p-value of the interaction[c]
Constant (drug treatment)	-1.49	<0.001	-2.08	-0.91	1.23	<0.001	0.81	1.66	<0.001
Drug administration (base level tablet once a month):									
tablet once a week	0.24	0.223	-0.15	0.63	-0.31	<0.001	-0.45	-0.17	<0.008
injection every 4 months	0.05	0.755	-0.27	0.37	-0.21	0.027	-0.41	-0.02	0.161
injection once a month	-0.78	<0.001	-1.11	-0.44	-0.44	<0.001	-0.64	-0.25	0.093
Effectiveness (10% risk reduction)	1.00	<0.001	0.81	1.18	0.28	<0.001	0.23	0.34	<0.001
Adverse effect nausea	-0.86	<0.001	-1.28	-0.44	-1.10	<0.001	-1.30	-0.89	0.323
Treatment duration (1 year)	-0.09	<0.001	-0.14	-0.04	-0.04	<0.001	-0.06	-0.02	<0.001
Cost (€100)	-0.10	0.006	-0.17	-0.03	-0.15	<0.001	-0.18	-0.11	0.228

[a] Number of observations 1,863 (39 GPs x 16 choices x 3 options per choice, minus 9 missing values)
[b] Number of observations 5,589 (117 patients x 16 choices x 3 options per choice, minus 27 missing values)
[c] Number of observations 7,452 (156 (i.e. 39 GPs + 117 patients) respondents x 16 choices x 3 options per choice, minus 36 missing values)

Table 4.4 GPs' and patients' ranks and utility scores of hypothetical preventive osteoporosis drug treatment profiles

Drug profile from fractional factorial design (administration route / effectiveness / adverse effect nausea / treatment duration / total costs)		**Ranks and utility scores**			
		GPs		**patients**	
Profile nr	Description	rank	utility[a]	rank	utility[b]
9	weekly tablet / 50% effect / no nausea / 5yrs / €240	1	3.03	1	1.81
6	monthly tablet / 50% effect / no nausea / 2yrs / €720	2	2.59	3	1.53
15	injection 4 months / 50% effect / nausea / 1yr / €120	3	2.47	6	1.13
14	monthly injection / 50% effect / nausea / 10yrs / €0	4	0.94	7	0.74
4	injection 4 months / 25% effect / no nausea / 5 yrs / €0	5	0.59	2	1.54
8	monthly injection / 25% effect / no nausea / 2yrs / €120	6	-0.08	5	1.25
12	weekly tablet / 25% effect / nausea / 1yr / €720	7	-0.43	15	-0.55
3	monthly tablet / 25% effect / nausea / 10yrs / €240	8	-1.02	10	0.12
11	monthly tablet / 5% effect / no nausea / 1yr / €0	9	-1.09	4	1.34
13	weekly tablet / 10% effect / nausea / 2yrs / €0	10	-1.30	12	0.04
10	monthly injection / 10% effect / no nausea / 1yr / €240	11	-1.60	8	0.69
1	weekly tablet / 5% effect / no nausea / 10yrs / €120	12	-1.79	9	0.52
5	monthly tablet / 10% effect / nausea / 5yrs / €120	13	-1.93	11	0.06
16	injection 4 months / 10% effect / no nausea / 10yrs / €720	14	-2.07	13	-0.12
7	injection 4 months / 5% effect / nausea / 2yrs / €240	15	-2.23	14	-0.36
2	monthly injection / 5% effect / nausea / 5yrs / €720	16	-3.80	16	-1.40

a General practitioners' utility = -1.49 + (0.24 * tablet weekly) + (0.05 * injection four months) - (0.78 * injection monthly) + (1.0 * effectiveness / 10) - (0.86 * nausea) - (0.09 * time) - (0.1 * cost / 100)

b Patients' utility = 1.23 - (0.31 * tablet weekly) - (0.21 * injection four months) - (0.44 * injection monthly) + (0.28 * effectiveness / 10) - (1.1 * nausea) - (0.04 * time) - (0.15 * cost / 100)

The rank orders of the groups were significantly correlated (Spearman's rank correlation coefficient, r=0.78). For both groups, the optimum hypothetical treatment was a weekly tablet with a 50% relative risk reduction of hip fracture, a total treatment duration of 5 years, that was not associated with the adverse effect of nausea, and relative low out-of-pocket costs (€240 for the total treatment of 5 years; thus, €4 per month). The four hypothetical preventive osteoporosis drug treatment scenarios that were most preferred by the GPs for osteoporosis patients had optimal levels of effectiveness, but some of them were associated with nausea as an adverse effect. The four hypothetical preventive osteoporosis drug treatment scenarios that were most preferred by the patients were not associated with nausea as an adverse effect, but the effectiveness of the treatment was not always optimal. Holding the route of drug administration constant, GPs preferred a more effective and shorter duration of the drug treatment for a patient, and accepted that it could cause nausea and required some out-of-pocket costs (drug profile 15) over one that was less effective and had a longer treatment duration (drug profile 4), whereas patients showed the reverse pattern (similarly, drug profile 12 compared with drug profile 1). This means that patients accepted a less effective and longer duration of the drug treatment if the drug treatment did not cause nausea (for up to 2 hours after intake) and did not require out-of-pocket costs.

4.4 Discussion

This study shows that the participating GPs had a much less favourable attitude towards preventive osteoporosis drug treatment than the patients. Treatment effectiveness (hip fracture risk reduction), adverse effects (nausea), total treatment duration, route of drug administration and out-of-pocket costs were all relevant to GPs' and patients' preferences for drug treatment; however, GPs placed higher relative values on effectiveness of drug treatment and shorter total treatment duration than patients.

There are no previous DCEs directly comparing preferences of health professionals and patients for preventive osteoporosis drug treatment. A DCE showed that health professionals were less focused on the process attributes of haemophilia care (e.g. infusion frequency) than patients [16]; this suggested that aspects of the treatment process was more relevant for patients than for health professionals. This is in line with our study, which showed that GPs did not prefer 'a monthly tablet' over 'a weekly tablet' or over 'an injection every 4 months' as a route of drug administration, whereas patients did. Two other preference studies, using other methods than formal DCE, also showed that health professionals had a less positive attitude towards treatment than patients (chemotherapy for solid tumours [26], and antibacterial treatment for acute respiratory illness [27]).

Our study illustrated the willingness of GPs and patients to participate in the relatively complex exercise of a DCE to weigh up the advantages and disadvantages of various

hypothetical preventive osteoporosis drug treatments. It provides further evidence that the DCE approach can be applied successfully in healthcare to directly compare (the determinants of the) preferences from health professionals and patients [15-18].

The inclusion of a 'no drug' treatment option in each choice set was warranted in the context of this study. GPs may prefer not to prescribe preventive osteoporosis drug treatment because Dutch treatment guidelines do not currently recommend preventive osteoporosis drug treatment for women who did not yet have an osteoporotic fracture or an abnormal low bone mass. For patients, the decision to start such treatment should be based on autonomous choice. Forcing GPs and patients to choose only between active treatments was likely to inflate any estimates obtained. However, including a 'no drug' treatment option raises issues about how non-participation is accounted for when analysing the data [28]. The conditional logit regression model (i.e. a multinomial model, MNL) is a relatively simple model. It assumes that errors are independent and identically distributed, that observed choices are independent, and that preferences are homogenous. A mixed logit model represents the most advanced and flexible discrete model to date, because it relaxes the assumptions embodied in MNL models. However, the MNL model was the best option for analysis of our dataset because of the relatively small number of GPs.

This study had some limitations. First, we used a main-effects only design, assuming that all attributes were valued independently of each other (i.e. all interactions between attributes were zero); this may be reasonable since main effects typically account for 70–90% of explained variance in DCE [24]. Second, this study used a postal questionnaire whereby the answers of the GPs and patients were collected by post and by telephone, respectively; this could result in bias in data collection. Third, the sample of GPs was selective, because they had already participated in a study on osteoporosis case finding; however, it is not known whether this has (largely) influenced the results. Fourth, the results of this study may not be applicable in healthcare systems in which patients never pay any out-of-pocket costs for drug treatment. Finally, the current results could gain importance if it were possible to validate the stated preferences by revealed preferences from actual behaviour, i.e. testing for external validity [29].

4.5 Conclusions

Our study found that GPs placed more importance on greater effectiveness and shorter treatment duration of preventive osteoporosis treatment than patients did.

The GPs had a much less favourable attitude towards preventive osteoporosis drug treatment than patients. This finding is consistent with other studies that show that health professionals generally have a less positive attitude towards curative treatment than patients, although our study is the first one to show this for preventive treatment. Essentially, the benefits of preventive treatment are uncertain in the individual patient

case. The decision to initiate preventive treatment should ideally be taken in a process of shared decision making in which both the patient and health professional are actively involved. Awareness and explicit address of differences in personal values regarding preventive treatment in general may have a positive effect on the process and outcomes of shared decision making on treatment initiation in individual patients.

References

1. Montgomery AA, Fahey T (2001). How do patients' treatment preferences compare with those of clinicians? Qual Health Care 10 Suppl 1:i39-43.
2. Consensus development conference: prophylaxis and treatment of osteoporosis (1991). Am J Med 90(1):107-110.
3. Riggs BL, Melton LJ (1995). The worldwide problem of osteoporosis: insights afforded by epidemiology. Bone 17(5 Suppl):505S-511S.
4. Guideline on osteoporosis (2002). Utrecht: The Dutch Institute for Healthcare Improvement (CBO).
5. Atik OS, Gunal I, Korkusuz F (2006). Burden of osteoporosis. Clin Orthop Relat Res 443:19-24.
6. Silverman SL (2005). Quality-of-life issues in osteoporosis. Curr Rheumatol Rep 7(1):39-45.
7. Black DM, Cummings SR, Karpf DB, Cauley JA, Thompson DE, Nevitt MC, et al. (1996). Randomised trial of effect of alendronate on risk of fracture in women with existing vertebral fractures. Fracture Intervention Trial Research Group. Lancet 348(9041):1535-1541.
8. Ensrud KE, Black DM, Palermo L, Bauer DC, Barrett-Connor E, Quandt SA, et al (1997). Treatment with alendronate prevents fractures in women at highest risk: results from the Fracture Intervention Trial. Arch Intern Med 157(22):2617-2624.
9. Harris ST, Watts NB, Genant HK, McKeever CD, Hangartner T, Keller M, et al (1999). Effects of risedronate treatment on vertebral and nonvertebral fractures in women with postmenopausal osteoporosis: a randomized controlled trial. Vertebral Efficacy With Risedronate Therapy (VERT) Study Group. JAMA. 282(14):1344-1352.
10. Pols HA, Felsenberg D, Hanley DA, Stepan J, Munoz-Torres M, Wilkin TJ, et al (1999). Multinational, placebo-controlled, randomized trial of the effects of alendronate on bone density and fracture risk in postmenopausal women with low bone mass: results of the FOSIT study. Fosamax International Trial Study Group. Osteoporos Int 9(5):461-468.
11. Gyrd-Hansen D, Sogaard J (2001). Analysing public preferences for cancer screening programmes. Health Econ 10(7):617-634.
12. Ryan M, Farrar S (2000). Using conjoint analysis to elicit preferences for health care. BMJ 320(7248):1530-1533.
13. Sculpher M, Bryan S, Fry P, de Winter P, Payne H, Emberton M (2004). Patients' preferences for the management of non-metastatic prostate cancer: discrete choice experiment. BMJ 328(7436):382.
14. Watson V, Ryan M, Brown CT, Barnett G, Ellis BW, Emberton M (2004). Eliciting preferences for drug treatment of lower urinary tract symptoms associated with benign prostatic hyperplasia. J Urol 172(6 Pt 1):2321-2325.
15. Bishop AJ, Marteau TM, Armstrong D, Chitty LS, Longworth L, Buxton MJ, et al (2004). Women and health care professionals' preferences for Down's Syndrome screening tests: a conjoint analysis study. BJOG 111(8):775-779.

16. Mantovani LG, Monzini MS, Mannucci PM, Scalone L, Villa M, Gringeri A (2005). Differences between patients', physicians' and pharmacists' preferences for treatment products in haemophilia: a discrete choice experiment. Haemophilia 11(6):589-597.

17. Lee A, Gin T, Lau AS, Ng FF (2005). A comparison of patients' and health care professionals' preferences for symptoms during immediate postoperative recovery and the management of postoperative nausea and vomiting. Anesth Analg 100(1):87-93.

18. Lewis SM, Cullinane FM, Carlin JB, Halliday JL (2006). Women's and health professionals' preferences for prenatal testing for Down syndrome in Australia. Aust N Z J Obstet Gynaecol 46(3):205-211.

19. Kanis JA, Delmas P, Burckhardt P, Cooper C, Torgerson D (1997). Guidelines for diagnosis and management of osteoporosis. The European Foundation for Osteoporosis and Bone Disease. Osteoporos Int 7(4):390-406.

20. de Bekker-Grob EW, Essink-Bot ML, Meerding WJ, Pols HA, Koes BW, Steyerberg EW (2008). Patients' preferences for osteoporosis drug treatment: a discrete choice experiment. Osteoporos Int 19(7):1029-1037.

21. Farrar S, Ryan M, Ross D, Ludbrook A (2000). Using discrete choice modelling in priority setting: an application to clinical service developments. Soc Sci Med 50(1):63-75.

22. Ryan M, Scott DA, Reeves C, Bate A, van Teijlingen ER, Russell EM, et al (2001). Eliciting public preferences for healthcare: a systematic review of techniques. Health Technol Assess 5(5):1-186.

23. Hahn GJ, Shapiro SS (1966). A catalogue and computer program for the design and analysis of orthogonal symmetric and asymmetric fractional factorial experiments. Schenectady, NY, USA: General Electric Research and Development Centre. Report No.: Technical report No. 66-C 165.

24. Louviere JJ, Hensher DA, Swait JD (2000). Stated choice methods: analysis and application. Cambridge: Cambridge University Press.

25. Sloane NJA. A library of orthogonal arrays. http://wwwresearchattcom/~njas/oadir/. [Accessed 2009 Feb 17]

26. Slevin ML, Stubbs L, Plant HJ, Wilson P, Gregory WM, Armes PJ, et al (1990). Attitudes to chemotherapy: comparing views of patients with cancer with those of doctors, nurses, and general public. BMJ 300(6737):1458-1460.

27. Macfarlane J, Holmes W, Macfarlane R, Britten N (1997). Influence of patients' expectations on antibiotic management of acute lower respiratory tract illness in general practice: questionnaire study. BMJ 315(7117):1211-1214.

28. von Haefen RH, Massey DM, Adamowicz W (2005). Serial non-participation in repeated discrete choice models. Am J Agric Econ 87(4):1061-76.

29. Ryan M, Gerard K (2003). Using discrete choice experiments to value health care programmes: current practice and future research reflections. Appl Health Econ Health Policy. 2(1):55-64.

30. Street D, Burgess L. Discrete choice experiments [computer software]. Sydney: University of Technology, 2007 [online]. Available from URL: http://maths.science.uts.edu.au /maths/ wiki/SPExpts { Accessed 2009 Mar 05].

5

Patients' preferences for scoliosis brace treatment: a discrete choice experiment

de Bekker-Grob EW*, Bunge EM*, van Biezen FC, Essink-Bot ML, de Koning HJ

* equal contribution

Spine; *in press*

Abstract

Study Design Discrete choice experiment

Objective To investigate the reduction in the risk of surgery that scoliosis patients would require in order to consider brace treatment as acceptable, and to elicit the trade-offs individuals make between characteristics of brace treatment.

Summary of Background Data The effectiveness of brace treatment in idiopathic scoliosis patients has not been established in randomized controlled trials (RCTs). Treatment with a brace can be quite bothersome. Patients' preferences for brace treatment are unknown. Insight into patients' preferences for (characteristics of) brace treatment will be useful for future trials and for the development of braces that may optimize compliance with brace treatment.

Methods A total of 197 patients who had completed treatment (brace and/or surgery) for scoliosis were approached for the study, of which 135 gave informed consent. A discrete choice experiment was designed in which patients had to choose between hypothetical brace treatment profiles that differed in four treatment attributes: effectiveness, visibility, discomfort, and treatment duration. A multinomial logit model was used to analyze the relative importance of these attributes. Subgroup analyses were conducted for brace-only, brace-surgery, and surgery-only patients.

Results The response rate was 86% (116/135). All treatment attributes proved to be important for patients' choices. All subgroups were prepared to initiate treatment with a Boston brace if the brace would reduce the need for surgery by 53%. Risk reductions from 32-74% were required for acceptance of a treatment duration of 3 years.

Conclusions Scoliosis patients stated to be prepared to undergo brace treatment only if it provides sizeable reduction of the risk of surgery. Effectiveness and discomfort in wearing a brace were the most important determinants of the choices. These results are important if RCTs would conclusively establish that bracing is effective, and show directions for the further technical development of braces to increase the compliance with brace treatment.

5.1 Introduction

Idiopathic scoliosis (IS) is defined as a lateral curvature of the spine with a minimal Cobb angle of 10º of unknown origin. The Cobb angle is the angle between the upper most inclined vertebra and the lower most inclined vertebra. Besides a lateral curvature of the spine, there is a fixed rotation of one or more vertebrae, and a rotational deformation of that vertebra [1]. About 0.3% of all children aged between 10 and 16 years have IS that progresses to curvatures with Cobb angles over 20-25 degrees [2]. In this situation, the chance is small that the curvature will disappear spontaneously, and brace treatment is usually applied. Brace treatment is supposed to prevent further progression of the curvature and thereby the need for surgery, which is indicated when a patient has a Cobb angle of more than about 45-50 degrees [1]. Therefore, the 'effectiveness of brace treatment' is commonly defined as the reduction of the risk of surgical intervention [3-5]. The effectiveness of brace treatment has not yet been convincingly established in RCTs. In both the US and the Netherlands, an RCT on (Boston) bracing IS patients was designed; in the US the trial is currently running (BrAIST) [6]. The Dutch trial was halted recently because of insurmountable problems with patient inclusion.

Treatment with a brace can be rather bothersome [7, 8]. Patients (usually aged 10-16 years) have to wear the brace for 18-23 hours a day, during several years of puberty [1] , the brace is often visible, and can be uncomfortable to wear. In conclusion, brace treatment has serious disadvantages and uncertain benefit.

The present study had two aims. First, we investigated the size of the reduction of the risk of surgery required by scoliosis patients to consider brace treatment, with its drawbacks, as acceptable. Second, we elicited the trade-offs individuals make between characteristics of brace treatment. The results of the study are potentially important for trials that will investigate the effectiveness of brace treatment. A trial may show that the effectiveness of a brace treatment is in fact below the threshold for acceptance required by scoliosis patients. Furthermore, the results may provide directions for the further technical development of braces, and thereby for optimising compliance with brace treatment. If brace treatment appears to be effective, but not sufficiently effective for scoliosis patients to be preferred over observation, improving brace characteristics (e.g. comfort or visibility of the brace) may decrease the required risk reduction of surgery so that brace treatment will be preferred over observation.

A discrete choice experiment (DCE) was used. DCEs have increasingly been used in health care as an approach to elicit patient preferences [9-13]. We hypothesised that patients will value brace treatments more highly once they have experienced them, and that patients, irrespective of their type of experienced treatment, will 'defend' their own treatment. These phenomena are known as "status quo bias" and "cognitive discordance", respectively, and can bias results. To offset these potential biases, and attempt to capture the full range of preferential attitudes towards brace treatments, we elicited preferences from brace-only, brace-surgery, and surgery-only patients. We used the most conservative estimates for the risk reduction required by patients to accept brace treatment.

5.2 Materials and methods

The study was conducted according to the principles of the Declaration of Helsinki. Under Dutch law, observational health surveys are exempted from approval from a Medical Ethics Committee.

5.2.1 Study sample

The 197 patients who gave consent for participation in the former NESCIO (Netherlands Evaluation Study on Screening for scoliosis) study [14], were approached again for their consent to send them a new questionnaire. These patients had completed treatment with a brace, surgery, or with a brace followed by surgery, and were recruited in 12 Dutch hospitals. For the braced patients, Boston braces had been used. Four weeks after the first mailing, a reminder was sent. A total of 135 patients gave informed consent regarding the questionnaire; for seven patients we had the incorrect address, one had moved to another country, and two did not give consent. The non-response rate at this stage of the study was 26.4%.

5.2.2 Variables

The following data concerning patient characteristics were collected: age at filling out the questionnaire, highest current or completed education, Cobb angle at diagnosis, Cobb angle after treatment, type(s) of treatment and, if applicable, total bracing period. Data on age and education were collected to assure that differences between groups could not be explained by differences in age and/or education. Data on post-treatment attitudes about choosing/not choosing brace treatment were collected to test for convergent validity of the DCE results.

5.2.3 Discrete Choice Experiment

DCEs assume that a given healthcare intervention can be described by its characteristics (attributes) and that any subject's preferences for an intervention are determined by the levels of the attributes [15]. Attributes should be identified beforehand as potentially important for the choice of an intervention [16]. The relative importance of attributes and the trade-offs patients make between these can be assessed when patients are offered a series of choices between brace treatment alternatives that have different combinations of attribute levels [17].
We identified four attributes of brace treatment, with specific focus on the attributes of the Boston brace. The attributes and the attribute levels were chosen based on literature, expert interview (specialist in orthopaedics), and personal interviews with patients who had experienced brace treatment and/or surgery for scoliosis (i.e., the target group). The attributes were: effectiveness of brace treatment, brace comfort, total treatment duration, and visibility of the brace (Table 5.1).

Table 5.1 Attributes and attribute levels for brace treatment

Attributes and attribute levels	**Coefficients in regression analysis**
Risk reduction of an operation (%) (EFFECTIVENESS):	β_1
12.5	
25	
50	
75	
Brace is visible (VISIBLE):	β_2
No (0)	
Yes (1)	
Total treatment duration (years) (TIME)	β_3
2	
4	
6	
8	
Brace is uncomfortable (DISCOMFORT)	β_4
No (0)	
Yes (1)	

The combination of attributes and attribute levels (2 attributes with 4 levels, and 2 attributes with 2 levels) resulted in 64 hypothetical brace treatment profiles ($4^2 * 2^2$). For practical reasons, not all of these could be used in a questionnaire. Therefore, we generated a sample of hypothetical brace treatment profiles from all these 64 brace treatment profiles for the questionnaire (i.e., we used a fractional factorial design) [18]. A sample of 16 hypothetical brace treatment profiles was sufficient to estimate at least all main effects in a regression analysis [19]. Based on these 16 profiles, choice sets were created. Each choice set consisted of two brace treatment profiles between which the patients could choose. If the patients considered both brace treatments as not acceptable, the patient could opt-out. Figure 5.1 shows an example. The first brace treatment profile (i.e. Brace treatment A) of each choice set was always one of the 16 hypothetical brace treatment profiles selected for the fractional factorial design. We created the second brace treatment profile (i.e. Brace treatment B) of each choice set by means of a specific technique (cycle 'fold-over') to ensure minimal overlap of attribute levels (i.e., Brace treatment A and Brace treatment B always had different attribute levels in each choice set) [20]. Too much overlap would reduce the information obtained on trade-offs between attribute levels.

Before participants started with the 16 choice sets of the DCE in the questionnaire, they were asked to rank the four attributes of a hypothetical brace treatment (total treatment

	Brace treatment A	Brace treatment B	No brace treatment
Total treatment duration	4 years	6 years	0 years
Brace is visible to wear	Yes	No	Not applicable
Risk of surgery	Reduces from 40% to 20%	Reduces from 40% to 10%	Remains 40%
Brace is uncomfortable to wear	No	Yes	Not applicable
Which brace treatment do you prefer?	**□ A**	**□ B**	**□ None**

Figure 5.1 Example of a choice set as presented in questionnaire

duration of two years, 30% risk reduction of surgery, brace is comfortable to wear, and invisible under clothing) from most important to least important. The questionnaire included a detailed written description of each attribute and its levels. While answering the choice sets of the DCE, the patients had to imagine that they were an 11-year-old patient with scoliosis that is eligible for brace treatment. A dominant choice set was included in the questionnaire to test for rationality (i.e. a choice set including one brace treatment profile characterized by logically preferable levels on all attributes). Finally, patients who had been treated with a brace (whether or not before surgery) were asked if they would choose for brace treatment again if they would face the same situation. The questionnaire was pilot tested to check for any problems in interpretation and face validity (n=10).

5.2.4 Analyses

Because some data were skewed and for some variables there were less than 30 patients in a subgroup, non-parametric tests (Mann Whitney U-test) were used to determine significant differences in age, education, Cobb angle, and brace duration between the different subgroups.

Data from respondents who failed the dominant question in the DCE were excluded from further analyses. The DCE was analyzed by taking each choice among the three options (two brace treatment alternatives, and a no brace treatment option) as an observation. The observations were analyzed by a multinomial logit regression model. Assuming that all attributes have an independent influence on a patient's preference, the following model was estimated:

$$V = \beta_0 + \beta_1 \text{TIME} + \beta_2 \text{ EFFECTIVENESS} + \beta_3 \text{VISIBLE} + \beta_4 \text{DISCOMFORT}$$

Where:

- V represents the observed preference score for a (hypothetical) brace treatment as derived from the respondents' choice behaviour;
- β_0 is an alternative specific constant reflecting the respondents' preference for receiving brace treatment relative to no brace treatment;
- β_1 to β_4 are coefficients that indicate the relative importance of each attribute (Table 5.1).

The values of V have a relative interpretation, i.e., an observed preference score for a brace treatment profile with a higher value of V is preferred over a brace treatment profile with a lower value of V. A positive preference score value of a specific brace profile indicates a preference for that treatment over no brace treatment (i.e., the preference score from the no brace treatment option was normalized to zero). The statistical significance of a coefficient indicates that the respondents considered the attribute important in their choices. A priori, we expected all attributes to be significant. The sign of a coefficient reflects whether the attribute has a positive or negative effect on preference score. We expected that only the attribute 'effectiveness' would have a positive effect (i.e., a positive sign)
The *value* of a coefficient indicates the relative contribution of the corresponding attribute to total preference score. For a correct interpretation of the comparison of the coefficients of the attributes, we need to consider the different units of measurement. For instance, a change in effectiveness of brace treatment of 1 percent may not be as important as a marginal change in any other three attributes. Assuming a linear benefit function, the change in benefit resulting from a 20 percent change in effectiveness of brace treatment is 20 times larger than a 1 percent change in effectiveness of brace treatment, which may outweighs the benefit of a marginal change in other three attributes. Subgroup analyses were conducted by using interaction terms in the multinomial logit regression model to assess whether brace-only, brace-surgery, and surgery-only patients had different preferences for each brace treatment attribute.

5.3 Results

5.3.1 Respondents

The response rate was 116/135 (86%). In total, 113 of 116 patients (97%) passed the dominant question. Of these 113 respondents, 41 had been treated with a brace only (brace-only group), 41 had been treated with a brace followed by surgery (brace-surgery group), and 31 had been treated surgically only (surgery-only group). The respondents of

each subgroup had a mean age of about 20 years at the time of completion of the DCE questionnaire (Table 5.2). There were no significant differences between the subgroups, except, as expected, the Cobb angle at diagnosis. Results of direct ranking showed that effectiveness and discomfort of the brace were considered the most important attributes of brace treatment (Table 5.3).

Table 5.2 Characteristics of the study population

	Brace-only	**Brace-surgery**	**Surgery-only**
	(n=41)	(n=41)	(n=31)
	n (%)	**n (%)**	**n (%)**
Girls	36 (87.8)	37 (90.2)	27 (87.1)
Lower education level	4 (9.8)	7 (17.5)	2 (6.5)
Intermediate education level	18 (43.9)	21 (52.5)	19 (61.3)
Higher education level	19 (46.3)	12 (30.0)	10 (32.3)
	Mean (SD)	**Mean (SD)**	**Mean (SD)**
Age at survey (years)	20.2 (1.6)	19.5 (2.5)	19.7 (2.4)
Cobb angle at diagnosis (°)[a]	27.0 (10.3)[1,2]	34.0 (13.8)[1,3]	50.0 (15.5)[2,3]
Cobb angle after treatment (°)[a]	31.0 (10.2)	30.0 (13.3)	34.0 (10.1)
Total bracing period (years)[b]	2.7 (1.8)	2.1 (1.7)	NA

[1,3] $p < 0.05$; [2] $p < 0.01$; [a] on average, 73.5% of the Cobb angles were available; [b] for five brace-only patients and five brace-surgery patients data on total bracing period were missing

Table 5.3 Direct ranking of the attributes from most important to least important

	Effectiveness	**Discomfort**	**Visible**	**Time**
	%	%	%	%
Brace-only (n=39)	53.8	23.1	20.5	2.6
Brace-surgery (n=40)	35.0	40.0	22.5	2.5
Surgery-only (n=31)	51.6	25.8	9.7	12.9

Table 5.4 Preferences for brace treatment per patient group

Attribute	All patients		Brace-only patients (BO)		Brace-surgery patients (BS)		Surgery-only patients (SO)		Differences between patients groups (p-value of the interaction)		
	coefficient[a]	p-value	coefficient[b]	p-value	coefficient[c]	p-value	coefficient[d]	p-value	BO versus SO	BO versus BS	SO versus BS
Constant (brace treatment)	0.79	**	1.99	**	0.60		-0.12		**	**	
Treatment duration (per 1 yr)	-0.35	**	-0.51	**	-0.33	**	-0.29	**	**	*	
Visible	-0.82	**	-0.94	**	-0.82	**	-0.78	**			
Effectiveness (per 10% risk reduction of surgery)	0.40	**	0.56	**	0.35	**	0.40	**	*	**	
Discomfort	-1.16	**	-1.29	**	-1.20	**	-1.05	**			

* a < 0.05

** a < 0.01

[a] Number of observations 5,376 (113 patients x 16 choices x 3 options per choice, minus 48 missing values)

[b] Number of observations 1,962 (41patients x 16 choices x 3 options per choice, minus 6 missing values)

[c] Number of observations 1,956 (41 patients x 16 choices x 3 options per choice, minus 12 missing values)

[d] Number of observations 1,488 (31 patients x 16 choices x 3 options per choice, minus 30 missing values)

5.3.2 DCE results for total group

Most patients found the DCE questions (very) clear and had no difficulties in completing the questionnaire. All coefficients were significant (Table 5.4, column 3); thus, all attributes were important for patients' choices. All signs were consistent with a priori expectations. The positive constant term (0.79) suggests a positive attitude towards brace treatment.
Patients weighted uncomfortable wearing of a brace about 1.4 times more important than visibility of wearing of a brace (Table 5.4, column 2; 1.16/0.82=1.4). As an example based on these results, the change in preference score in moving from 'no brace treatment' to a hypothetical brace treatment profile characterized by a total treatment duration of 5 years, 50% reduction in the risk of undergoing surgery, comfortable wearing, and invisibility under clothing; can be estimated as:

V = 0.79 + 5*-0.35(TIME) + 0*-0.82(VISIBLE) + 5*0.40(EFFECTIVENESS) + 0*-1.16(DISCOMFORT) = 1.04;

To this end, we used the coefficients of Table 5.4, column 2. A treatment duration of 5 years contributes negatively to V by -1.75 (5*-0.35). In contrast to visible wearing, invisible wearing of the brace under clothing does not have a negative relative contribution to V (i.e., value is zero). An effectiveness of brace treatment of 50% contributes positively to V by 5 * 0.40 per 10% risk reduction of surgery (i.e. value +2.00). And finally, in contrast to uncomfortable wearing, comfortable wearing of the brace does not contribute negatively to V (the value zero). The resulting estimate of V (0.79 + (5*-0.35) + (0*-0.82) + (5*0.40) + (0*-1.16)) results in a positive preference score of 1.04, suggesting that the participants would accept this particular brace treatment profile.

5.3.3 Subgroup analyses

Table 5.4 shows the results of multinomial logit regression modelling of data from the three subgroups and between the three subgroups (columns 4-12). The constant term was only significant for the brace-only group (Table 5.4, columns 4-5); thus, only the brace-only patients had an unconditionally positive attitude towards brace treatment. If the effectiveness of the brace treatment was 30% (as in the hypothetical brace treatment in the direct ranking exercise) the brace-only group and the surgery-only group considered the effectiveness of brace treatment to be the most important attribute (coefficients 1.68 (3*0.56) and 1.20 (3*0.40), respectively) followed by discomfort of the brace (coefficients 1.29 and 1.05, respectively) (Table 5.4, columns 4 and 8). In contrast, the brace-surgery patients considered comfort of the brace as the most important attribute (coefficient 1.20)

followed by effectiveness of the brace treatment (coefficient 1.05 (3*0.35)) (Table 5.4, column 6).

There were no significantly different preferences for attributes of brace treatment between brace-surgery patients and surgery-only patients (Table 5.4, columns 11-12). Brace-only patients were significantly less prepared to undergo long treatment duration compared to brace-surgery patients and surgery-only patients, but were more prepared to wear a less effective brace.

Brace-only patients chose significantly more often for brace treatment again if they would be in the same situation, than did patients who were treated with a brace followed by surgery (70.0% vs. 24.4%, respectively).

Figure 5.2 shows the relationship between effectiveness (%) of a Boston brace (assuming this brace is visible and uncomfortable to wear) and acceptance of maximal treatment duration (years) of bracing for brace-only, brace-surgery, and surgery-only patients. If the brace would reduce the relative risk of surgical intervention by 53%, then all sub-groups were prepared to initiate brace treatment (treatment duration of 0 years). However, brace-only, brace-surgery, and surgery-only patients were only prepared to accept a Boston brace for 3 years if the brace would reduce the relative risk of surgical intervention by 32%, 69% and 74%, respectively.

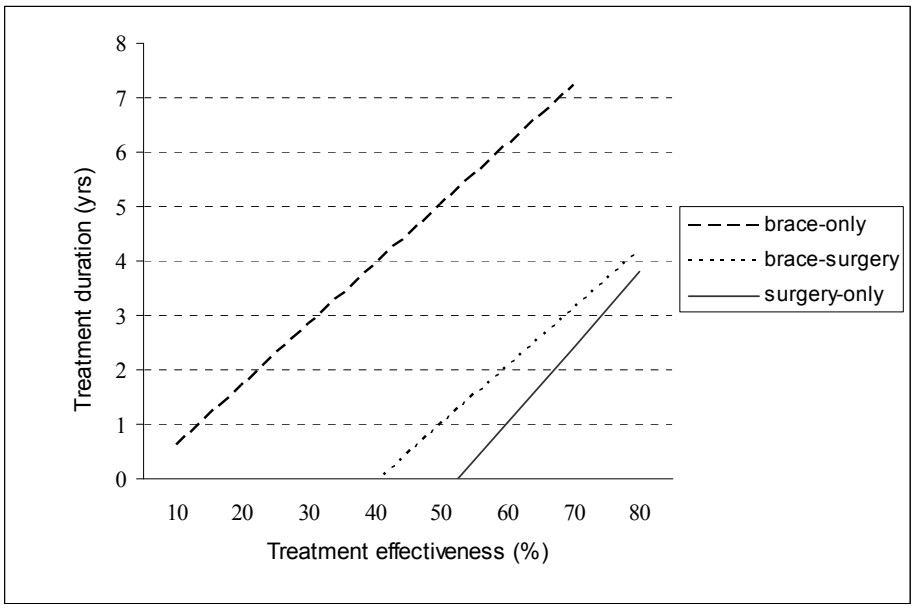


Figure 5.2 Regression lines showing the relationship between effectiveness (%) and acceptation of maximal treatment duration (years) of bracing for brace-only, brace-surgery, and surgery-only patients

5.4 Discussion

The present study shows that scoliosis patients, including patients who underwent surgery, reported that they would be prepared to undergo brace treatment only if it is associated with sizeable reduction of the risk of surgery. Effectiveness and comfortable wearing of a brace played the most important role in the patients' choices. The risk reduction required by these patients for acceptance of a Boston was in the range of 32-74%.

There are no previous DCEs investigating how large the risk reduction on surgery must be for scoliosis patients to consider brace treatment as acceptable, or eliciting the trade-offs individuals make between characteristics of brace treatment with which to compare our data. However, Dolan et al. [21] investigated the required risk reduction (number of surgeries avoided by the use of a brace before the participant would favour bracing over observation, given their understanding of the side-effects) desired by parents of scoliosis patients and parents of a surrogate group. They found a median desired risk reduction of 50% (from 60% reduced to 30%), which is quite similar to our study results. In an orthopaedic context Snoek et al. [22] used a DCE to study the effects of health outcomes and non-health outcomes on the decision for reconstructive interventions for the upper extremities in subjects with tetraplegia. They showed that process effects, which relate to the intensity of treatment, were equally important or even more important than functional outcome in the decision for reconstructive upper extremities surgery in subjects with tetraplegia. This is in line with our study, which also showed that besides health outcome (i.e., effectiveness of brace treatment) other characteristics (treatment duration, comfort and visibility) proved to be important for patients' choices.

The 'real' risk of progression and the need for surgery in case of no brace treatment is not well established. Surgical rates of 0-38% have been reported, with higher rates in not braced patients than in braced patients; however, these studies were not based on randomized controlled trials [23-25]. By including hypothetical levels of effectiveness we could extend the relative importance of the effectiveness of brace treatment in comparison with other attributes of brace treatment.

This study has illustrated the feasibility of DCE to investigate whether a given attribute is important for patients' choices for brace treatment; the relative importance of these various attributes; and the trade-offs individuals are made between these attributes. A reasonable proportion of potential respondents agreed to participate in the investigation, and only 3 of 116 failed the dominant question. This study therefore adds to the available literature on the usefulness of DCE to investigate preferences for treatment [28-30]. Comparing our DCE results with the results of the direct ranking exercise in our questionnaire, the brace-only group and the surgery-only group considered the effectiveness of brace treatment to be the most important attribute followed by comfort of the brace in both methods, whereas the brace-surgery patients considered comfort of the brace as the most important attribute followed by effectiveness of the brace treatment. These results support convergent validity of the DCE results. Prior rank-

ordering of the attributes was included to familiarize respondents with the attributes, but some (expectedly small) effects on the eventual results due to anchoring-and-adjustment bias cannot be excluded.

This study had various limitations. First, we used a main effects only design, assuming that all attributes were value-independent of each other (i.e., all interactions between attributes were zero). This may, however, be reasonable since main effects typically account for 70-90% of the explained variance in DCE [19]. Second, this DCE used attributes of the Boston brace only, because that is the most commonly used brace in the Netherlands at present. It is conceivable that other brace types, like SpineCor braces® or TriaC® may be preferred differently, because these are less rigid than the Boston brace.

Third, all patients in our study population had experienced treatment for scoliosis and therefore knew what they were choosing for in this DCE. This strength is, however, also a limitation. These patients may have 'defended' their own treatment (i.e., cognitive discordance), and this may have biased the results. For example, brace-only patients had no experience of surgery and may have been more afraid of surgery than the surgery patients and may, therefore, have expressed more willingness to wear a brace. Brace-surgery patients have experienced both forms of treatments, but this group may have been disappointed in brace treatment because eventually they had to undergo surgery. Surgery patients might be too positive towards surgery, because their surgery had been successful. The differences in the preferences we found were in the expected directions, i.e., the surgery-only patients expressed the least positive preferences for brace treatment. Therefore, we may conclude that the general attitude for brace treatment was positive. Regardless of the variations in preferences between the patient groups in this study, new IS patients who have to decide about brace treatment might have other preferences. The current results could gain importance if it were possible to ask new IS patients what preferences they state and then compare their stated preference with their actual behaviour later on.

In conclusion, our study shows that scoliosis patients, including patients who underwent surgery, reported that they would be prepared to undergo brace treatment only if it is associated with sizeable reduction of the risk of surgery. Effectiveness and comfortable wearing of a brace played the most important role in the patients' choices. The risk reduction required by these patients for acceptance of a Boston brace was in the range of 32-74%. These results are important, because they give more insight into minimal clinical important differences, which are important for trials that will investigate the effectiveness of brace treatment.

References

1. Campbell W, Canale ST, Daugherty K, et al. (2003). Scoliosis and Kyphosis. In: ST C, editor. Campbell's Operative Orthopaedics. St. Louis, MO: Mosby:1751-1984.

2. Weinstein SL (1999). Natural history. Spine 24(24):2592-2600.
3. Dolan LA, Weinstein SL (2007). Surgical rates after observation and bracing for adolescent idiopathic scoliosis: an evidence-based review. Spine 32(19 Suppl):S91-S100.
4. Goldberg CJ, Dowling FE, Hall JE, et al. (1993). A statistical comparison between natural history of idiopathic scoliosis and brace treatment in skeletally immature adolescent girls. Spine 18(7):902-908.
5. Weinstein SL, Dolan LA, Cheng JC, et al. (2008). Adolescent idiopathic scoliosis. Lancet 371(9623):1527-1537.
6. http://clinicaltrials.gov/ct2/show/NCT00448448.
7. Matsunaga S, Hayashi K, Naruo T, et al. (2005). Psychological management of brace therapy for patients with idiopathic scoliosis. Spine 30(5):547-550.
8. Tones M, Moss N, Polly DW, Jr. (2006). A review of quality of life and psychosocial issues in scoliosis. Spine 31(26):3027-3038.
9. Gyrd-Hansen D, Sogaard J (2001). Analysing public preferences for cancer screening programmes. Health Econ 10(7):617-634.
10. Ryan M, Hughes J (1997). Using conjoint analysis to assess women's preferences for miscarriage management. Health Econ 6(3):261-273.
11. Ryan M, Farrar S (2000). Using conjoint analysis to elicit preferences for health care. BMJ 320(7248):1530-1533.
12. Sculpher M, Bryan S, Fry P, et al. (2004). Patients' preferences for the management of non-metastatic prostate cancer: discrete choice experiment. BMJ 328(7436):382.
13. de Bekker-Grob EW, Essink-Bot ML, Meerding WJ, et al. (2008). Patients' preferences for osteoporosis drug treatment: a discrete choice experiment. Osteoporosis Int 19(7):1029-1037.
14. Bunge EM, Juttmann RE, de Kleuver M, et al. (2007). Health-related quality of life in patients with adolescent idiopathic scoliosis after treatment: short-term effects after brace or surgical treatment. Eur Spine J 16(1):83-89.
15. Watson V, Ryan M, Brown CT, et al. (2004). Eliciting preferences for drug treatment of lower urinary tract symptoms associated with benign prostatic hyperplasia. J Urol 172(6 Pt 1):2321-2325.
16. Farrar S, Ryan M, Ross D, et al. (2000). Using discrete choice modelling in priority setting: an application to clinical service developments. Soc Sci Med 50(1):63-75.
17. Ryan M, Scott DA, Reeves C, et al. (2001). Eliciting public preferences for healthcare: a systematic review of techniques. Health Technol Assess 5(5):1-186.
18. Hahn GJ, Shapiro SS (1966). A catalogue and computer program for the design and analysis of orthogonal symmetric and asymmetric fractional factorial experiments. Schenectady, NY, USA: General Electric Research and Development Centre.
19. Louviere JJ, Hensher DA, Swait JD (2000). *Stated choice methods: analysis and application*. Cambridge: Cambridge University Press.
20. Street DJ, Burgess L, Louviere J. (2005). Quick and easy choice sets: constructing optimal and nearly optimal stated choice experiments. International Journal of Research in Marketing 22: 459-470.
21. Dolan LA, Sabesan V, Weinstein SL, et al (2008). Preference assessment of recruitment into a randomized trial for adolescent idiopathic scoliosis. J Bone Joint Surg Am 90:2594-2605.
22. Snoek GJ, van Til JA, Krabbe PF, et al (2008). Decision for reconstructive interventions of the upper limb in individuals with tetraplegia: the effect of treatment characteristics. Spinal Cord 46:228-33.
23. Danielsson AJ, Hasserius R, Ohlin A, et al (2007). A prospective study of brace treatment versus observation alone in adolescent idiopathic scoliosis: a follow-up mean of 16 years after maturity. Spine 32(20):2198-2207.

24. Fernandez-Feliberti R, Flynn J, Ramirez N, et al (1995). Effectiveness of TLSO bracing in the conservative treatment of idiopathic scoliosis. J Pediatr Orthop 15(2):176-181.

25. Miller JA, Nachemson AL, Schultz AB (1984). Effectiveness of braces in mild idiopathic scoliosis. Spine 9(6):632-635.

26. Lloyd A, McIntosh E, Price M (2005). The importance of drug adverse effects compared with seizure control for people with epilepsy: a discrete choice experiment. Pharmacoeconomics 23(11):1167-1181.

27. Mahadevia P, Shah S, Mannix S, et al (2006). Willingness to pay for sensory attributes of intranasal corticosteroids among patients with allergic rhinitis. J Manag Care Pharm 12(2):143-151.

28. Lancsar EJ, Hall JP, King M, et al (2007). Using discrete choice experiments to investigate subject preferences for preventive asthma medication. Respirology 12(1):127-136.

6

Patients' preferences for breast reconstruction: a discrete choice experiment

Damen THC, de Bekker-Grob EW, Mureau MAM, Menke-Pluijmers MB, Seynaeve C, Hofer SOP, Essink-Bot ML

Submitted

Abstract

Background Patients' preferences are important determinants in the decision for a specific type of breast reconstruction (BR). Understanding patients' motivational factors in the decision for a specific type of BR can contribute to further improve patient information and to develop demand-led healthcare. We explored patients' preferences for three BR modalities.

Methods We approached 386 patients who previously underwent a therapeutic or prophylactic mastectomy, of whom 247 had also undergone a BR. A discrete choice experiment (DCE) was designed. BRs were characterized by six treatment attributes. Relative importance of attributes and trade-offs patients were willing to make between them were analyzed using a multinomial logit regression model.

Results Overall response rate was 71%. All treatment characteristics proved important for patients' choices, with material and aesthetic result being most important. Patients who had not undergone a BR had a stronger preference for BR with autologous tissue. Genetically predisposed patients expressed more negative consequences (disutility) from long-term complications than those without an increased risk for breast cancer. A unique model was developed to estimate utilities derived from different types of BR. Autologous free flap BR fitted in best with patients' preferences.

Conclusions Material and aesthetic result are the most important determinants for patients' choice for a specific type of BR after mastectomy. Our study provides insight into the relative weight patients place on various aspects of BR and the trade-offs they make between BR characteristics. Patients' preferences need to be taken seriously in the development of demand-led healthcare for breast cancer patients.

6.1 Introduction

Breast reconstruction (BR) is aimed at restoring patients' quality of life and body image after mastectomy, and increasingly becomes an integral part of breast cancer (BC) treatment. Multiple techniques are available, differing in characteristics such as material used, duration of the operation(s), recovery period, complication rates, and aesthetic result. Each technique has (dis)advantages and therefore its own place in current practice [1].

Multiple factors have to be considered to determine the optimal treatment modality for individual patients. Procedural characteristics have to be regarded in the context of patient-related factors (e.g., age, medical history, body habitus) and surgeon-related factors (e.g., expertise, experience) to assess which methods are technically feasible and which risks acceptable. Adjuvant radiation therapy, for example, is a relative contra-indication for implant BR due to increased complication rates, while sufficient excess abdominal tissue is a prerequisite for autologous BR using abdominal tissue [2, 3].

Complementary to the medical analysis performed by the plastic surgeon regarding which BR type individual patients could undergo, patients' preferences for the procedure they would opt for are also important determinants of the treatment choice. Understanding women's motivational factors and personal views can contribute to further improve patient-centered and demand-led healthcare.

The aim of this study was to explore patients' preferences for different BR modalities after (prophylactic) mastectomy using a discrete choice experiment (DCE). We also evaluated whether preferences differed between patients who had or had not undergone a BR and between women with or without a genetic predisposition to develop BC.

6.2 Patients and methods

6.2.1 Breast reconstruction

There are essentially three types of BR after mastectomy, using implant material, autologous tissue, or a combination of both [1]. At our institution, the most frequently performed methods per category are: 1) breast implants preceded by tissue expansion, 2) free Deep Inferior Epigastric Artery Perforator (DIEP) flap, and 3) pedicled Latissimus Dorsi (LD) flap with an implant, respectively. BR can be performed directly after mastectomy (primary BR) or at a later stage (delayed or secondary BR). The Daniel den Hoed Cancer Center specifically attracts patients with a genetic predisposition to develop BC, who frequently opt for prophylactic bilateral mastectomy. In the Netherlands, BR is covered by basic health insurance and no out-of-pocket expenses are required from patients.

6.2.2 Discrete choice experiment

DCEs are a rather novel approach to elicit patient preferences, with their origin in mathematical psychology. They assume that healthcare interventions can be described by their characteristics (attributes) and that patients' preferences are determined by attribute levels [4]. Relative importance of attributes and trade-offs that respondents make between them can be assessed by offering a series of choices between treatment alternatives with different combinations of attribute levels [5]. Concept and definitions are illustrated in Figure 6.1. A DCE was designed to determine trade-offs that women who underwent a prophylactic or therapeutic mastectomy make when they are offered a choice between different BR types.

6.2.3 Attributes and attribute levels

Choice of attributes and their levels was based on data from literature and interviews with specialists and with women having had a mastectomy or BR [2, 3, 6-9]. Six attributes were selected: 1) material used for reconstruction; 2) number and duration of operations; 3) short- and 4) long-term complication rate; 5) aesthetic result; 6) waiting time.
Three levels were determined per attribute. Some hypothetical levels were included to allow assessment of preferences to be extended beyond currently available treatments and to potentially guide the development of new techniques (Table 6.1).

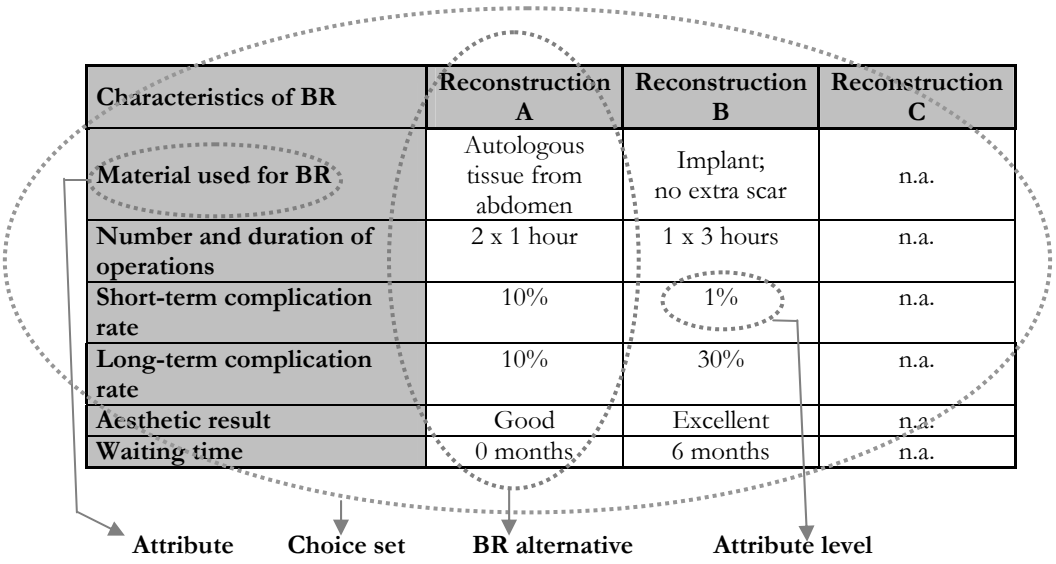


Figure 6.1 Concept of DCEs illustrated using a choice set as presented in questionnaire – Which type of breast reconstruction would you choose?

Table 6.1 Six attributes with three levels each used for Discrete Choice Experiment to assess women's preferences for three different types of BR

Attributes and levels	**Coefficient in regression analysis**
Material used for reconstruction	
Implant	β_1
Autologous tissue from abdomen	β_2
Autologous tissue from back with implant	
Number and duration of operation(s)	
1 x 3 hours	
2 x 1 hour	β_3
1 x 7 hours	β_4
Short-term complication rate	β_5
1 percent	
5 percent	
10 percent	
Long-term complication rate	β_6
0 percent	
10 percent	
30 percent	
Aesthetic result	
Moderate	
Good	β_7
Excellent	β_8
Waiting time	β_9
0 months	
6 months	
12 months	

6.2.4 Study design and questionnaire

The combination of attributes and attribute levels (6 attributes with 3 levels each) resulted in 729 (3^6) hypothetical BR alternatives. Eighteen BR alternatives proved sufficient to estimate all main effects in a regression analysis, while guaranteeing orthogonality (attributes being independent of each other) and level balance (levels occurring with equal frequency) [10, 11]. Choice sets were designed using the ´fold-over´ technique to ensure minimal overlap of attribute levels between BR alternatives and optimal efficiency [12]. All main effects were uncorrelated [13]. Choice sets consisted of two BR alternatives and a 'no reconstruction' option to allow patients to 'opt out' (Figure 6.1); BR is elective surgery and patients should not be forced to choose BR.

Previous studies demonstrated that more than 16 choice sets per respondent are associated with lower response rate and/or response reliability [14, 15]. Therefore, 18 choice sets were divided over two questionnaires containing nine choice sets each (i.e., two questionnaires constituted one full dataset) by using a blocked design [16].
Each questionnaire started with a detailed description of attributes and their levels. Pictures were included to demonstrate 'moderate' and 'excellent' aesthetic results. Short-term complications (e.g., infection, haematoma, or flap failure) were defined as arising within 6 weeks after BR, while long-term complications (e.g. capsular contracture, abdominal herniation) were stated to arise between 6 weeks and 5 years after BR. A dominant choice set (i.e., a choice set in which both alternatives used implant material, but one was characterized by logically preferable levels on all other attributes) was included to test for rationality.
The main part of each questionnaire comprised of nine choice sets. Patients were asked to consider all three options as realistic alternatives and to choose the option that appealed most to them. Furthermore, eight questions covered medical history, satisfaction with previous BR, whether or not patients would opt for (the same type of) BR again, marital status, and level of education.
The questionnaire was pilot tested (n=10) to check for problems in interpretation and face validity. As none of the respondents raised any problems, no alterations were made.

6.2.5 Study sample

We randomly approached 386 women from a total group of 820 who had undergone a mastectomy with (n=247) or without (n=139) BR between 2002 and 2006. The vast majority of patients had been diagnosed with sporadic BC. Seventy-nine patients were genetically predisposed to develop BC due to a BRCA 1/2 mutation and chose to undergo a contralateral or bilateral prophylactic mastectomy. None of the respondents had signs of BC recurrence at the time of the study. Women over 70 years of age were not eligible.
The number of respondents, choices per respondent, attributes, and attribute levels determine the power of a DCE. A formal power analysis is not feasible, however, as patients' preferences are hard to predict. Earlier studies have shown that the sample size of our study is sufficient for reliable statistical analyses [4, 17].

6.2.6 Procedure

The questionnaire and a prepaid return envelope were mailed to 386 patients. After 4 weeks non-responders were sent a reminder and after two months remaining non-responders were contacted by telephone. Written informed consent was obtained from all patients. The study was approved by the medical ethical review committee (MEC-2007-406).

6.2.7 Statistical analyses

Differences in age were analyzed using Student's t-tests for independent samples, while differences in categorical variables were analyzed using Pearson chi-square tests.
The DCE was analysed taking each choice among the three options as an observation. Data from respondents who failed the dominant question were excluded from further analysis. Remaining observations were analyzed by a multinomial logit regression model to determine the relative importance of treatment attributes. This model was implemented in SAS software (Version 9.1, SAS Institute Inc., Cary, NC, USA). Assuming that all attributes have an independent influence on women's preferences, the following model was estimated [11]:

$$V = \beta_0 + (\beta_1 * \text{implant}) + (\beta_2 * \text{autologous tissue from abdomen}) + (\beta_3 * (\text{2x1 hour operations})) + (\beta_4 * (\text{1x7 hour operation})) + (\beta_5 * \text{short term complications}) + (\beta_6 * \text{long term complications}) + (\beta_7 * \text{good aesthetic result}) + (\beta_8 * \text{excellent aesthetic result}) + (\beta_9 * \text{waiting time})$$

Where:

- V represents total relative utility or relative positive consequences derived from a certain type of BR, which can be viewed as the preference for that particular health state [18].
- β_0 is a constant reflecting respondents' preference for receiving BR relative to 'no BR';
- β_1 to β_9 are coefficients indicating the relative importance of each attribute or attribute level; β_1 to β_4, β_7 and β_8 are dummy variables of 'material used', 'duration of operation(s)', and 'aesthetic result', with 'autologous tissue from the back with implant', '1x3 hour operation', and 'moderate aesthetic result' as respective base levels;
- short- and long-term complication rates are scored as a fraction of 10%;
- waiting time is scored in months.

Using these coefficients, relative utility scores of BR profiles can be generated and subsequently ranked. The higher a relative utility score, the stronger the preference for that particular BR alternative. Relative values allow comparison of strengths of preferences for specific BR types. Absolute values of V, however, have no direct interpretation [11]. The sign of a coefficient reflects whether the attribute has a positive or a negative effect on utility, and the value indicates the relative importance of the corresponding attribute to total relative utility. A statistically significant coefficient is interpreted to indicate that respondents considered the attribute important. A priori all attributes were expected to be important, and material and aesthetic result were expected to have a positive effect. Trade-offs that respondents were willing to make between attributes were estimated by the ratios of the coefficients. For example, β_2/β_9 estimates

how much longer respondents are willing to wait (in months) to undergo autologous BR instead of BR with autologous tissue and an implant.
Subgroup analyses were conducted using interaction terms in the multinomial logit regression model to assess whether patients who previously underwent BR had different preferences than patients who had not. We also compared patients who had undergone BR after sporadic BC to genetically predisposed patients who had undergone BR.

6.3 Results

6.3.1 Respondents

Of 386 invited patients, 320 responded and 272 agreed to participate (overall response rate: 71%). Patients who had undergone a BR (n=186; 69%) were significantly younger ($p<0.01$) and more frequently had an increased risk of BC ($p<0.001$) than patients who had not undergone a BR. Of all respondents who underwent BR, mutation carriers were also younger than patients not being at increased risk of hereditary BC ($p<0.01$), and more frequently lived with a partner ($p=0.04$; Table 6.2). Table 6.3 gives an overview of the experiences of respondents who underwent BR.

6.3.2 DCE results

In total, 270 of 272 patients (99%) passed the dominant question. All treatment attribute coefficients were significant and all signs were in keeping with a priori expectations (Table 6.4). The positive constant term suggests that respondents generally had a positive attitude towards BR. Respondents preferred autologous tissue over implants, while autologous tissue combined with an implant was least popular. A short operation was preferred over a long operation and patients would rather undergo two short operations than one long one. Patients were less likely to choose options with increasing complication rates, both short- and long-term, but short-term complications were more important than long-term complications (higher regression coefficient for short-term complications). An 'excellent' aesthetic result was preferred over 'good' or 'moderate'.
Noteworthy, the magnitude of attribute coefficients corresponds with their relative importance. For example, the coefficient of -0.30 for long-term complication rate implies the decrease in utility per 10% rise. A risk reduction of 30% therefore corresponds with a utility increase of 0.90.
Nearly 75% of respondents who underwent BR were willing to make a trade-off with regard to material used for BR (data not shown). Only 27% consistently chose their preferred material, which in over 90% of these cases matched their own BR type.
The results of subgroup analyses are presented in Table 6.5. Patients with a BR had a significantly more positive attitude towards BR than women without a previous BR. Except for women without BR having a significantly stronger preference for autologous

Table 6.2 Respondent characteristics

	All patients		Patients with BR		Patients without BR		p-value[a]	Patients with BR, without genetic predisposition		Patients with BR, with genetic predisposition		p-value[b]
	n	%	n	%	n	%		n	%	n	%	
	270	100.0	186	100.0	84	100.0		134	100.0	52	100.0	
Age (years)							<0.01					<0.01
20-29	2	0.7	2	1.1	0	0.0		1	0.7	1	1.9	
30-39	35	13.0	31	16.7	4	4.8		10	7.5	21	40.4	
40-49	72	26.7	57	30.6	15	17.9		40	29.9	17	32.7	
50-59	108	40.0	73	39.2	35	41.7		63	47.0	10	19.2	
60-69	53	19.6	23	12.4	30	35.7		20	14.9	3	5.8	
Genetic predisposition							<0.01					
Yes	55	20.4	52	28.0	3	3.6		0	0.0	52	100.0	
Household							0.31					0.04
With partner	222	82.2	150	80.6	72	85.7		103	76.9	47	90.4	
Eductional level							0.52					0.21
Low	98	36.3	65	34.9	33	39.3		52	38.8	13	25.0	
Intermediate	87	32.2	64	34.4	23	27.4		43	32.1	21	40.4	
High	85	31.5	57	30.6	28	33.3		39	29.1	18	34.6	

[a] Differences between patients who did and did not undergo breast reconstruction (t-test or Pearson chi-squared test)

[b] Differences between patients with and without a genetic predisposition to develop breast cancer (t-test or Pearson chi-squared test)

Table 6.3 Characteristics and experiences of respondents who previously underwent breast reconstruction

	All Patients with BR		**TE / Implant****		**LD +/- Implant****		**DIEP, fTRAM****		**p-value**
	n	%	n	%	n	%	n	%	
	186	100	70	100	47	100	49	100	
Type of BR*									
Tissue expander. implant	87	42.6							
LD +/- implant	60	29.4							
Abdominal autologous tissue (DIEP, fTRAM)	55	27.0							
Other	2	1.0							
Laterality of BR									<0.01
Unilateral	99	53.2	21	30.0	38	80.9	34	69.4	
Bilateral	87	46.8	49	70.0	9	19.1	15	30.6	
Timing of BR									<0.01
Immediate (primary)	85	45.7	50	71.4	16	34.0	9	18.4	
Delayed (secondary)	91	48.9	18	25.7	30	63.8	38	77.6	
Combination	10	5.4	2	2.9	1	2.1	2	4.1	
Complications (perceived)									<0.01
Short-term complications	48	25.8	7	10.0	16	34.0	17	34.7	
Long-term complications	31	16.7	12	17.1	9	19.1	2	4.1	
Satisfaction with BR									<0.01
Satisfied	118	63.4	32	45.7	33	70.2	41	83.7	
Reasonably satisfied	58	31.2	31	44.3	13	27.7	8	16.3	
Not satisfied	10	5.4	7	10.0	1	2.1	0	0.0	
Would opt for BR again									0.50
Yes	184	98.9	69	98.6	47	100.0	49	100.0	
If yes, would choose same BR procedure again									0.03
Yes	154	83.7	60	85.7	38	80.9	48	98.0	

* Eighteen patients underwent more than one breast reconstruction, resulting in different totals; ** Eighteen patients who underwent more than one type of breast reconstruction were excluded from the subgroup analysis; BR = breast reconstruction; TE / Implant = implant preceded by tissue expansion; LD +/- Implant = Latissimus Dorsi transposition with implant; DIEP = Deep Inferior Epigastric Artery Perforator flap; fTRAM = free Transverse Rectus Abdominis Myocutaneous flap

tissue, there were no differences between both groups. Mutation carriers who had undergone BR experienced more negative consequences (disutility) from long-term complications than those without a genetic predisposition.
The relative importance of attributes was subsequently used to estimate and compare the relative utility derived from different BR profiles (Table 6.6). In contrast to (relatively) fixed levels for material and duration of operation, levels for complication rates and aesthetic result are more variable and prone to discussion. The isolated effect of material and number and duration of operations on utility was therefore evaluated first by assuming equal complication rates (5%), aesthetic results (good), and waiting time (0 months) (Table 6, top lines). Under these assumptions, utility derived by DIEP flap BR was higher than utilities derived by implants preceded by TE and LD transpositions (relative utility of 0.93 compared to 0.78 and 0.66, respectively). The effect of changes in attribute levels on relative utilities and the subsequent ranking of different treatment scenarios are also presented.

Table 6.4 Women's preferences for breast reconstruction

Attribute	**Coefficient in regression analysis**	**p-value**
	n=270	
Constant (breast reconstruction)	0.20	0.05
Material used for reconstruction (base level: autologous tissue from back with implant)		
Implant	0.33	<0.01
Autologous tissue from abdomen	0.76	<0.01
Number and duration of operation(s) (base level: 1 x 3 hours)		
2 x 1 hour	-0.21	<0.01
1 x 7 hours	-0.49	<0.01
Short-term complication rate (per 10 %)	-0.43	<0.01
Long-term complication rate (per 10 %)	-0.30	<0.01
Aesthetic result (base level: moderate)		
Good	0.82	<0.01
Excellent	1.18	<0.01
Waiting time (per month)	-0.01	0.05

Table 6.5 Differences in preferences for breast reconstruction between patients who did and did not undergo breast reconstruction and between patients with and without a genetic predisposition to develop breast cancer (t-test or Pearson chi-squared)

Attribute	Coefficient of patients with BR	Coefficient of patients without BR	p-value	Coefficient of patients with BR, without genetic predisposition	Coefficient of patients with BR, with genetic predisposition	p-value
	n=186	n=84		n=134	n=52	
Constant (breast reconstruction)	1.43	-1.44	<0.01	1.19	1.79	0.08
Material used for reconstruction (base level: autologous tissue from back with implant)						
Implant	0.31	0.43	0.55	0.20	0.56	0.08
Autologous tissue from abdomen	0.67	1.28	<0.01	0.83	0.60	0.32
Number and duration of operation(s) (base level: 1 x 3 hours)						
2 x 1 hour	-0.15	-0.45	0.11	-0.13	-0.20	0.75
1 x 7 hours	-0.49	-0.49	0.99	-0.38	-0.55	0.40
Short-term complication rate (per 10 %)	-0.44	-0.39	0.80	-0.38	-0.74	0.13
Long-term complication rate (per 10 %)	-0.30	-0.43	0.06	-0.25	-0.47	<0.01
Aesthetic result (base level: moderate)						
Good	0.85	0.69	0.43	0.84	1.13	0.18
Excellent	1.27	1.12	0.45	1.25	1.66	0.08
Waiting time (per month)	-0.02	<0.01	0.26	-0.01	<0.01	0.71

Table 6.6 Relative utility scores and ranks of several breast reconstruction profiles

Material used for BR	Number and duration of operation	Short-term complications	Long-term complications	Aesthetic result	Waiting time	Utility score[a]	Rank
TE / implant							
Implant	2 x 1 hour	5%	5%	Good	0 months	0.78	9
Implant	2 x 1 hour	5%	**10%**	Good	0 months	0.63	12
Implant	2 x 1 hour	5%	**20%**	Good	0 months	0.33	15
Implant	2 x 1 hour	5%	**30%**	Good	0 months	0.03	17
Implant	2 x 1 hour	5%	**10%**	**Moderate**	0 months	-0.20	18
Implant	2 x 1 hour	5%	**10%**	**Excellent**	0 months	0.99	5
LD + implant							
Autologous + implant	1 x 3 hour	5%	5%	Good	0 months	0.66	11
Autologous + implant	1 x 3 hour	5%	**10%**	Good	0 months	0.51	14
Autologous + implant	1 x 3 hour	5%	**20%**	Good	0 months	0.21	16
Autologous + implant	1 x 3 hour	5%	5%	**Excellent**	0 months	1.02	4
Autologous + implant	1 x 3 hour	5%	**10%**	**Excellent**	0 months	0.87	7
Autologous + implant	1 x 3 hour	5%	**20%**	**Excellent**	0 months	0.57	13
DIEP							
Autologous	1 x 7 hour	5%	5%	Good	0 months	0.93	6
Autologous	1 x 7 hour	**10%**	5%	Good	0 months	0.71	10
Autologous	1 x 7 hour	5%	**1%**	Good	0 months	1.05	3
Autologous	1 x 7 hour	5%	5%	**Excellent**	0 months	1.29	2
Autologous	1 x 7 hour	5%	5%	Good	**12 months**	0.81	8
Autologous	**1 x 3 hour**	5%	5%	Good	0 months	1.42	1

[a] Breast reconstruction relative utility = 0.20 + (0.33 * implant) + (0.76 * autologous) - (0.21 * 2 x 1 hour) - (0.49 * 1 x 7 hour) - (0.43 * short term complication rate) - (0.30 * long term complication rate) + (0.82 * good) + (1.18 * very good) - (0.01 * waiting time)

BR = breast reconstruction; TE / implant = implant preceded by tissue expansion; LD + implant = Latissimus Dorsi transposition with implant; DIEP = Deep Inferior Epigastric Perforator flap

6.4 Discussion

This discrete choice experiment showed that autologous material and an excellent aesthetic result were the most important determinants in women's choices for (a specific type of) BR.

Patients' motivation for and satisfaction with BR have been previously assessed, mostly focusing on general determinants of patients' preferences for BR and its timing [19-22]. These studies, however, did not distinguish between different types of BR, nor did they evaluate the impact of specific procedural characteristics on patients' choices. Specific comparative data are therefore not available.

DCEs are increasingly being used in healthcare to explore trade-offs women make between different treatment modalities [4, 23, 24]. With a high participation rate and only 2 of 272 respondents failing the dominant question, this DCE demonstrated its feasibility to elicit (determinants of) women's preferences for BR as well as patients' willingness to participate in a relatively complex study to weigh up pros and cons of various BR treatments.

Characteristics of subgroups with regard to age, genetic predisposition, complication rates, and patient satisfaction were in line with other publications. Patients who undergo a BR tend to be younger than patients who refrain from BR [25, 26]. In addition, a BRCA mutation is generally recognized at a younger age than the mean age at which BC develops in the population, and women who opt for prophylactic mastectomy tend to request BR as well. Short-term complications are more frequent in autologous BR, while long-term complications (e.g., capsular contracture) are typical for implant reconstructions [7]. Patient satisfaction after BR is generally high, but several studies reported higher satisfaction rates after autologous reconstructions [19, 27]. This unequal division of complication and satisfaction rates could potentially affect our results: preference for autologous tissue may be overrated due to higher satisfaction levels in women who previously underwent an autologous BR compared to women who experienced other types of BR. It was technically not possible to fully adjust for this. Patients' backgrounds should therefore be kept in mind when analyzing our results.

Respondents having undergone a LD transposition were least likely to choose the same type of BR again. DCE results showed that both implant material and autologous tissue were preferred over the combination of autologous tissue with an implant. These findings are in line with each other and demonstrate our study's (internal) validity.

Genetically predisposed women having undergone mastectomy and BR reported more expected disutility from long-term complications than those not being at increased risk of hereditary BC. A generally younger age combined with a prophylactic rather than a therapeutic mastectomy might explain why the former group considers long-term complications more cumbersome than older women with a history of BC who are more preoccupied with survival and short-term results of BR.

In estimating the relative utilities for existing BR techniques (Table 6.6), long-term complication rates of 5% and 10% were firstly used. In reality, however, long-term

complication rates after implant reconstructions have been reported to be as high as 40% [7]. Similarly, we based our estimates on 'good' aesthetic results while autologous reconstructions are frequently said to yield superior aesthetic results [19]. In our own experience, DIEP flap surgery can frequently be performed in less than five hours. Such changes in attribute levels have a major impact, and would drastically increase the differences between autologous and implant BR. In contrast, waiting time hardly affected utility levels. In our opinion, this finding is remarkable as waiting time for autologous BR has recently been a big issue in the Dutch media. We had therefore expected a more substantial effect for this variable.

All patients in our study population had experienced mastectomy and a majority of patients had also undergone BR. Respondents could therefore identify themselves with the presented choices, which is one of the prerequisites for a successful DCE. This design, however, may simultaneously bias the results, as patients who previously underwent BR are likely to incorporate their experiences in their answers and 'defend' their own choice and treatment (cognitive dissonance) [28]. This mechanism could explain why patients who had not undergone a BR had an even stronger preference for autologous material than women with a BR. Respondents who previously had a positive experience with implant BR (nearly 60% of all respondents with a BR) are less likely to demonstrate a strong preference for autologous tissue, regardless of their original preferences. Nevertheless, only 27% of respondents who had experienced BR were not willing to make a trade-off with regard to material used and consistently chose their preferred material.

The choices our respondents had to make were realistic, yet in real life decisions had already been made. It is possible, therefore, that preferences presented here differ from those of patients facing a forthcoming mastectomy. The current study design does not allow stated preferences to be validated by revealed preferences from actual behavior (i.e., testing for external validity) [29]. This interesting comparison would be possible if the study was repeated prospectively in women who are about to undergo mastectomy and are currently deciding on BR. A prospective design would also prevent previous experience from interfering with patients' choices. Further research could shine a light on this.

Given cross-national differences in health care organization and financing, it may not be possible to completely generalize these results to other countries, such as the US.

In conclusion, this DCE showed that patients' choices for BR after (prophylactic) mastectomy are influenced by all investigated BR attributes, with material and aesthetic result being the most important ones. Autologous free flap BR (e.g., DIEP flap) seemed to fit in best with patients' preferences. More insight into the relative weight patients place on various aspects of BR and the trade-offs they make between BR attributes could enable healthcare workers to improve counseling and may have a positive effect on the process and outcomes of shared decision-making of BC treatment.

References

1. Cordeiro PG. Breast reconstruction after surgery for breast cancer. N Engl J Med 2008;359(15):1590-601.
2. Contant CM, van Geel AN, van der Holt B, et al (2000). Morbidity of immediate breast reconstruction (IBR) after mastectomy by a subpectorally placed silicone prosthesis: the adverse effect of radiotherapy. Eur J Surg Oncol 26:344-50.
3. Allen RJ, Treece P (1994). Deep inferior epigastric perforator flap for breast reconstruction. Ann Plast Surg 32:32-38.
4. Watson V, Ryan M, Brown CT, et al (2004). Eliciting preferences for drug treatment of lower urinary tract symptoms associated with benign prostatic hyperplasia. J Urol 172:2321-2325.
5. Ryan M, Scott DA, Reeves C, et al (2001): Eliciting public preferences for healthcare: a systematic review of techniques. Health Technol Assess 5:1-186
6. Gabriel SE, Woods JE, O'Fallon WM, et al (1997). Complications leading to surgery after breast implantation. N Engl J Med 336:677-682.
7. Handel N, Cordray T, Gutierrez J, et al (2006). A long-term study of outcomes, complications, and patient satisfaction with breast implants. Plast Reconstr Surg 117:757-767; discussion 768-772.
8. Hofer SOP, Damen THC, Mureau MAM, et al (2007). A critical review of perioperative complications in 175 free Deep Inferior Epigastric Perforator flap breast reconstructions. Annals of Plastic Surgery 59:137-142.
9. Moore TS, Farrell LD (1992). Latissimus dorsi myocutaneous flap for breast reconstruction: long-term results. Plast Reconstr Surg 89:666-672; discussion 673-674.
10. Hahn GJ, Shapiro SS (1966). A catalogue and computer program for the design and analysis of orthogonal symmetric and asymmetric fractional factorial experiments. Schenectady, NY, USA, General Electric Research and Development Centre.
11. Louviere JJ, Hensher DA, Swait JD (2000). Stated choice methods: analysis and application. Cambridge, Cambridge University Press.
12. Street DJ, Burgess L, Louviere JJ. Quick and easy choice sets: constructing optimal and nearly optimal stated choice experiments. Intern J of Research in Marketing 2005;22:459-470.
13. Street DJ, Burgess L. The Construction of Optimal Stated Choice Experiments: Theory and Methods. http://maths.science.uts.edu.au/maths/wiki/SPExpts
14. Hall J, Fiebig DG, King MT, et al (2006). What influences participation in genetic carrier testing? Results from a discrete choice experiment. J Health Econ 25:520-537.
15. Pearmain D, Swanson J, Kroes E, et al (1991). Stated preferences techniques: a guide to practice. The Hague, Steer Davis Gleave and Hague Consulting Group.
16. Hensher DA, Rose JM, Greene WH (2005). Applied choice analysis: a primer. Cambridge, Cambridge University Press.
17. Weston A, Fitzgerald P (2004). Discrete choice experiment to derive willingness to pay for methyl aminolevulinate photodynamic therapy versus simple excision surgery in basal cell carcinoma. Pharmacoeconomics 22:1195-1208.
18. Chew RT, Sprague S, Thoma A (2005). A systematic review of utility measurements in the surgical literature. J. Am. Coll. Surg. 200: 954-964.
19. Alderman AK, Wilkins EG, Lowery JC, et al (2000). Determinants of patient satisfaction in postmastectomy breast reconstruction. Plast Reconstr Surg 106:769-776.

20. Ananian P, Houvenaeghel G, Protiere C, et al (2004). Determinants of patients' choice of reconstruction with mastectomy for primary breast cancer. Ann Surg Oncol 11:762-771.
21. Harcourt D, Rumsey N (2004). Mastectomy patients' decision-making for or against immediate breast reconstruction. Psychooncology 13:106-115.
22. Keith DJ, Walker MB, Walker LG, et al (2003). Women who wish breast reconstruction: characteristics, fears, and hopes. Plast Reconstr Surg 111:1051-1056; discussion 1057-1059.
23. Lloyd A, McIntosh E, Price M (2005). The importance of drug adverse effects compared with seizure control for people with epilepsy: a discrete choice experiment. Pharmacoeconomics 23:1167-81.
24. De Bekker-Grob EW, Essink-Bot ML, Meerding WJ, et al (2008). Patients' preferences for osteoporosis drug treatment: a discrete choice experiment. Osteoporos Int 19:1029-1037.
25. Rowland JH, Desmond KA, Meyerowitz BE, et al (2000). Role of breast reconstructive surgery in physical and emotional outcomes among breast cancer survivors. J Natl Cancer Inst 92:1422-1429.
26. Reaby LL (1998). Reasons why women who have mastectomy decide to have or not to have breast reconstruction. Plast Reconstr Surg 101:1810-1818.
27. Damen THC, Timman R, Kunst HH, Gopie JP, Bresser PJC, Seynaeve C, et al. (2008). High satisfaction rates in women after DIEP flap breast reconstruction. J Plast Reconstr Aesthet Surg;doi:10.1016/j.bjps.2008.08.019.
28. Festinger L (1957). A theory of cognitive dissonance. Stanford CA, Stanford University Press.
29. Ryan M, Gerard K (2003). Using discrete choice experiments to value health care programmes: current practice and future research reflections. Appl Health Econ Health Policy 2:55-64.

7

Preferences for colorectal cancer screening strategies in the Netherlands; A discrete choice experiment

Hol L, de Bekker-Grob EW, van Dam L, Donkers B, Kuipers EJ, Habbema JDF, Steyerberg EW, van Leerdam ME, Essink-Bot ML

Submitted

Abstract

Objectives To determine the influence of different screening tests and their characteristics on individual preferences for colorectal cancer (CRC) screening.

Design A discrete choice experiment (DCE).

Setting Region Rotterdam in the Southwest of the Netherlands.

Participants The DCE questionnaire was sent to a population based random sample (n=1,498) of the screening naïve population (50-74 years old) and to a random sample (n=769) of previously screened subjects of a CRC screening pilot programme using faecal occult blood test (FOBT, n=512) or flexible sigmoidoscopy (FS, n=257).

Main outcome measures Preferences and expected uptake for and trade-offs between different screening strategies FOBT, FS and total colonoscopy (TC), screening intervals and risk reduction of CRC related mortality.

Results In total 489/1,498 (33%) screening-naïve (52% male; mean age±SD 61±7yrs) and 545/769 (71%) screened subjects (52% male; mean age±SD 61±6yrs) returned the questionnaire. Type of screening test, screening interval, and risk reduction of CRC related mortality influenced subjects' preferences for CRC screening (all $p<0.05$). Screening-naïve and screened subjects equally preferred five-yearly FS and ten-yearly TC screening ($p=0.24$; $p=0.11$). They preferred both endoscopic screening options to FOBT screening if, based on the literature, realistic risk reduction of CRC related mortality was applied (all p-values <0.001). Both screening naïve and screened subjects were willing to undergo a ten-yearly TC instead of a five-yearly FS to obtain an additional risk reduction of CRC mortality of 25% ($p<0.001$).

Conclusions Assuming realistic test characteristics, subjects in the target population preferred endoscopic screening to FOBT screening primarily due to the more favourable risk reduction of CRC related mortality. Increasing knowledge on risk reduction by different screening strategies is therefore warranted to prevent unrealistic expectations and to optimise informed choice.

7.1 Introduction

Colorectal cancer (CRC) is the second cause of cancer-related death in the Western world. Screening can reduce CRC related mortality by removal of adenomas and early detection of CRC [1-5]. There is compelling evidence to support screening of average-risk individuals over 50 years of age [3-7]. Guidelines underline the role of individual preferences in the selection of a screening test [8-10], since insufficient evidence is available to recommend one screening test to another. Individual preferences for a certain screening test have been found to influence uptake in a CRC screening programme [11]. Uptake is a key factor determining the effectiveness of such a screening programme. However, uptake levels are fairly low in many countries (<60%) [3, 4, 12-14]. Several countries, including The Netherlands, are considering introduction of a nation-wide CRC screening programme. It is therefore essential to obtain insight into individual preferences for available screening strategies prior to the implementation of a nation-wide screening programme.

Previous surveys demonstrated a broad variation in preferences for CRC screening tests, since tests differ in benefit (CRC mortality reduction) on the one hand and potential harms on the other hand (perceived burden, complications). Subjects who valued effectiveness most highly chose for colonoscopy screening, whereas others preferred faecal occult blood testing (FOBT) because of the less invasive character [11, 15-17]. These studies however did not provide data on the relative importance of test characteristics on preferences, for example, how much potential health gain does a subject require to undergo invasive endoscopic screening?

Discrete choice experiments (DCEs) are becoming more widely used in health care research [18-22]. A DCE is capable of establishing preferences and to predict uptake in controlled experimental conditions, through responses to realistic and hypothetical scenarios. DCEs may be valuable for patient centred evaluations of health technologies [23].

This study was conducted to determine individuals' preferences and to predict uptake for CRC screening programmes with various screening tests, and the relative importance of different test characteristics for these preferences in an average risk population. Furthermore, we aimed to identify differences in preference structures between subgroups in the population.

7.2 Methods

7.2.1 Data sources

A total of 1,498 screening-naïve individuals aged 50-74 years old were randomly selected from municipal registries of the Rotterdam region in the Southwest of the Netherlands. We also invited a random sample 769 screened subjects of a CRC screening trial comparing guaiac-based FOBT (gFOBT), faecal immunochemical test (FIT) and flexible sigmoidoscopy (FS) (Figure 7.1). This screening trial was carried out in the same target population as mentioned above [24].

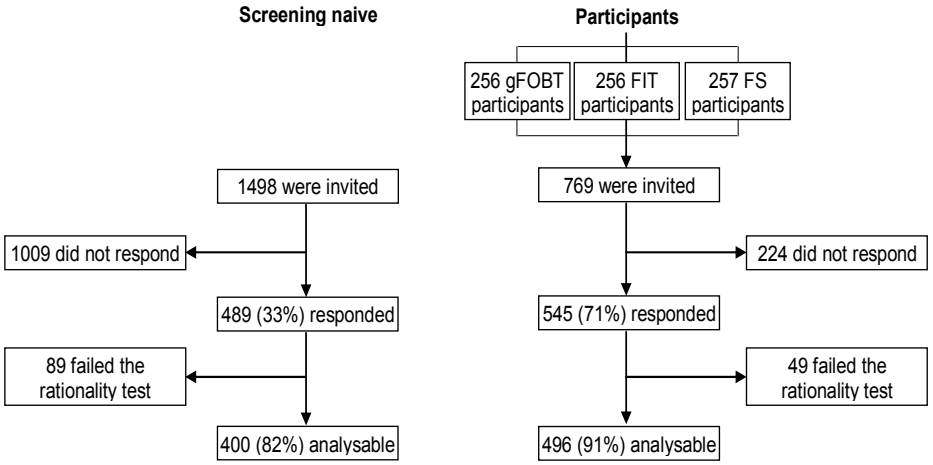


Figure 7.1 Study profile

7.2.2 Discrete choice experiments

DCEs can measure individuals' preferences for health care interventions. DCEs are based on the assumptions that a healthcare intervention can be described by its characteristics (attributes) (e.g. frequency of undergoing the intervention) and that the individual valuation of the intervention is determined by pre-defined levels (e.g. monthly, yearly) of those attributes. The health care intervention (e.g., screening test) as well as its test characteristics have to be specified before generating an experimental design. In a DCE individuals choose between several realistic and hypothetical scenarios. Preference estimates can be obtained from the choice data and describe the relative preference for characteristics of the health care intervention.

7.2.3 Attributes and levels

FOBT, FS and total colonoscopy (TC) are most widely used as CRC screening test and therefore incorporated in this study design. The characteristics and their levels were derived from the literature, expert opinions (n=3) and interviews with potential screenees (n=40). Experts were asked to comment on a list of characteristics derived from literature review. Potential screenees could also comment on the list of characteristics and rank them in order of importance. Based on these data we selected the two most important characteristics as identified by both groups: risk reduction of CRC related mortality (RR) and screening interval. Noteworthy, characteristics that are directly related to the test (e.g. oral bowel cleansing solution is not required for FOBT and always for TC) were already

captured by the specific screening test (FOBT, FS and TC). All subjects were informed about the incorporated test characteristics of the three screening tests (Appendix 7.A). The specific values (levels; e.g. amount of risk reduction, or length of screening interval) for each test characteristic incorporated the range of possible test outcomes of a specific screening test (FOBT, FS and TC) based on the current literature (Table 7.1). The levels were test specific to create realistic scenarios (Table 7.1).

7.2.4 Study design and questionnaire

The design contained three tests (FOBT, FS and TC) and two characteristics (risk reduction of CRC related mortality and screening interval) with three levels each (Table 7.1). The test specific levels (e.g. screening interval of FOBT between four months and triennial) were required to select realistic combinations. Furthermore, unrealistic combinations of the characteristics' levels were blocked (i.e. a combination of the lowest RR with the shortest screening interval as well as the highest RR combined with the longest screening interval). The combination of the characteristics and levels resulted in 21 (i.e. 7*3) possible test scenarios, and thus 343 (i.e. 7^3) possible combinations of the scenarios (i.e. full factorial design). It is not feasible to present a single individual with all these combinations. We therefore reduced the design in such away that two-way interactions could be estimated (i.e. we created a fractional factorial design). We therefore used SAS software (Version 9.1, SAS Institute Inc., Cary, NC, USA), which is capable of generating designs that are highly efficient (i.e. maximizing D-efficiency or minimizing D-error) in such circumstances [25]. We chose a design with 84 choice sets divided over 7 versions of the questionnaire (D-error 0.573). Each choice set included two CRC screening tests and an option not to be screened (opt-out) (Appendix 7.B).
A rationality test was included in the questionnaire to determine the understanding of the questionnaire by each subject. This test was a choice set of which one screening option was logically preferable to the other option given the levels of each test characteristic

Table 7.1 Alternatives, attributes and the alternative specific levels based on the literature

Alternatives	Alternative specific levels	Literature	References
Screening interval (yr)			
FOBT	1/4 - 1 - 3	1 - 2	
Sigmoidoscopy	1 - 5 - 10	5 - 10	[8, 28, 29]
Colonoscopy	2 - 5 - 10	5 - 10	
Risk reduction (%)			
FOBT	10 - 25 - 40	13 - 33	[3-5]
Sigmoidoscopy	40 - 50 - 70	49 - 62	[1, 2, 28, 30]
Colonoscopy	75 - 85 - 95	80 - 84	[28, 31, 32]

(biennial FS screening resulting in 40% RR against biennial FS screening resulting in 70% RR). It is common practice to exclude irrational responses from the analysis [26-28], and that was why this approach was adopted here. However, some recent discussions in the literature suggest that these responses could be included [29, 30]. Further sensitivity analyses were conducted. Inclusion of irrational responses led to similar results.
Subjects' social economic status (SES), previous endoscopy experience and experience with CRC in family and close friends were determined. Furthermore, the generic health status (EQ-5D summary score) was assessed. This is a validated classification of subject's own health [31].
We conducted a pilot study (n=20) to ascertain subjects could manage the length of the questionnaire and to examine the intelligibility, acceptability and validity. The questionnaire was mailed to all subjects. Background information on the three screening tests (Appendix 7.A) was printed on the first page of the questionnaire. A reminder was sent to non-responders four weeks later.

7.2.5 Data Analysis

Each choice between two tests and the opt-out was considered as a specific observation. The DCE was analysed using multinomial logit regression models with test specific parameters. The model was implemented in SAS software (Version 9.1, SAS Institute Inc., Cary, NC, USA). A priori we expected the test as well as the two characteristics to be important for subjects' choices and that a higher risk reduction would have a positive effect on preferences and lengthening of 'screening interval' a negative effect.
We assumed that there was no linear relationship between the different levels of the characteristics. On this basis, we estimated the following models for the DCE:

$$U_{\text{no test}} = 0$$

$$U_{FOBT} = \beta_0 + \beta_1 \text{Interval1yr} + \beta_2 \text{Interval3yr} + \beta_3 \text{RR25} + \beta_4 \text{RR40} + \varepsilon_{FOBT}$$

$$U_{FS} = \beta_5 + \beta_6 \text{Interval5y}_r + \beta_7 \text{Interval10yr} + \beta_8 \text{RR50} + \beta_9 \text{RR70} + \varepsilon_{FS}$$

$$U_{TC} = \beta_{10} + \beta_{11} \text{Interval5yr} + \beta_{12} \text{Interval10yr} + \beta_{13} \text{RR85} + \beta_{14} \text{RR95} + \varepsilon_{TC}$$

Utility (U) represents the preference score for a (hypothetical) CRC screening programme consisting of the deterministic and observable component (V) and the random component (ε) to the analysis, accounting for unobserved or unobservable components of choice. The constant terms (screening test; β_0, β_5, β_{10}) are alternative specific constants that indicate the general attitude of subjects towards screening with a specific screening test compared to no screening. $\beta_{1,2}$, $\beta_{6,7}$, $\beta_{11,12}$ are coefficients of the levels of the test characteristic 'screening interval' and $\beta_{3,4}$, $\beta_{8,9}$, $\beta_{13,14}$ are coefficients of the levels of the test characteristic 'risk reduction of CRC related mortality'; each coefficient indicates the relative weight individuals place on that test specific level compared with the reference

level for that test specific test characteristic (for the reference levels see Table 7.3). A two-sided p-value smaller than 0.05 was considered statistically significant.
Generic health status was dichotomized to an EQ-5D summary score of '1', representing full health, versus an EQ-5D summary score '<1', indicating sub-optimal health. Aggregate data on socio-economic status (SES) were available at the level of the area postal code (www.cbs.nl) of the subject, weighted by population size and classified into three groups (high, intermediate, low).
Chi-square and Student t-tests were used to assess the differences in the value of characteristics between screening naïve and screened subjects as well as between subgroups (age, gender, SES, EQ-5D, prior endoscopy experience, or knowing someone affected by CRC) within the screening-naïve population.
To examine the predicted uptake of CRC screening based on our results, we applied previously proposed models to our data [32, 33]. We also investigated the effect of changing the characteristics, as identified by the results of our multinomial logit model, on the expected uptake of CRC screening.

7.2.6 Ethical approval

The study was approved by the Institutional Review Board of the Erasmus MC, University Medical Centre (MEC-2007-224).

Table 7.2 Subjects' characteristics

	Screening naïve	**Screened subjects**	**p-value**
Invited subjects (n)	1,498	769	
Sex (male; n-%)	209 (52)	260 (52)	0.96
Age (mean-SD)	60.7 (6.6)	61.1 (6.4)	0.36
EQ5D score (mean-SD)	0.94 (0.11)	0.93 (0.10)	0.76
Social economic status (n-%)			<0.01
High	195 (49)	196 (40)	
Intermediate	77 (19)	96 (19)	
Low	128 (32)	204 (41)	
Endoscopy experience (n-%)			<0.01
Yes	92 (23)	242 (49)	
No	307 (76)	251 (50)	
Unknown	1 (1)	3 (1)	
Related to CRC patient (n-%)			0.78
Yes	53 (13)	67 (13)	
No	285 (71)	381 (77)	
Unknown	62 (16)	48 (10)	

7.3 Results

A total of 489/1,498 (33%) screening-naïve and 545/769 (71%) screened subjects returned the questionnaire. Screening-naïve subjects were of higher SES than screened subjects ($p<0.001$, Table 7.2). A higher proportion of screened subjects previously underwent an endoscopy compared to screening-naïve subjects (49% vs. 23%; $p<0.001$). Among the subjects that participated in the CRC screening trial 22% (70/324) of the screenees that performed a FOBT previously underwent an endoscopy and obviously all (172/172) FS screenees.

7.3.1 DCE

A significantly higher proportion of the screened subjects (91%) passed the rationality test compared to the screening-naïve subjects (82%; $p<0.001$).
Screening-naïve subjects did not prefer FOBT to no screening. They expressed a positive attitude towards FS and TC (positive and statistically significant sign, Table 7.3, Figure 7.2). A high RR was preferred to intermediate and low RR for all screening tests (p-values <0.01). Screening-naïve subjects expressed a more positive attitude towards an intermediate (FOBT: annually; FS: five-yearly; TC: five-yearly) compared to a short screening interval (FOBT: three monthly; FS: annual; TC: biennial). Further lengthening of the screening interval (FOBT: triennial; FS: ten-yearly; TC: ten-yearly) had only a small positive effect on subjects' preferences for FOBT ($p=0.02$) and FS ($p=0.02$), and no effect on subjects' preferences for TC screening.
Screened subjects had a positive attitude towards all screening tests ($p<0.001$). A high RR was preferred to intermediate and low RR for all screening tests, and an intermediate screening interval was preferred to a short screening interval (Table 7.3, Figure 7.2). Screened subjects did not prefer an intermediate to a long interval for all screening tests (FOBT $p=0.67$; FS $p=0.99$; TC $p=0.10$).

7.3.2 Screening-naïve versus screened subjects

Screened subjects had a more positive attitude towards all screening tests than screening-naïve subjects (Table 7.3, $p<0.001$). The differences in preferences regarding RR and screening interval between screening-naïve and screened subjects were statistically not significant, except for preferences regarding five- and ten-yearly FS screening. The more positive attitude of screening-naïve subjects towards longer screening intervals (Five-yearly $p<0.001$; ten-yearly $p<0.001$) indicated that screening naïve-subjects valued infrequent screening more positively than screened subjects.

Table 7.3 Coefficients of the different tests and test characteristics

	Screening naïve		Screened subjects		
Attribute levels	Coeff	(95%CI)	Coeff	(95%CI)	p-value
Screening test (ref level 'no screening')					
FOBT	-0.18	(-0.44-0.08)	0.38	(0.15-0.62)*	<0.001
Sigmoidoscopy	0.30	(0.06-0.54)*	0.94	(0.72-1.16)*	<0.001
Colonoscopy	0.33	(0.08-0.57)*	1.05	(0.84-1.27)*	<0.001
Risk reduction of CRC related mortality					
FOBT					
3% to 2.7% (RR 10%) *(ref level)*					
3% to 2.4% (RR 25%)	0.19	(-0.01-0.38)	0.17	(-0.01-0.34)	0.45
3% to 1.8% (RR 40%)	0.78	(0.54-1.02)*	0.65	(0.44-0.87)*	0.88
Sigmoidoscopy					
3.0 to 1.8% (RR 40%) *(ref level)*					
3.0 to 1.5% (RR 50%)	0.10	(-0.09-0.29)	0.33	(0.16-0.50)*	0.97
3.0 to 0.9% (RR 70%)	0.65	(0.42-0.89)*	0.65	(0.44-0.86)*	0.08
Colonoscopy					
3.0 to 0.8% (RR 75%) *(ref level)*					
3.0 to 0.5% (RR 85%)	0.16	(-0.03-0.35)	0.19	(0.02-0.36)*	0.95
3.0 to 0.1% (RR 95%)	0.40	(0.17-0.62)*	0.41	(0.20-0.61)*	0.79
Screening interval					
FOBT					
Four-monthly *(ref level)*					
Annual	0.73	(0.52-0.93)*	0.64	(0.44-0.83)*	0.50
Triennial	0.96	(0.72-1.20)*	0.67	(0.46-0.89)*	0.07
Sigmoidoscopy					
Annual *(ref level)*					
Five-yearly	0.92	(0.74-1.11)*	0.55	(0.39-0.72)*	<0.001
Ten-yearly	1.14	(0.91-1.37)*	0.56	(0.36-0.75)*	<0.001
Colonoscopy					
Biennial *(ref level)*					
Five-yearly	0.71	(0.52-0.90)*	0.56	(0.39-0.73)*	0.22
Ten-yearly	0.72	(0.48-0.95)*	0.42	(0.21-0.63)*	0.06

7.3.3 Differences in preferences between subgroups

No differences in preferences were found between men and women, apart from a more positive attitude towards FS and TC among men (FS p=0.06; TC p=0.02). Men, in contrast to women, did prefer FS and TC to no screening (men: FS p<0.001; TC p<0.001; women FS p=0.07; TC p=0.84). Respondents' age, SES and EQ-5D summary score did not influence the attitude towards a screening test, interval or RR. Subjects who

reported to have a close friend or family member with CRC expressed a more positive attitude towards TC screening than subjects without (p=0.01). Experience with FS or TC was positively associated with the willingness to undergo a TC (p<0.001). Subjects that underwent FS screening had a more positive attitude towards FS and TC screening than subjects who performed a FOBT (p<0.001).

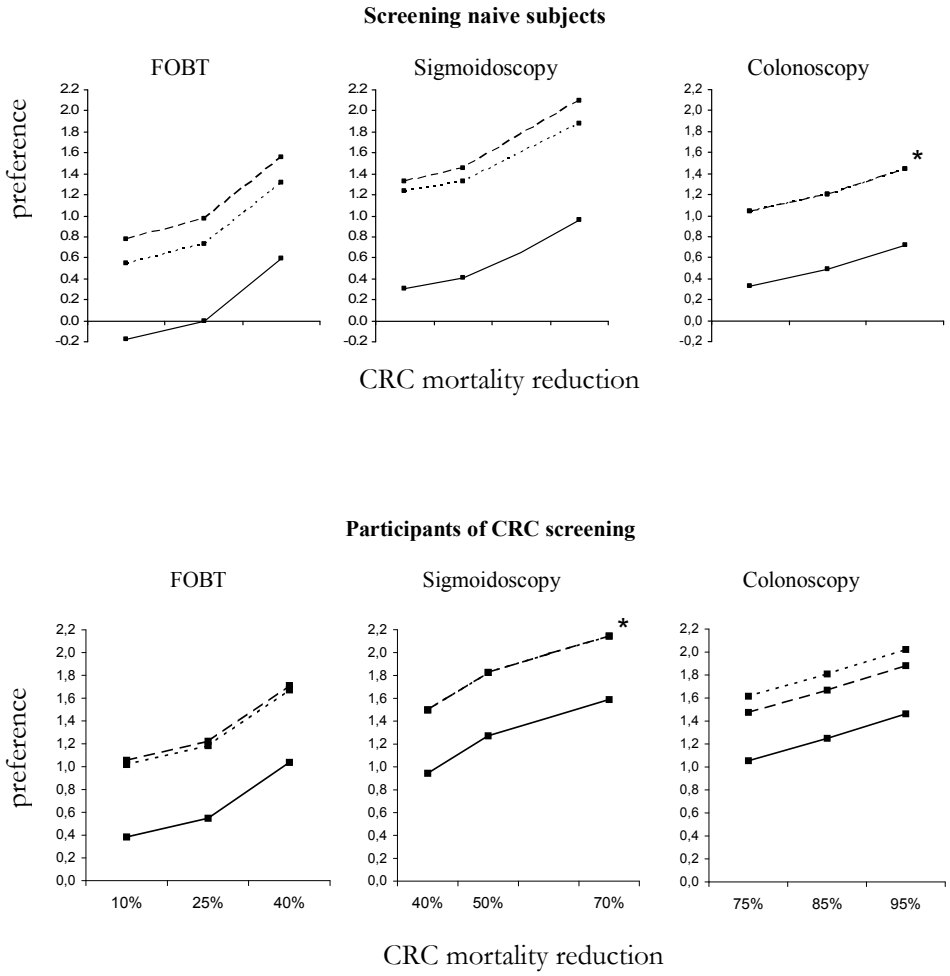


Figure 7.2 Preferences for the different screening strategies at a long (--), intermediate (···) and short (—) screening interval and different levels of mortality risk reduction for screening-naïve and screened subjects.

*Preferences for long and intermediate screening interval were similar

7.3.4 Trade-offs

Screening-naïve subjects were, when assuming the same interval (annual) and RR (40%), more willing to undergo FOBT than FS screening (preferences: FOBT=1.32; FS=0.30; p<0.001). Preferences were similar for a five-yearly FS and an annual FOBT if both tests would generate a RR of 40% (preferences: FOBT=1.32; FS=1.23; p=0.40). A five-yearly FS was preferred to annual FOBT if FOBT was associated with a less favourable RR than FS screening (preferences: FOBT 25%RR = 0.73; FS 40%RR = 1.23; p<0.001).
A five-yearly FS was preferred to a ten-yearly TC if the difference in RR was 25% in favour of TC (e.g. FS = RR 50%; TC =RR 75%; p<0.001). The preferences for a five-yearly FS and a ten-yearly TC were similar if TC would achieve an additional 35% RR (p=0.24), while more than 45% difference in RR was associated with a preference for 10-yearly TC (p<0.001).
Screening-naïve subjects equally preferred FS and TC screening, but did prefer both endoscopic screening options to FOBT screening if, based on the literature, the most realistic screening intervals and mortality reduction were applied (preferences: annual FOBT RR 25% = 0.77; five-yearly FS RR 50% = 1.33; ten-yearly colonoscopy RR 85% = 1.22 ; FS vs. FOBT p<0.001, TC vs. FOBT p<0.001; TC vs. FS p=0.24).
Screened subjects made similar trade-offs between the screening test, interval and RR as screening-naïve subjects.

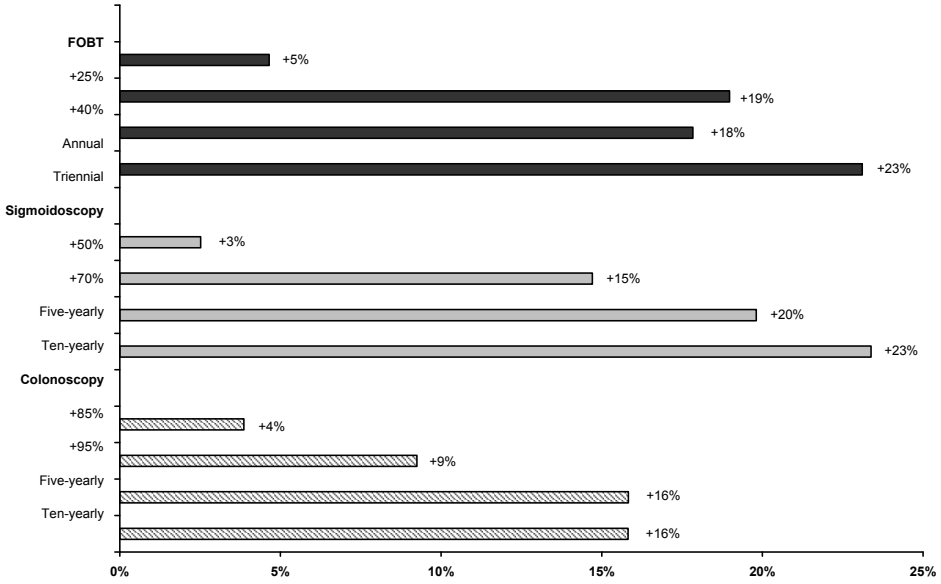


Figure 7.3 Effects of changing the screening programme characteristics on the average probability of uptake for respectively FOBT (45%), FS (58%) and TC (58%) in screening-naïve subjects.

7.3.5 Predicted uptake

Predicted uptake of screening naïve subjects for FOBT, FS and TC screening was 45%, 58% and 58% respectively, assuming screening with the reference level for RR and screening interval. Based on realistic screening intervals and mortality reduction from the literature, these numbers were 68% for FOBT, 79% for FS, and 81% for TC. The screening programme characteristics had substantial impact on the expected uptake among screening naïve subjects (Figure 7.3).

7.4 Discussion

7.4.1 Principle findings

In this population-based study we found that the type of screening test, screening interval, and risk reduction of CRC related mortality significantly influenced individual preferences among screening naïve and screening experienced subjects in the target population (aged 50-74 years old). Both screened and screening-naïve subjects preferred FS and TC to FOBT screening if, based on the literature [3-7, 9, 34, 35], the most realistic screening interval and risk reduction on CRC related mortality were applied (annual FOBT with 25% RR; five-yearly FS with 50% RR, ten-yearly colonoscopy with 85% RR). This underlines the importance of adequate information on those aspects of CRC screening to achieve informed decision-making by potential screenees.

Five studies investigated preferences in CRC screening using a DCE [18, 19, 36-38], with two studies investigating preferences among available screening tests [19, 36]. This is the first DCE including both a screening naïve and screening experienced population. In agreement with previous DCE studies, we found that RR dominated preferences for a screening test. Both FS and TC screening were therefore preferred to FOBT screening when associated with sufficient RR [19, 36].

The literature on preferences for the optimal screening interval per test is limited. One study reported a preference for five or ten-yearly to annual screening irrespective of the screening test [36]. However, deciding on screening interval without information on the screening test leads to unrealistic choices, since an annual FOBT is less burdensome than an annual TC. We therefore used test specific screening intervals, which add to the validity of our results. In our study, screened subjects equally preferred intermediate and long screening interval for all tests. Reassurance may be a reason for preferring frequent screening [39]. However, both intermediate and long interval of all three screening tests were preferred to a short interval, suggesting that subjects trade-off between reassurance and frequency of undergoing a screening test.

Men had a more positive attitude towards FS and colonoscopy screening than women. This finding is in accordance with FS screening programmes which described a lower uptake among women than among men [12, 13, 24]. Known barriers for women to participate in endoscopy screening are male endoscopists [40], and anxiety prior to

screening [41]. A different approach to inform both sexes on screening or sex-specific screening strategies might be considered in a nation-wide screening programme to improve acceptance.

The results of this study may be relevant to predict population preferences for newer screening tests with a similar profile or an improved version of a screening test. For example, recently randomised trials demonstrated more favourable detection rates for FIT than gFOBT [14, 24, 42] suggesting a larger reduction of CRC related mortality. According to our data, informing people in the target population about a more favourable effect on CRC related mortality of FIT would lead to a higher acceptance of FIT screening and most likely a higher uptake.

Predicted uptake of FS or TC screening based on our model was significantly higher than uptake of FOBT screening, given realistic levels. This finding is in contrast to the observed higher uptake of FOBT than FS screening in the randomised screening trial performed in the same population as this DCE. Screenees in this trial were however not specifically informed on test efficacy. This suggests that increasing awareness on the efficacy of a screening test may enhance uptake. It is therefore of paramount importance to improve the level of awareness on achievable risk reduction of CRC related mortality to obtain a higher uptake, especially for the more effective endoscopic screening tests. This is further underlined by two European studies. A Swiss study [43], in which the majority (75%) of all screenees chose to undergo a TC, and only a small proportion (25%) preferred FOBT or FS screening after they were informed about the efficacy of the three screening tests. A large population-based Italian study found simialar participation rates for FS and FOBT when subjects were offered a choice between both strategies [44].

7.4.2 Strengths and weaknesses of this study

In contrast to previous DCE studies we used a labelled instead of an unlabelled DCE design. In a labelled design the specific screening test is mentioned in each choice option (FOBT, FS, TC; Appendix 7.B), while in an unlabelled design the screening test is presented as 'screening test A', 'B' or 'C' and is further described by certain characteristics that are presented in the choice set. CRC screening tests may evoke individual feelings, which can not be described in a questionnaire (e.g. anxiety for an endoscopy). It is therefore difficult to adequately convey the essential differences from a subject's perspective between FOBT and endoscopic tests in terms of, for example, 'more burdensome' or 'less burdensome'. Using a labelled design, the scenarios are more realistic, which adds to the validity of the results. Furthermore, we assessed preferences among screening-naïve and screened subjects within the target population (aged 50-74 yrs old) including all social economic classes, which adds to the generalisability of the results. Experienced subjects stated a more positive attitude towards all screening tests than screening-naïve subjects. A selection bias may explain this difference in attitude, as experienced subjects have already demonstrated interest in screening and therefore express a more positive attitude towards screening. There is however also an experience

effect, i.e. anticipated discomfort and pain might be higher than actually experienced. This experience might reduce anticipated pain and discomfort for successive screening round. Additionally, there may also be an expose effect, i.e. people tend to develop a preference merely because they are familiar with it. Our results suggest that subjects who underwent screening are willing to return for a successive screening round, which is of vital importance for efficacy of a screening programme. Costs of screening were not included as a test characteristic in this study. All CRC screening programmes in Europe including the Netherlands do not require out-of-pocket costs. Including cost would therefore influence the results in an unrealistic manner. A limitation of this study is the significantly lower response rate in screening-naïve than in screened subjects. This may have led to selection bias and therefore a more positive attitude towards screening in the latter group.

7.4.3 Conclusions

Both screening-naïve and screened subjects stated a more positive attitude towards both endoscopic screening strategies than FOBT if, based on the literature, the most realistic screening interval and risk reduction on CRC related mortality were applied. Risk reduction of CRC related mortality mainly determined the preference for endoscopic screening tests. This underlines the importance of awareness on achievable risk reduction of CRC related mortality of the different screening test to enhance uptake particularly for endoscopic screening tests and to optimise informed choice.

References

1. Selby JV, Friedman GD, Quesenberry CP, Jr., Weiss NS (1992). A case-control study of screening sigmoidoscopy and mortality from colorectal cancer. N Engl J Med 326(10):653-657.
2. Newcomb PA, Norfleet RG, Storer BE, Surawicz TS, Marcus PM (1992). Screening sigmoidoscopy and colorectal cancer mortality. J Natl Cancer Inst 84(20):1572-1575.
3. Kronborg O, Fenger C, Olsen J, Jorgensen OD, Sondergaard O (1996). Randomised study of screening for colorectal cancer with faecal-occult-blood test. Lancet; 348(9040):1467-1471.
4. Hardcastle JD, Chamberlain JO, Robinson MH, Moss SM, Amar SS, Balfour TW et al (1996). Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. Lancet; 348(9040):1472-1477.
5. Mandel JS, Bond JH, Church TR, Snover DC, Bradley GM, Schuman LM et al (1993). Reducing mortality from colorectal cancer by screening for fecal occult blood. Minnesota Colon Cancer Control Study. N Engl J Med; 328(19):1365-1371.
6. Kewenter J, Brevinge H, Engaras B, Haglind E, Ahren C (1994). Results of screening, rescreening, and follow-up in a prospective randomized study for detection of colorectal cancer by fecal occult blood testing. Results for 68,308 subjects. Scand J Gastroenterol; 29(5):468-473.
7. Hoff G, Grotmol T, Skovlund E, Bretthauer M (2009). Risk of colorectal cancer seven years after flexible sigmoidoscopy screening: randomised controlled trial. BMJ; 338:b1846.

8. Sung JJ, Lau JY, Young GP, Sano Y, Chiu HM, Byeon JS et al (2008). Asia Pacific consensus recommendations for colorectal cancer screening. Gut; 57(8):1166-1176.
9. Levin B, Lieberman DA, McFarland B, Andrews KS, Brooks D, Bond J et al (2008). Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. Gastroenterology; 134(5):1570-1595.
10. Council Recommendation on Cancer Screening (2003). 2003/0093. Commission of the European Communities Brussels.
11. Wolf RL, Basch CE, Brouse CH, Shmukler C, Shea S (2006). Patient preferences and adherence to colorectal cancer screening in an urban population. Am J Public Health; 96(5):809-811.
12. Single flexible sigmoidoscopy screening to prevent colorectal cancer: baseline findings of a UK multicentre randomised trial. (2002) Lancet; 359(9314):1291-1300.
13. Segnan N, Senore C, Andreoni B, Aste H, Bonelli L, Crosta C et al (2002). Baseline findings of the Italian multicenter randomized controlled trial of "once-only sigmoidoscopy"--SCORE. J Natl Cancer Inst; 94(23):1763-1772.
14. van Rossum LG, Van Rijn AF, Laheij RJ, van Oijen MG, Fockens P, Van Krieken HH et al (2008). Random Comparison of Guaiac and Immunochemical Fecal Occult Blood Tests for Colorectal Cancer in a Screening Population. Gastroenterology.
15. Frew EJ, Wolstenholme JL, Whynes DK (2005). Eliciting relative preferences for two methods of colorectal cancer screening. Eur J Cancer Care (Engl); 14(2):124-131.
16. Pignone M, Bucholtz D, Harris R (1999). Patient preferences for colon cancer screening. J Gen Intern Med; 14(7):432-437.
17. DeBourcy AC, Lichtenberger S, Felton S, Butterfield KT, Ahnen DJ, Denberg TD (2008). Community-based preferences for stool cards versus colonoscopy in colorectal cancer screening. J Gen Intern Med; 23(2):169-174.
18. Gyrd-Hansen D, Sogaard J (2001). Analysing public preferences for cancer screening programmes. Health Econ; 10(7):617-634.
19. Marshall DA, Johnson FR, Phillips KA, Marshall JK, Thabane L, Kulin NA (2007). Measuring patient preferences for colorectal cancer screening using a choice-format survey. Value Health; 10(5):415-430.
20. Marshall DA, Johnson FR, Kulin NA, Ozdemir S, Walsh JM, Marshall JK et al (2009). How do physician assessments of patient preferences for colorectal cancer screening tests differ from actual preferences? A comparison in Canada and the United States using a stated-choice survey. Health Econ.
21. Hur C, Broughton DE, Ozanne E, Yachimski P, Nishioka NS, Gazelle GS (2008). Patient preferences for the chemoprevention of esophageal adenocarcinoma in Barrett's esophagus. Am J Gastroenterol; 103(10):2432-2442.
22. Sculpher M, Bryan S, Fry P, de WP, Payne H, Emberton M (2004). Patients' preferences for the management of non-metastatic prostate cancer: discrete choice experiment. BMJ; 328(7436):382.
23. Ryan M (2004). Discrete choice experiments in health care. BMJ; 328(7436):360-361.
24. Hol L, van Leerdam ME, van Ballegooijen M, van Vuuren AJ, reijerink-verheij JC, van der Togt-van Leeuwen AC et al (2009). Screening for colorectal cancer; randomised trial comparing guaiac-based and immunochemical faecal occult blood testing and flexible sigmoidoscopy. Gut; in press.
25. Street DJ, Burgess L, Louviere JJ (2005). Constructing. Optimal and Nearly Optimal Stated Choice Experiments. Intern J of Research in Marketing; 22:459-470.
26. Ryan M, Major K, Skatun D (2005). Using discrete choice experiments to go beyond clinical outcomes when evaluating clinical practice. J Eval Clin Pract; 11(4):328-338.
27. Weston A, Fitzgerald P (2004). Discrete choice experiment to derive willingness to pay for methyl aminolevulinate photodynamic therapy versus simple excision surgery in basal cell carcinoma. Pharmacoeconomics; 22(18):1195-1208.

28. Langenhoff BS, Krabbe PF, Ruers TJ (2007). Computer-based decision making in medicine: A model for surgery of colorectal liver metastases. Eur J Surg Oncol; 33 Suppl 2:S111-S117.
29. Ryan M, Watson V, Entwistle V (2009). Rationalising the 'irrational': a think aloud study of discrete choice experiment responses. Health Econ; 18(3):321-336.
30. Lancsar E, Louviere J (2006). Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? Health Econ; 15(8):797-811.
31. Dolan P (1997). Modeling valuations for EuroQol health states. Med Care; 35(11):1095-1108.
32. Hall J, Kenny P, King M, Louviere J, Viney R, Yeoh A (2002). Using stated preference discrete choice modelling to evaluate the introduction of varicella vaccination. Health Econ; 11(5):457-465.
33. Gerard K, Shanahan M, Louviere J (2008). In: Ryan M, Gerard K, Amaya-Amaya M, editors. Using Discrete Choice Experiments to Value Health and Health Care. Dordrecht: Springer; 117-137.
34. Winawer SJ, Zauber AG, Ho MN, O'Brien MJ, Gottlieb LS, Sternberg SS et al (1993). Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. N Engl J Med; 329(27):1977-1981.
35. Faivre J, Dancourt V, Lejeune C, Tazi MA, Lamour J, Gerard D et al (2004). Reduction in colorectal cancer mortality by fecal occult blood screening in a French controlled study. Gastroenterology; 126(7):1674-1680.
36. Hawley ST, Volk RJ, Krishnamurthy P, Jibaja-Weiss M, Vernon SW, Kneuper S (2008). Preferences for colorectal cancer screening among racially/ethnically diverse primary care patients. Med Care; 46(9 Suppl 1):S10-S16.
37. Howard K, Salkeld G (2008). Does Attribute Framing in Discrete Choice Experiments Influence Willingness to Pay? Results from a Discrete Choice Experiment in Screening for Colorectal Cancer. Value Health.
38. Salkeld G, Solomon M, Short L, Ryan M, Ward JE (2003). Evidence-based consumer choice: a case study in colorectal cancer screening. Aust N Z J Public Health; 27(4):449-455.
39. Cantor SB, Volk RJ, Cass AR, Gilani J, Spann SJ (2002). Psychological benefits of prostate cancer screening: the role of reassurance. Health Expect; 5(2):104-113.
40. Menees SB, Inadomi JM, Korsnes S, Elta GH (2005). Women patients' preference for women physicians is a barrier to colon cancer screening. Gastrointest Endosc; 62(2):219-223.
41. Farraye FA, Wong M, Hurwitz S, Puleo E, Emmons K, Wallace MB et al (2004). Barriers to endoscopic colorectal cancer screening: are women different from men? Am J Gastroenterol; 99(2):341-349.
42. Hol L, Wilschut JA, van BM, van Vuuren AJ, van D, V, Reijerink JC et al (2009). Screening for colorectal cancer: random comparison of guaiac and immunochemical faecal occult blood testing at different cut-off levels. Br J Cancer; 100(7):1103-1110.
43. Marbet UA, Bauerfeind P, Brunner J, Dorta G, Valloton JJ, Delco F (2008). Colonoscopy is the preferred colorectal cancer screening method in a population-based program. Endoscopy; 40(8):650-655.
44. Segnan N, Senore C, Andreoni B, Arrigoni A, Bisanti L, Cardelli A et al (2005). Randomized trial of different screening strategies for colorectal cancer: patient response and detection rates. J Natl Cancer Inst; 97(5):347-357.
45. Frazier AL, Colditz GA, Fuchs CS, Kuntz KM (2000). Cost-effectiveness of screening for colorectal cancer in the general population. JAMA; 284(15):1954-1961.
46. Salkeld G, Solomon M, Short L, Ryan M, Ward JE (2003). Evidence-based consumer choice: a case study in colorectal cancer screening. Aust N Z J Public Health; 27(4):449-455.
47. Muller AD, Sonnenberg A (1995). Prevention of colorectal cancer by flexible endoscopy and polypectomy. A case-control study of 32,702 veterans. Ann Intern Med; 123(12):904-910.
48. Zauber AG, Winawer SJ, O'Brien MJ, shi W, Bayuga S (2007). Significant Long Term Reduction in Colorectal Cancer Mortality with Colonoscopic Polypectomy: Findings of the National Polyp Study. Gastroenterology; 132 (Suppl 2): A–50.

Appendix 7.A

Background information on all screening test as applied to all subjects.

	Faecal occult blood test	**Sigmoidoscopy**	**Colonoscopy**
Preparation	None.	- One or two enemas (bowel preparation). - No fasting.	- You have to drink 4 litres of special cleansing solution the day before the procedure. - You have to fast for 12 hours prior to the procedure. - You cannot work the afternoon prior to and the day of the procedure.
The procedure	**How do I carry out the test?** At home, several (1-3) samples of stool are collected by using the test set. The test can be returned by mail to the laboratory. **What does the test measure?** The test measures if there are (in)visible traces of blood present in the stools. **What happens if the test results are abnormal?** You will be advised to undergo a colonoscopy.	**The procedure** T*he last 60 cm of the large bowel* is examined by using a flexible tube with a small camera on the tip. This tube is inserted through the anus. During the procedure the large bowel will be filled with air in order to carefully examine the bowel. **What do I feel of the investigation?** Because of the air put into your bowel you may feel abdominal cramps. **What happens if abnormalities are found?** Precursors of colon carcinoma (polyps) are removed during the procedure (this is painless). You will be advised to undergo a colonoscopy to see if there are other abnormalities in the remainder large bowel.	**The procedure** You will be given conscious sedation ('short narcosis'). Therefore, you may fall into a light sleep. T*he entire large bowel (100-120 cm)* is examined by using a flexible tube with a small camera on the tip. This tube is inserted through the anus. During the procedure the large bowel will be filled with air in order to carefully examine the bowel. **What do I feel of the investigation?** Due to the air and tube in your bowel you may feel abdominal pressure and cramps. **What happens if abnormalities are found?** Precursors of colon carcinoma (polyps) are removed during the procedure (this is painless).
After the procedure	- You can return to your daily activities immediately.	- You may eat and drink again immediately and go home.	- You may eat and drink again and go home after one hour. - You cannot drive a car, ride a motorcycle or bicycle.
Perceived burden	Low.	High.	High.
Results	- You will receive the result by mail within two weeks.	- Directly after the procedure. - When tissue has been removed, you will receive the pathology results by mail within two weeks.	- Directly after the procedure. - When tissue has been removed, you will receive the pathology results by mail within two weeks.
Test at home or in the hospital	At home.	Hospital.	Hospital.
Total duration of the procedure	30 minutes.	15 minutes.	1 hour and 45 minutes.
Complications	Never.	In 1 in 10.000 individuals: severe blood loss or a perforation or a tear through the bowel wall.	In 1 in 1.000 individuals: severe blood loss or a perforation or a tear through the bowel wall.

Appendix 7.B

Choice set

Choice options:	The test you will be examined with:	In the following 10 years you will undergo the test:	The chance of dying from colon cancer decreases from:
A	Sigmoidoscopy	1x	3%  to 1,8% 
B	Colonoscopy	1x	3%  to 0,8% 
C	None	0x	3% 

Suppose screening for colon cancer is introduced. Which situation do you prefer?

(fill in: A, B or C) ☐

8

What determines individuals' preferences for colorectal cancer screening programmes? A discrete choice experiment

Van Dam L, Hol L, de Bekker-Grob EW, Steyerberg EW, Kuipers EJ, Habbema JDF, Essink-Bot ML, van Leerdam ME

European Journal of Cancer; *in press*

Abstract

Introduction In many countries uptake of colorectal cancer (CRC) screening remains low.

Aim To assess how procedural characteristics of CRC screening programmes determine preferences for participation and how individuals weigh these against the perceived benefits from participation in CRC screening.

Methods A discrete choice experiment was conducted among subjects in the age-group of 50 – 75 years, including both screening-naïve subjects as well as participants of a CRC screening programme. Subjects were asked on their on preferences for aspects of CRC screening programmes using scenarios based on: pain, risk of complications, screening location, preparation, duration of procedure, screening interval and risk reduction of CRC related death.

Results The response was 31% (156/500) for screening-naïve and 57% (124/210) for CRC screening participants. All aspects proved to significantly influence the respondents' preferences. For both groups combined, respondents required an additional relative risk reduction of CRC related death by a screening programme of 1% for every additional 10 minutes of duration, 5% in order to expose themselves to a small risk of complications, 10% to accept mild pain, 10% to undergo preparation with an enema, 12% to use 0.75 litres of oral preparation combined with 12 hours fasting and 32% to use an extensive bowel preparation. Of note, screening intervals shorter than 10 years were significantly preferred to a 10-year screening interval.

Conclusions This study shows that especially type of bowel preparation, risk reduction and length of screening interval influence CRC screening preferences, improving awareness on CRC mortality reduction by screening may increase CRC screening uptake.

8.1 Introduction

Colorectal carcinoma (CRC) is the second most frequently occurring malignancy in the European Union, and the second leading cause of cancer related death in the Western world [1]. A recent study demonstrates that for many European countries CRC mortality rates are decreasing while incidence is rising, suggesting an increasing CRC prevalence [2]. CRC screening is effective in reducing CRC mortality [3-11]. Screening can reduce CRC mortality by early detection of CRC and endoscopic removal of premalignant precursors of CRC (adenomas) [5, 11, 12]. There are several methods available for CRC screening. The various types of faecal occult blood tests (FOBTs) primarily aim at the early detection of CRC, whereas endoscopic and radiologic screening tests (flexible sigmoidoscopy (FS), colonoscopy) are effective at both early detection and removal of premalignant lesions [12]. Different screening methods are expected to have a different impact on CRC mortality reduction due to these differences in preventive potential. CRC screening methods also differ with respect to procedural characteristics, which determine the subject's burden of a screening method. CRC screening methods perceived as the most burdensome (FS, colonoscopy) also have the largest potential for prevention of CRC [12]. Currently, insufficient evidence is available to recommend one screening method over another.

Attendance is an important determinant of the effectiveness of CRC screening programmes. Uptake of CRC screening in a pilot screening programme in the Netherlands has remained lower than uptake of breast and cervical cancer screening [13-15]. In many countries, uptake of CRC screening, as well as continuing adherence to CRC screening, has also remained suboptimal [3, 4, 13, 16-18]. It has been established that increasing colorectal cancer screening uptake, in comparison with other targets, has a large potential for reducing CRC related mortality [19]. Attendance rates depend on the willingness of individuals to undergo a certain screening test. This willingness may be influenced by perceived advantages and drawbacks of CRC screening tests and furthermore, by knowledge and awareness of CRC, CRC risk and CRC screening [18, 20, 21]. Individuals may be willing to undergo a screening test despite several drawbacks in order to maximize health benefit or vice versa (to accept a lower health benefit in order to avoid several burdensome test characteristics). To optimise a CRC screening programme it is of paramount importance to gain insight in factors that influence population preferences for CRC screening programmes and the trade-offs individuals are willing to make between benefits and drawbacks of a CRC screening programme. Research has shown that patient preferences can have a major impact on their willingness to use services and furthermore, there is an increasing emphasis on involvement of patients in health care decisions [22].

This study therefore investigated preferences for CRC screening using a discrete choice experiment (DCE). DCE is a survey methodology with its origin in market research. DCEs are widely used for the assessment of preferences in transport and environmental economics and marketing research [23]. They are increasingly used for health care purposes [24, 25].

It has been demonstrated that awareness of CRC and CRC screening in the Netherlands

has remained low [21]. There is currently no organised CRC screening programme in the Netherlands, except for hereditary or familial CRC. A similar situation is encountered in many countries in the EU, in fact, only approximately 50% of the target population is offered any type of screening for CRC. It is of particular importance to study preferences in a screening naïve population, since they may guide the introduction and adjustment of new CRC screening programmes in these countries.
The aim of our study was to determine how procedural characteristics of various CRC screening methods determine preferences for participation, and how individuals weigh these against the expected health benefits from CRC screening. We compared the relative importance of aspects of the three most commonly used CRC screening tests: FOBT, FS and colonoscopy.

8.2 Materials and methods

8.2.1 Study population

We conducted the study in two groups. The first group included a total of 500 screening-naïve individuals aged 50-74 years old who were randomly selected from the population registry of the region Rijnmond in the Southwest of the Netherlands. The region includes Rotterdam and surrounding suburbs and harbours 338,000 inhabitants in the target age groups. The second group included 210 participants of a randomised screening trial for CRC in the Netherlands from the same target population as mentioned above. This screening trial invited average risk individuals to participate in a CRC screening programme with guiac-based FOBT (gFOBT), faecal immunochemical test (FIT) or FS [26].

8.2.2 Invitation of subjects

Subjects were contacted by mail. They received a questionnaire and an information brochure with general and background information about CRC and CRC screening. Individuals could return the questionnaire in a postage-paid self-addressed envelope that was included in the mailing package. A reminder was sent four weeks later in case of non-response.

8.2.3 DCE

DCE is a formal technique to assess preferences, assuming that a healthcare intervention (e.g. a screening programme) can be described by its characteristics (attributes; e.g. test duration) [27]. Those attributes are further specified by variants of that attribute (levels; for test duration: 10, 20, 30 minutes). The DCE assumes that the individual preference for a test is determined by the levels of those attributes [27]. Individuals are presented with

a number of choice sets containing several scenarios (screening programmes). Those programmes are described by several attributes with varying levels (Figure 8.1). The results of a DCE provide information on the relative importance of the attributes and the trade-offs individuals are willing to make between these attributes. The DCE design will be explained in more detail further on.

8.2.4 Attributes and attribute levels

The attributes and attribute levels of the DCE were derived from literature review, expert opinions, interviews with screening naïve (n=10) and screened (n=10) individuals of the target population. In the interviews we asked individuals to point out which of these attributes they expected to be important or had been important in their decision to participate in a CRC screening programme. The attributes identified as most relevant were: pain, risk of complications, location of the screening test, preparation for the procedure, duration of the procedure, screening interval and risk reduction of CRC related death (Table 8.1). Attribute levels were derived from the literature. The levels for each attribute incorporated the range of characteristics or possible test outcomes of all different screening methods (FOBT, FS and colonoscopy). The attribute 'interval' was related to a CRC screening programme, the other attributes were test-related.

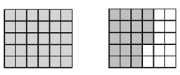
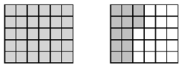
Choice options:	**Screening A**	**Screening B**	**No screening**
Preparation:	Enema, no fasting	Drinking 0.75l of fluid, 12 hours fasting	None
Location:	At home	Hospital	None
Pain:	None	Mild pain	None
Risk of complications:	None	Small	None
The chance of dying from colon cancer decreases from:	3% to 1.8% 	3% to 1.2% 	remains 3% 
In the following 10 years you will undergo the test:	5x	2x	0x
Duration:	30 min	60 min	None
Which screening programme do you prefer?	□ **A**	□ **B**	□ **None**

Figure 8.1 Choice set example

Table 8.1 Attributes and levels for CRC screening

Attributes and levels	Coefficient in regression analysis
Pain	
No pain (reference level)	
Mild pain	β_1
Risk of complications	
None (reference level)	
Small	β_2
Location	
At home (reference level)	
Hospital	β_3
Preparation	
None (reference level)	
Enema, no fasting	β_4
Drinking of 0.75l of fluid, 12h fasting	β_5
Drinking of 4l of fluid, 18h fasting	β_6
Duration	β_7
10 min	
30 min	
60 min	
90 min	
Interval	
1x in 10 years (reference level)	
2x in 10 years	β_8
5x in 10 years	
10x in 10 years	β_{10}
Risk reduction of death from CRC	β_{11}
3% to 2.7% (10% relative risk reduction)	
3% to 1.8% (40% relative risk reduction)	
3% to 1.2% (60% relative risk reduction)	
3% to 0.3% (90% relative risk reduction)	

8.2.5 Study design and questionnaire

The design contained three attributes with two levels and four attributes with four levels. The combination of those attributes and levels resulted in 2048 (i.e. $2^3 * 4^4$) possible test scenarios. Since it is not feasible to present a single individual with all these scenarios, we reduced the model to 16 scenarios (a fractional factorial design) by means of a website, containing a library of orthogonal arrays [28]. These 16 scenarios were used to create 16 choice sets. Each choice set contained two screening programmes and an opt-out (the option to choose 'no screening', see Figure 8.1). A special technique (fold-over [29]) was used to create the second programme of each choice set. As a result, our design was an

efficient orthogonal design; there was no correlation between any pairs of attributes (orthogonality), all levels of each attribute were represented in the same frequency (level balance), and similar levels of an attribute did not occur within the same choice set (minimal overlap). A rationality test was included in the DCE to investigate the understanding of the questionnaire. This was a choice set of which one screening programme was logically preferable over the other given the attribute levels.
The questionnaire further contained questions on background variables (e.g. generic health status (EQ-5D [30])) and a question assessing experienced difficulty of the questionnaire (5-point scale). A written description of the attributes and levels was given at the beginning. We conducted a pilot study (n=20) to ascertain respondents could manage the length of the questionnaire and to examine the intelligibility, acceptability and validity of the questionnaire.
The study was approved by the Medical Ethical Committee of the Erasmus MC (MEC, MEC-2007-224).

8.2.6 Analyses

Each choice between three options (two screening programmes and the opt-out) was considered as a specific observation. A multinomial logit model was used to analyse the data. Individuals who failed the rationality test were not excluded from analyses.
We assumed that there was no linear relationship between the different levels of the attributes 'preparation' and 'screening interval' and that all attributes had independent effects on preferences. On this basis, we estimated the following model for the DCE:

$$U = V + \varepsilon = \beta_0 + \beta_1 \text{pain} + \beta_2 \text{complications} + \beta_3 \text{location} + \beta_4 \text{enema} + \beta_5\ 0{,}75\text{lfluid} + \beta_6\ 4\text{lfluid} + \beta_7 \text{duration} + \beta_8 \text{interval2} + \beta_9 \text{interval5} + \beta 10\ \text{interval10} + \beta_{11} \text{mortalityreduction} + \varepsilon$$

U represents the preference for a (hypothetical) CRC screening programme. V ($\beta_{0.....}\beta_{11}$ mortalityreduction) is the observable utility that is composed of the preference scores for the individual β-coefficients of the model. The constant term (screening programme; β_0) is an 'alternative specific constant' and indicates the relative weight individuals place on screening programmes compared to no screening. β_1-β_{11} are coefficients of the attributes indicating the relative weight individuals place on a certain attribute(level). ε is the random component to the analysis, accounting for unobserved or unobservable components of choice. The value of each coefficient represents the importance respondents assign to a certain level. However, different attributes utilise different units of measurement. For example, the coefficient for 'risk reduction of death from CRC' represents the importance per relative 10% risk reduction. When looking at a screening programme that generates a

50% risk reduction, the coefficient should be multiplied five times in order to enable comparison to the coefficients of other levels. An attribute with a two sided p-value smaller than 0.05 was considered to be important in the decision to participate in a certain screening programme.

Given the current controversy in the literature on whether to include individuals who failed the rationality test or not, the impact of removing respondents who failed the rationality test was explored by removing such individuals from the sample and rerunning the analysis. The results of the model indicated that removing irrational respondents had no major impact on the size or relative importance of the attributes, except for the attitudes of CRC screening participants towards preparation with '0.75 litres of fluid'. We therefore included individuals who failed the rationality tests as currently advised in the DCE literature [31, 32].

The trade-offs respondents were willing to make between the attributes were calculated by the ratios of the coefficients of the different attributes with risk reduction as the denominator. For example, β_1/β_{11} indicates how much additional relative risk reduction respondents think a test should generate in order to undergo a test that causes mild pain instead of a test that causes no pain.

To examine the expected uptake of CRC screening based on our results, we applied the model as presented by Gerard and colleagues and Hall and colleagues to our data [33, 34].

$$P_{participation} = \frac{1}{(1 + e^{\wedge -V})}$$

The model assumes that a preference score of 0 indicates that individuals have an equal preference for either participation or non-participation, hence the expected participation rate equals 50%. Additionally we investigated the effect of changing the most important CRC screening programme characteristics, as identified by the results or our multinomial logit model, on the expected uptake of CRC screening.

Aggregate data on socio-economic status (SES) were available at the level of the respondents' area zip code, weighted by the number of inhabitants per postal code and classified into three groups (high, average, low).

Characteristics of the different groups were compared using parametric and non-parametric tests. For categorical data, we used Chi-square and Fisher Exact Test to test for differences between screening naïve individuals and CRC screening participants. For continuous variables, we used the independent Samples T-test. To assess whether there were differences in preferences among participants of the FOBT (either gFOBT or FIT) and FS screening programme and those with and without endoscopy experience, we performed subgroup analyses.

Table 8.2 Respondent characteristics

Characteristics	Screening naïve	Participants	Difference
Respons (n respondents/n invited - %)	156/500 (31.0)	124/210 (59.0)	p<0.01
Analyzable questionnaires (n - %)	152 (97.4)	120 (96.8)	p=0.74
Age (mean - SD)	59.9 (5.7)	62.2 (6.4)	p<0.01
Gender (male; n - %)	74 (48.7)	59 (49.2)	p=0.94
Socio economic status (n - %)			p=0.49
High	78 (51.3)	53 (44.2)	
Intermediate	21 (13.8)	20 (16.7)	
Low	53 (34.9)	47 (39.2)	
Endoscopy experience (n - %)			p<0.01
Yes	33 (21.7)	64 (53.3)	
No	117 (77.0)	54 (45.0)	
Unknown	2 (1.3)	2 (1.5)	
Knowing someone affected by CRC (n - %)			p=0.84
Yes	19 (12.5)	18 (15.0)	
No	115 (75.7)	88 (73.3)	
Unknown	18 (11.8)	14 (11.6)	
EQ-5D summary score (mean - SD)	0.92 (0.11)	0.93 (0.12)	p=0.48

8.3 Results

8.3.1 Respondents

The response rate was higher among CRC screening participants (59%; 124/210) compared to screening naïve individuals (31%; 155/500) (Table 8.2). The characteristics of the respondents are shown in Table 8.2. Among the screening naïve group, 22% had undergone an endoscopy in the past. Within the group of CRC screening participants, 53% had previous endoscopy experience including 23% (16/70) of FOBT screenees and logically all FS screening subjects (48/48).

8.3.2 DCE results

Forty-three percent of the screening-naïve individuals and 50% of the CRC screening participants rated the questionnaire as ' easy' (p=0.24).
The signs of all coefficients of the attributes were consistent with our initial hypotheses (see Table 8.3). The positive sign given to the coefficient 'risk reduction of death from CRC' indicated that respondents preferred a test generating a higher risk reduction over a test that generates a lower risk reduction. The positive sign of the coefficients for shorter screening intervals indicated that individuals preferred those screening intervals over screening once every 10 years. The negative signs for all other attributes indicate that individuals preferred a screening test of shorter duration, with no preparation, no pain, and no risk of complications.

Table 8.3 Preferences of the screening naïve individuals and CRC screening participants

	Screening naïve		**Participants**	
	β-coeff	95% CI	β -coeff	95% CI
Constant (screening)	0.25	(-0.00 to 0.50)	0.62	(0.35 to 0.90)*
Pain				
No pain (ref)				
Mild pain	-0.31	(-0.42 to -0.20)*	-0.23	(-0.34 to -0.11)*
Risk of complications				
None (ref)				
Small	-0.16	(-0.28 to -0.05)*	-0.13	(-0.25 to -0.01)*
Location				
At home (ref)				
Hospital	-0.09	(-0.20 to 0.02)	-0.01	(-0.13 to 0.10)*
Preparation				
Enema. no fasting	-0.37	(-0.57 to -0.16)*	-0.23	(-0.45 to -0.02)*
Drinking 0.75l fluid.12h fasting	-0.51	(-0.72 to -0.29)*	-0.22	(-0.45 to 0.01)
Drinking 4l fluid. 18 h fasting	-0.98	(-1.18 to -0.77)*	-0.88	(-1.10 to -0.67)*
Duration (per 10 min)	-0.03	(-0.05 to -0.01)*	-0.03	(-0.06 to -0.01)*
Interval				
1x in 10 years (ref)				
2x in 10 years	0.28	(0.11 to 0.45)*	0.24	(0.06 to 0.42)*
5x in 10 years	0.4	(0.21 to 0.59)*	0.33	(0.13 to 0.53)*
10x in 10 years	0.33	(0.18 to 0.49)*	0.27	(0.10 to 0.44)*
Risk reduction of death from CRC (per relative 10% risk reduction)	0.32	(0.29 to 0.35)*	0.26	(0.24 to 0.29)*

* significant at the 5% level; (ref) = reference level; β-coeff = β-coefficient; CI = confidence interval

The non-significant coefficient of the constant term in the screening-naïve group indicated that these subjects had, if assuming a screening programme with the reference level for all the attributes, no preference for either screening or no screening, whereas the group of CRC screening participants expressed a positive attitude towards screening compared to no screening (positive significant coefficient). All screening attributes proved to be important determinants of the preferences in each of the respondent groups, except for location of the screening test, which only significantly influenced preferences of CRC screening participants and not those of the screening naïve individuals and a preparation with '0.75 litres of fluid and 12 hours fasting', that did not influence preferences of CRC screening participants.

The differences in preferences *between* screening naïve-individuals and participants of a CRC screening programme statistically not significant, except for preferences regarding risk reduction of CRC related death. Screening naïve individuals demanded more effectiveness from a CRC screening programme compared to participants ($p<0.01$). We performed subgroup analyses, analysing FOBT and FS screenees separately, which showed that participants of FOBT and FS screening did differ in preferences: FS screenees expressed a positive attitude, while FOBT screenees expressed a negative attitude towards a test in the hospital ($p<0.001$). Furthermore, FS screenees attached more importance to a 5-yearly screening interval ($p=0.01$) and to the effectiveness of a screening test ($p<0.001$) than FOBT screenees.

When comparing those with previous endoscopy experience to those without endoscopy experience, it could be seen that pain had a significant greater influence on preferences for those without previous endoscopy experience ($p=0.02$). The location hospital was negatively associated with preferences for those without endosocpy experience, but it had a positive affect on preferences for those who had undergo a previous endoscopy (difference: $p<0.01$). Individuals without endoscopy experience also demanded more effectiveness from a screening test ($p<0.01$).

Screening naïve individuals and CRC screening participants significantly preferred no preparation to all other preparations (p-values <0.03). Both groups significantly preferred preparation with an 'enema' or '0.75 litres of fluid' instead of a preparation with '4 litres of fluid' (p-values <0.001). Preparation with an 'enema' and '0.75 litres of fluid' were valued equally by both groups (p-values > 0.09).

8.3.3 Trade-offs

It can be seen in Table 8.4, that based on the expressed preferences, screening-naïve individuals required an additional relative risk reduction of 30% (95% confidence interval (CI) 24-37%) for participation in a screening programme with a test requiring a preparation with '4 litres of fluid and 18 hours fasting' instead of a test that required 'no preparation'. Respondents preferred shorter screening intervals and they were willing to give up a 12% (CI 7-18%) relative risk reduction if the screening interval was *shortened*

Table 8. 4 Individuals' tradeoffs between risk reduction and different aspects of a screening programme

	Screening naïve	Participants	Interpretation note
	% of additional relative risk reduction respondents think a test should generate....		
Pain	10%	9%	.. in order to undergo a test that causes mild pain instead of a test that causes no pain
Risk of complications	5%	5%	.. in order to undergo a test that carries a small risk of complications instead of a test with no risk of complications
Preparation			.. in order to accept a test that requires a preparation with one of these three methods instead of a test requiring no preparation at all
Enema. no fasting	11%	9%	
Drinking 0.75l, 12h fasting	16%	8%	
Drinking 4l, 12h fasting	30%	33%	
Duration	1%	1%	.. in order to accept a test with an additional 10 minutes of duration compared to the standard duration
Interval			.. if the screening interval is lengthened from one of the shorter, more preferred, screening intervals (5-yearly, biennial, annual) to the longest screening interval (once every 10 years)
2x in 10 years	9%	9%	
5x in 10 years	12%	13%	
10x in 10 years	10%	10%	

from once every 10 years to a 2-yearly screening interval. Participants of a CRC screening programme made trade-offs that were comparable to those of the screening naïve individuals.

8.3.4 Expected uptake of CRC screening

The average expected uptake of CRC screening was 56% (CI 50-62%) for screening naïve individuals. If we calculate the expected uptake for a screening programme without taking into account risk reduction of CRC related death and screening interval it would be 55% for the FOBT, 31% for FS and 17% for colonoscopy. This corresponds well to the participation rates as observed in the Dutch pilot screening programme (gFOBT 49.5% (CI 48.1-50.9%); FIT 61.5% (CI 60.1-62.9%); FS 32.4% (CI 31.1-33.7%) [13]. Assuming that all screening tests would generate a 10% risk reduction of CRC related death, uptake would increase to 72% for biennial FOBT screening , 46% for 5-yearly FS screening and 22% for 10-yearly colonoscopy screening. We would expect that, if individuals are aware of the achievable risk reduction as currently known from the literature, the uptake would increase to 75% for biennial FOBT screening, 80% for five-yearly FS screening and 71% for 10-yearly colonoscopy screening (risk reduction of CRC related death respectively 16% [35], 59% [5] and 74.5% [36]. The effects of changing the CRC screening programme characteristics on average expected uptake of CRC screening are shown in Figure 8.2.

8.4 Discussion

Our study demonstrates the importance of several procedural characteristics of CRC screening programmes for the preferences of potential and actual screenees: risk reduction of CRC-related death, preparation for the procedure, procedure related pain and complications and screening interval. To optimise a screening program, the attendance rate should be high. A high attendance rate is only possible when the utilised screening strategy and the information given connect with the preferences of the target population. The results of this DCE in the first place indicate targets for improvement of CRC screening programmes. Secondly they stress the importance of several aspects of screening programmes regarding the information provided to screening invitees. To our knowledge, this is the first study assessing preferences for CRC screening among both screening-naïve subjects and CRC screening participants.

In our DCE, especially mortality reduction had an important positive influence on preferences for CRC screening methods. A few other studies have investigated preferences for CRC screening using a DCE [37-42]. Our finding that individuals attach much importance to CRC mortality reduction by a screening method is consistent with the results of previous studies [37, 42, 43]. The finding that individuals are prepared to undergo more burdensome screening tests if this results in sufficient additional risk

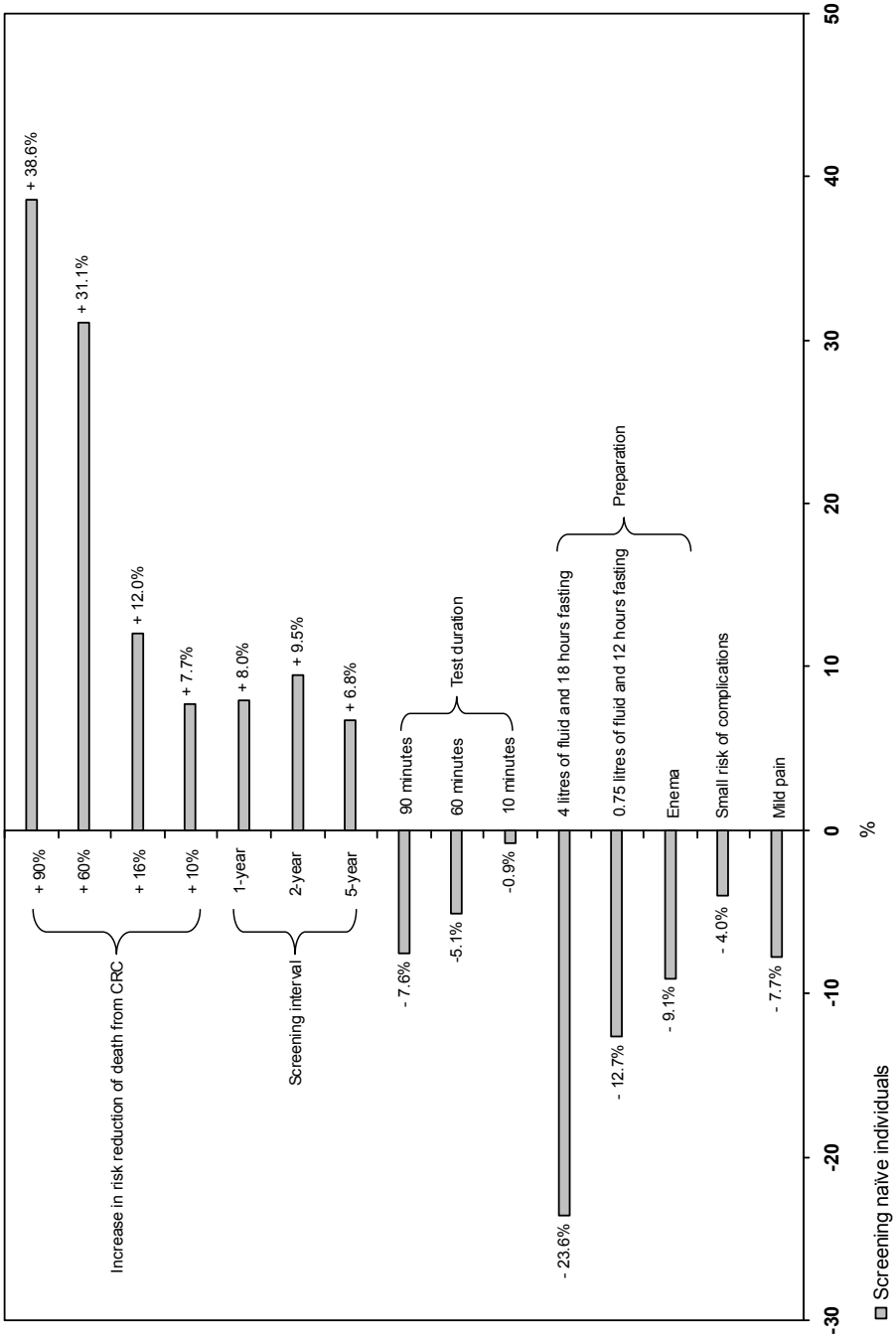


Figure 8.2 Effects of changing the screening programme characteristics on the average probability of participation in CRC screening (56.2%), as predicted by the multinomial logit model

reduction of CRC related mortality demonstrates that they trade benefits and harms of a screening test.

The burden of the required preparation was considered the main drawback of undergoing CRC screening. A preparation commonly used for colonoscopy (i.e. drinking 4 litres of fluid and 18 hours fasting) would only be chosen when an additional relative risk reduction of, on average, 33% would be achieved. In line with our results, Canadian investigators found that preparation was ranked as the most important process related attribute. In contrast, American investigators found that preparation was rated as the least important attribute [38]. The levels that were chosen for the attributes may explain those differences. The results of our DCE are of utmost importance when for example starting a colonoscopy screening programme with a burdensome preparation. Emphasis should be laid on adequate information that should be provided to the target population about the burden and benefits including expected CRC mortality reduction by colonoscopy screening, since this may compensate for a burdensome preparation.

Interestingly, we found that respondents significantly preferred shorter screening intervals to a 10-year screening interval irrespective of health benefit. This finding is consistent with a previous study suggesting that women preferred shorter (annual and biennial) over longer (3-, 4- or 5-year) screening intervals for cervical cancer screening [44]. One study among Danish individuals and another among both American and Canadian individuals could not confirm preferences for shorter CRC screening intervals [37, 41]. A second American study could not determine if individuals preferred shorter or longer screening intervals [38]. Several studies have showed that reassurance may be a motivation for and/or a result of undergoing cancer screening [45, 46]. The preference for shorter screening intervals found in our study may be associated with expected reassurance. This again stresses the importance of adequate information provided to potential screenees. It emphasises the need to adequately inform individuals that longer screening intervals for CRC screening do not imply lower reductions in mortality, but that specific CRC screening tests with longer screening intervals have more potential for CRC prevention and therefore require less frequent testing.

There were some differences in preferences between FOBT and FS screenees. Assessment of preference variations across subgroups is advisory because of status quo bias; in other words the tendency of people to value services higher once they have experienced them [47]. We conducted the study among both screening-naïve individuals and individuals who had prior experience with CRC screening tests, so that we were able to investigate if status quo bias was present. The preferences of screening-naïve subjects and participants of a CRC screening programme were not significantly different. The fact that FOBT screenees expressed a negative attitude towards a test in the hospital, while FS screenees expressed a positive attitude towards a test in the hospital may be explained by the phenomenon of status quo bias. However, it may also be a result of selection bias; that those subjects with a preference for the location 'home' do not participate in FS screening and vice versa. Interestingly, the same significant difference regarding the influence of screening location on preferences was observed when comparing those with

endoscopy experience to those without. A possible explanation might be that individuals on beforehand have a negative association with the location hospital, but develop a positive attitude towards a hospital-based examination once they have experienced it.
Research has consistently shown that expected pain is one of the most important reasons for declining the endoscopic screening offer [18, 48, 49]. The results from our study confirm that finding and furthermore they demonstrate that pain has significant less influence on preferences of those with endoscopy experience, suggesting that pain actually experienced during endoscopic screening is not as severe as expected on beforehand.
This study revealed uptake levels of the FOBT, FS and TC based on the characteristics in our model. We found that mainly risk reduction of CRC related death highly influenced the participation that could be expected for the different screening tests. If we took into account the fact that individuals in the randomised screening trial were not informed on achievable risk reduction or required frequency of testing, the uptake as calculated by the model corresponded to the results found in the screening trial [13]. However, participation rates based on our model increased if risk reduction of CRC related death increased, suggesting that increasing awareness on efficacy of the screening tests might enhance uptake.
Given the low levels of awareness of CRC screening in the Netherlands, it may be of vital importance to raise knowledge on achievable risk reduction of CRC related death in order to increase screening uptake especially for the more effective endoscopic screening tests. The importance of awareness on efficacy of the available screening tests is further underlined by data of a Swiss study, in which 75% of all screenees chose to undergo a TC and only 25% preferred FOBT or FS screening after they were informed about the efficacy of all screening methods [50]. This study involved testimonies from patients with CRC in their campaign in order to raise CRC awareness. This strategy has also been used in various other campaigns throughout the European Union, among others in the United Kingdom, Germany and the Netherlands. CRC patients and their relatives may be important advocates for raising awareness, and possibly also for increasing public familiarity with endoscopic screening which has been demonstrated to influence CRC screening preferences in our study.
There are some limitations to our study. The way we framed the information on risk reduction may have influenced our results.
Furthermore, there was a significant difference in response rate between screening-naïve individuals and CRC screening participants. This may have given a selection bias and thereby be a limitation regarding the interpretation of our results.
In conclusion, individuals are willing to trade-off benefits and harms of CRC screening programmes. Especially type of bowel preparation, length of screening interval and mortality reduction influenced individuals' trade-offs. The results provide insight in the decision-making process regarding the decision to participate in a CRC screening programme. This information can be used to improve information provided to CRC screening invitees, and identify targets for increasing participation rates.

References

1. Ferlay J, Autier P, Boniol M, Heanue M, Colombet M, Boyle P (2007). Estimates of the cancer incidence and mortality in Europe in 2006. Ann Oncol Mar;18(3):581-92.
2. Karim-Kos HE, de VE, Soerjomataram I, Lemmens V, Siesling S, Coebergh JW (2008). Recent trends of cancer in Europe: a combined approach of incidence, survival and mortality for 17 cancer sites since the 1990s. Eur J Cancer 44(10):1345-89.
3. Faivre J, Dancourt V, Lejeune C, Tazi MA, Lamour J, Gerard D, et al (2004). Reduction in colorectal cancer mortality by fecal occult blood screening in a French controlled study. Gastroenterology; 126(7):1674-80.
4. Hardcastle JD, Chamberlain JO, Robinson MH, Moss SM, Amar SS, Balfour TW, et al (1996). Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. Lancet; 348(9040):1472-7.
5. Hoff G, Grotmol T, Skovlund E, Bretthauer M (2009). Risk of colorectal cancer seven years after flexible sigmoidoscopy screening: randomised controlled trial. BMJ;338:b1846.
6. Kronborg O, Fenger C, Olsen J, Jorgensen OD, Sondergaard O (1996). Randomised study of screening for colorectal cancer with faecal-occult-blood test. Lancet; 348(9040):1467-71.
7. Lindholm E, Brevinge H, Haglind E (2008). Survival benefit in a randomized clinical trial of faecal occult blood screening for colorectal cancer. Br J Surg; 95(8):1029-36.
8. Mandel JS, Bond JH, Church TR, Snover DC, Bradley GM, Schuman LM, et al (1993). Reducing mortality from colorectal cancer by screening for fecal occult blood. Minnesota Colon Cancer Control Study. N Engl J Med; 328(19):1365-71.
9. Mandel JS, Church TR, Ederer F, Bond JH (1999). Colorectal cancer mortality: effectiveness of biennial screening for fecal occult blood. J Natl Cancer Inst; 91(5):434-7.
10. Scholefield JH, Moss S, Sufi F, Mangham CM, Hardcastle JD (2002). Effect of faecal occult blood screening on mortality from colorectal cancer: results from a randomised controlled trial. Gut; 50(6):840-4.
11. Winawer SJ, Zauber AG, Ho MN, O'Brien MJ, Gottlieb LS, Sternberg SS, et al (1993). Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. N Engl J Med; 329(27):1977-81.
12. Levin B, Lieberman DA, McFarland B, Smith RA, Brooks D, Andrews KS, et al (2008). Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. CA Cancer J Clin; 58(3):130-60.
13. Hol L, van Leerdam ME, van Ballegooijen M, van Vuuren AJ, van Dekken H, Reijerink JCIY, et al (2009). Screening for Colorectal Cancer; Randomised Trial Comparing Guiac-based and Immunochemical Faecal Occult Blood Testing and Flexible Sigmoidoscopy. Gut. In press.
14 Rebolj M, van Ballegooijen M., Berkers LM, Habbema D (2007). Monitoring a national cancer prevention program: successful changes in cervical cancer screening in the Netherlands. Int J Cancer; 120(4):806-12.
15. Schopper D, de WC (2009). How effective are breast cancer screening programmes by mammography? Review of the current evidence. Eur J Cancer.
16. Manfredi S, Piette C, Durand G, Plihon G, Mallard G, Bretagne JF (2008). Colonoscopy results of a French regional FOBT-based colorectal cancer screening program with high compliance. Endoscopy; 40(5):422-7.
17. van Rossum LG, van Rijn AF, Laheij RJ, van Oijen MG, Fockens P, van Krieken HH, et al (2008). Random comparison of guaiac and immunochemical fecal occult blood tests for colorectal cancer in a screening population. Gastroenterology; 135(1):82-90.
18. Vernon SW (1997). Participation in colorectal cancer screening: a review. J Natl Cancer Inst; 89(19):1406-22.
19. Vogelaar I, van BM, Schrag D, Boer R, Winawer SJ, Habbema JD, et al (2006). How much can current interventions reduce colorectal cancer mortality in the U.S.? Mortality projections for scenarios of risk-factor modification, screening, and treatment. Cancer; 107(7):1624-33.

20. Blalock SJ, DeVellis BM, Afifi RA, Sandler RS (1990). Risk perceptions and participation in colorectal cancer screening. Health Psychol; 9(6):792-806.
21 Keighley MR, O'Morain C, Giacosa A, Ashorn M, Burroughs A, Crespi M, et al (2004). Public awareness of risk factors and screening for colorectal cancer in Europe. Eur J Cancer Prev; 13(4):257-62.
22. Phillips KA, Van BS, Marshall D, Walsh J, Thabane L (2006). A review of studies examining stated preferences for cancer screening. Prev Chronic Dis; 3(3):A75.
23. Louviere JJ, Hensher DA, Swait JD (2000). Stated Choice Methods: Analysis and Applications. Cambridge, UK: Cambridge University Press.
24. Caldon LJ, Walters SJ, Ratcliffe J, Reed MW (2007). What influences clinicians' operative preferences for women with breast cancer? An application of the discrete choice experiment. Eur J Cancer; 43(11):1662-9..
25. Johnson FR, Ozdemir S, Mansfield C, Hass S, Miller DW, Siegel CA, et al (2007). Crohn's disease patients' risk-benefit preferences: serious adverse event risks versus treatment efficacy. Gastroenterology; 133(3):769-79.
26. Hol L, van Leerdam ME, van Ballegooijen M, van Vuuren AJ, Reijerink-Verheij JC, van der Togt-van Leeuwen AC, et al (2008). Attendance to Screening for Colorectal Cancer in the Netherlands; Randomized Controlled Trial Comparing Two Different Forms of Faecal Occult Blood Tests and Sigmoidoscopy. Gastroenterology; (Supplement 1):Abstract no 621 (A-87).
27. Ryan M (2004). Discrete choice experiments in health care. BMJ; 328(7436):360-1.
28. Sloane NJA. A library of orthogonal arrays. web-site 2008. URL: http://www.research.att.com/ ~njas/oadir/
29. Street DJ, Burgess L, Louviere JJ (2005). Quick and easy choice sets: Constructing optimal and nearly optimal stated choice experiments. Intern J of Research in Marketing; 22:459-70.
30. Dolan P (1997). Modeling valuations for EuroQol health states. Med Care; 35(11):1095-108.
31. Lancsar E, Louviere J (2006). Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? Health Econ; 15(8):797-811.
32. Ryan M, Watson V, Entwistle V (2009). Rationalising the 'irrational': a think aloud study of discrete choice experiment responses. Health Econ; 18(3):321-36.
33. Gerard K, Shanahan M, Louviere J (2008). Using Discrete Choice Modelling to Investigate Breast Screening Participation. In: Ryan M, Gerard K, Amaya-Amaya M, editors. Using Discrete Choice Experiments to Value Health and Health Care.Dordrecht: Springer;. p. 117-37.
34. Hall J, Kenny P, King M, Louviere J, Viney R, Yeoh A (2002). Using stated preference discrete choice modelling to evaluate the introduction of varicella vaccination. Health Econ; 11(5):457-65.
35. Hewitson P, Glasziou P, Watson E, Towler B, Irwig L (2008). Cochrane systematic review of colorectal cancer screening using the fecal occult blood test (hemoccult): an update. Am J Gastroenterol; 103(6):1541-9.
36. Zauber AG, Lansdorp-Vogelaar I, Knudsen AB, Wilschut J, van BM, Kuntz KM (2008). Evaluating test strategies for colorectal cancer screening: a decision analysis for the U.S. Preventive Services Task Force. Ann Intern Med; 149(9):659-69.
37. Gyrd-Hansen D, Sogaard J (2001). Analysing public preferences for cancer screening programmes. Health Econ; 10(7):617-34.
38. Hawley ST, Volk RJ, Krishnamurthy P, Jibaja-Weiss M, Vernon SW, Kneuper S (2008). Preferences for colorectal cancer screening among racially/ethnically diverse primary care patients. Med Care; 46(9 Suppl 1):S10-S16.
39. Howard K, Salkeld G (2008). Does Attribute Framing in Discrete Choice Experiments Influence Willingness to Pay? Results from a Discrete Choice Experiment in Screening for Colorectal Cancer. Value Health.
40. Marshall DA, Johnson FR, Phillips KA, Marshall JK, Thabane L, Kulin NA (2007). Measuring patient preferences for colorectal cancer screening using a choice-format survey. Value Health; 10(5):415-30.
41. Marshall DA, Johnson FR, Kulin NA, Ozdemir S, Walsh JM, Marshall JK, et al (2009). How do physician assessments of patient preferences for colorectal cancer screening tests differ from actual preferences? A comparison in Canada and the United States using a stated-choice survey. Health Econ.

42. Salkeld G, Solomon M, Short L, Ryan M, Ward JE (2003). Evidence-based consumer choice: a case study in colorectal cancer screening. Aust N Z J Public Health; 27(4):449-55.
43. Salkeld GP, Solomon MJ, Short L, Ward J (2003). Measuring the importance of attributes that influence consumer attitudes to colorectal cancer screening. ANZ J Surg; 73(3):128-32.
44. Holloway RM, Wilkinson C, Peters TJ, Russell I, Cohen D, Hale J, et al (2003). Cluster-randomised trial of risk communication to enhance informed uptake of cervical screening. Br J Gen Pract; 53(493):620-5.
45. Cantor SB, Volk RJ, Cass AR, Gilani J, Spann SJ (2002). Psychological benefits of prostate cancer screening: the role of reassurance. Health Expect; 5(2):104-13.
46. Whynes DK, Philips Z, Avis M (2007). Why do women participate in the English cervical cancer screening programme? J Health Econ; 26(2):306-25.
47. Salkeld G, Ryan M, Short L (2000). The veil of experience: do consumers prefer what they know best? Health Econ;9(3):267-70.
48. Codori AM, Petersen GM, Miglioretti DL, Boyd P (2001). Health beliefs and endoscopic screening for colorectal cancer: potential for cancer prevention. Prev Med; 33(2 Pt 1):128-36.
49. Janz NK, Lakhani I, Vijan S, Hawley ST, Chung LK, Katz SJ (2007). Determinants of colorectal cancer screening use, attempts, and non-use. Prev Med; 44(5):452-8.
50. Marbet UA, Bauerfeind P, Brunner J, Dorta G, Valloton JJ, Delco F (2008). Colonoscopy is the preferred colorectal cancer screening method in a population-based program. Endoscopy; 40(8):650-5.
51. Edwards A, Elwyn G, Mulley A (2002). Explaining risks: turning numerical data into meaningful pictures. BMJ; 324(7341):827-30.

Part III

Theory of discrete choice experiments in health care

9

Labelled versus unlabelled discrete choice experiments in health economics: an application to colorectal cancer screening

de Bekker-Grob EW, Hol L, Donkers B, van Dam L, Habbema JDF, van Leerdam ME, Kuipers EJ, Essink-Bot ML, Steyerberg EW

Submitted

Abstract

Objectives Discrete choice experiments (DCE) in health economics commonly present choice sets in an *unlabelled* form. *Labelled* choice sets are less abstract and may increase the validity of the results. We empirically compared the feasibility, respondents' trading behaviour, and convergent validity between a labelled and an unlabelled DCE for colorectal cancer (CRC) screening programs in the Netherlands.

Methods A labelled DCE version presented CRC screening test alternatives as 'fecal occult blood test', 'sigmoidoscopy', and 'colonoscopy', whereas the unlabelled DCE version presented them as 'screening test A' and 'screening test B'. Questionnaires were sent to participants and non-participants in CRC screening.

Results Total response rate was 276/712 (39%) and 1,033/2,262 (46%) for unlabelled and labelled DCEs respectively ($p<0.001$). The labels played a significant role in individual choices; approximately 22% of subjects had dominant preferences for screening test labels. The convergent validity was modest to low (participants in CRC screening: $r=0.54$; $p=0.01$; non-participants: $r=0.17$; $p=0.45$), largely due to different preferences for screening frequency.

Conclusions This study provides important insights in the feasibility and difference in results from labelled and unlabelled DCEs. The inclusion of labels appeared to play a significant role in individual choices, but reduced the attention respondents give to the attributes. As a result, unlabelled DCEs may be more suitable to investigate trade-offs between attributes and for respondents who do not have familiarity with the alternative labels, whereas labelled DCEs may be more suitable to explain real life choices such as uptake of cancer screening.

9.1 Introduction

Estimates of public and patients' preferences are of great importance to inform policy decision making and improve adherence with public health care interventions or programs [1]. Discrete choice experiments (DCEs) have become a commonly used technique in health economics to elicit preferences. The DCE is an attribute-based survey method for measuring benefits (utility) [2]. In a DCE, subjects are presented with a sequence of (hypothetical) scenarios (choice sets) and are asked to choose between two or more competing alternatives that vary along several characteristics or attributes of interest [2]. DCEs assume that subjects' preferences (as summarized by their utility function) are revealed through their choices [2]. (For further details see Bliemer and Rose [3]; Hensher et al. [4]; Louviere et al. [5]; and Ryan et al. [2]).

A fundamental question that arises in the application of DCE is whether to present the choice sets in a labelled or unlabelled form. The unlabelled form involves assigning unlabelled alternatives in the choice set, such as 'alternative A', 'alternative B' and so on. The labelled form involves assigning labels that communicate information regarding the alternative. In marketing applications, labels tend to consist of brand names and logos, which consumers have learnt to associate with different product characteristics and feelings. In the context of health economics, labels tend to consist of generic or brand-name medications, specific screening tests (e.g. colonoscopy, sigmoidoscopy), specific treatments (surgery vs conservative), or other descriptors. An advantage of assigning labels is that alternatives will be more realistic and the choice task less abstract for the subject, which add to the validity of the results. Hence, the results may be better suitable to support decision-making at policy level. However, by far most commonly applied DCEs in health economics used unlabelled alternatives.

The aim of our study was to empirically compare the feasibility, respondents' trading behaviour, and convergent validity between a labelled and an unlabelled DCE. All of these aspects were explored in the context of a DCE study directed at investigating population preferences for colorectal cancer (CRC) screening programs in the Netherlands. We were convinced that specific aspects of endoscopy (sigmoidoscopy, colonoscopy) or fecal occult blood test (FOBT) that determine its burden could not be totally captured by presenting an unlabelled 'screening test A' variant to patients [6]. For that very reason, we expected differences between an unlabelled and a labelled DCE.

The paper is structured as follows. The theoretical basis of labelled and unlabelled DCEs is briefly reviewed in Section 9.2. The methods including the case study are then presented in Section 9.3, and the results are presented in Section 9.4. Finally, a discussion and conclusions are drawn.

9.2 Theoretical basis of labelled and unlabelled DCEs

The aim of discrete choice modeling is to estimate the weights that respondents place on attributes of alternatives. An individual acting rationally is expected to evaluate the set of

available alternatives and will choose that alternative which gives the greatest relative utility [4]. Thus, an individual will choose alternative A over B, if U (X_A, Z) > U (X_B, Z), where U represents the individual's indirect utility function from certain alternatives, X_A the attributes of alternative A, X_B the attributes of alternative B, and Z socioeconomic and other characteristics of the individual that influence his/her utility. Choices made in DCEs are analyzed using random utility theory (i.e. an error term is included in the utility function to reflect the unobservable factors in the individual's utility function) [4]. Thus, an individual will choose alternative A over B, if V (X_A, Z) + ε_A > V (X_B, Z) + ε_B, where V is the measurable component of utility estimated empirically, and ε_A and ε_B reflect the unobservable factors in the individual's utility function of alternative A and B respectively (X_A, X_B and Z defined as above).

There are two general types of DCEs: 1) unlabelled and 2) labelled DCEs [5]. Unlabelled DCEs use generic titles for the alternatives (e.g. radio-imaging 'A' or 'B'). Labelled DCEs use alternative-specific titles for the alternatives (e.g. 'computer tomography' or 'MRI-scan'). The number of alternatives (irrespective labelled or unlabelled) in a choice set is unrestricted from a theoretical point of view [4]. The decision as to whether to use labelled or unlabelled DCEs is an important one [4]. The labelled alternative itself conveys information to respondents. This matters in choice and other decision tasks, because (a) respondents may use labelled alternatives to infer information that they perceive as missing, and (b) these inferences may be (and usually are) correlated with the random component [5]. Although we may not exactly know what respondents find relevant in the label, for forecasting uptake of, for example, a health care intervention it may be worthwhile to find out if respondents prefer one alternative label to another. A labelled DCE can take effects into account, which respondents may have learnt to associate with different health care intervention characteristics and feelings, and as a result may be more suitable to predict [6]. Unlabelled and labelled DCEs both have their merits. If each of the labelled options has A attributes with L levels and the choice sets are of size M then there are L^MA possible choice sets assuming that all labels are presented in a choice set and that the same label does not appear more than once in a choice set. If the options are unlabelled then there are L^A possible items that can be included in each position of each choice set. If the choice sets are of size M and we are not going to allow the same item to appear more than once in a choice set then there are "L^A choose M" possible choice sets of size M Therefore the designs of an unlabelled DCE can be much smaller. For example, two alternatives with 4 attributes and 3 levels yields 6561 (i.e. $3^{(2*4)} = 3^8$) possible alternative combinations for a labelled DCE compared with 'just' 81 (i.e. 3^4) possible alternative combinations for an unlabelled design. Other merits of unlabelled DCEs include that they 1) do not require the identification and use of all alternatives within the universal set of alternatives; namely, the attribute levels are sufficiently broad to represent all alternatives; and 2) might be more robust in terms of not violating IID-assumption (i.e. error terms are independent and identically distributed), because the alternatives may be less correlated with the attributes as in labelled DCEs [4]; and 3) encourage respondents to choose an alternative by trading-off attribute levels, which may be desirable from a

non-market valuation perspective [7]. On the other hand, merits of using labelled DCEs include that they 1) will be more realistic and less abstract, so that responses may better reflect the real preference structure; and 2) can study the main effect of the labels.

9.3 Methods

9.3.1 Case study

Colorectal carcinoma (CRC) is the most frequently occurring malignancy within the European Union, and the second leading cause of cancer related death in the Western world [8, 9]. Various countries have implemented a national screening program for CRC screening to detect CRC in an early stage or are investigating prerequisites for implementation [10, 11]. There are several screening tests eligible for use as a population-based screening program such as fecal occult blood tests (FOBTs), sigmoidoscopy, or colonoscopy. The DCE study aimed at investigating individual preferences for CRC screening.

9.3.2 Discrete choice experiment design

The questionnaire design phase involved extensive background research, expert opinions and interviews with screened individuals. Experts (n=3) were asked to comment on a list of test characteristics derived from our extensive literature review. Potential screenees (n=40), both participants of a CRC screening program (n=20) and screening naïve individuals (n=20), could also comment on the list of test characteristics and rank them in order of importance. Based on these data we selected the most important test characteristics. The levels for each test characteristic incorporated the range of possible test outcomes based on the current literature (for more detail how the qualitative data was used to select the final test labels, attributes and levels see Hol et al. [12]; and Van Dam et al. [13]). Table 9.1 lists the labels, attributes, and attribute levels chosen. The labelled CRC screening tests ('FOBT', 'sigmoidoscopy', and 'colonoscopy') may evoke individual feelings, which may not be captured in the unlabelled CRC screening tests ('CRC screening test A' and 'CRC screening test B'). Notably, the invasiveness of the alternative test was (indirectly) described by the levels of five attributes: 'side effects of the test', 'complication risk of the test', 'preparation for the patient', 'location of screening', and 'the duration of screening'. Giving directly the information 'how a sample is taken' is in our case totally equal to the screening test label: 'taking a sample from your motion' is equal to FOBT, and 'tube into your back passage throughout your colon' is equal to colonoscopy. If the unlabelled DCE would include directly this information about 'how the sample is taken' (thus actually naming the test), then the unlabelled DCE will be a labelled DCE as well; the attribute 'how the sample is taken' will have an interaction with all other attributes and a restricted design is needed to avoid implausible combinations of attribute levels (i.e. the attribute levels are alternative specific, and thus a labelled DCE).

Table 9.1 Attribute and levels for unlabelled and labelled discrete choice experiment between three alternatives A, B, and C.

Attribute	Levels in unlabelled model	Levels in labelled model
Reduction in mortality	Options A and B: from 3.0% to 0.3%, 1.2%, 1.8%, 2.7% Option C (no test): from 3.0% to 3.0%	FOBT: from 3.0% to 1.8%, 2.3%, 2.7% Sigmoidoscopy: from 3.0% to 0.9%, 1.5%, 1.8% Colonoscopy: from 3.0% to 0.1%, 0.5%, 0.8% No test (Option C; base): from 3.0% to 3.0%
Frequency of screening per 10 years	Options A and B: 1, 2, 5, 10 Option C: 0	FOBT: 3, 10, 30 Sigmoidoscopy: 1, 2, 10 Colonoscopy: 1, 2, 5 No test: 0
Complication risk	Options A and B: none, small Option C: none	FOBT: none Sigmoidoscopy: small Colonoscopy: small No test: none
Location of screening	Options A and B: at home, hospital Option C: none	FOBT: at home Sigmoidoscopy: hospital Colonoscopy: hospital No test: none
Screening duration	Options A and B: 10, 30, 60, 90 min Option C: 0 min	FOBT: 30min Sigmoidoscopy: 15min Colonoscopy: 105min No test: 0min
Preparation for patient	Options A and B: none, enema no fasting, drinking 0.75l + fasting, drinking 4l + fasting Option C: none	FOBT: none Sigmoidoscopy: enema, no fasting Colonoscopy: drinking 4l + fasting No test: none
Side effects of screening	Options A and B: none, mild pain Option C: none	FOBT: none Sigmoidoscopy: mild pain Colonoscopy: mild pain No test: none

Another point of notice is that the unlabelled experiment had for some attributes a smaller level range than the feasible options in the labelled experiment. As a result, we avoided some extreme combinations of 30 times a screening test resulting in a reduction in mortality from 3.0% to 2.7% in the unlabelled DCE, which added to utility balance.
The combination of the attributes and attribute levels of the unlabelled design resulted in 2048 CRC screening test alternatives (4^4*2^3). A fractional factorial design was used based on a web-site, which contained a library of over 200 orthogonal arrays [14], to reduce the number of alternatives to a manageable level of 16 alternatives in which orthogonality and level balance were fulfilled. These 16 alternatives were paired up with another orthogonal array by using the fold-over technique (i.e. cyclic design), which caused minimal overlap between attribute levels [15]. Each choice set (i.e. a set of available alternatives) contained two screening test alternatives and an opt-out (see Table 9.2a for an example). The unlabelled design had an efficiency of 95% compared with an optimal choice set design, and all main effects were uncorrelated according to the results of an analysis using the software of Street and Burgess [16].
Attribute levels in the labelled DCE were alternative-specific. In other words, different CRC screening test labels (FOBT, sigmoidoscopy, colonoscopy) were associated with different sets of outcomes. Implausible combinations of attribute levels and labels were minimized as a result. Furthermore, the implausible combinations of the attribute levels longest screening interval with simultaneous highest risk reduction as well as shortest screening interval with lowest risk reduction were blocked. Optimal designs for labelled DCEs, which require a design with two-way interactions, are not available for the general

Table 9.2a An example of an unlabelled choice set

	Screening test A **(A)**	**Screening test B** **(B)**	**No screening test** **(C)**
Preparation:	None	Enema, no fasting	None
Location:	At home	Hospital	None
Pain:	None	Mild pain	None
Complication risk:	None	Small	None
Mortality risk of colorectal cancer decrease:	from 3% to 2.7%	from 3% to 1.8%	remain 3%
Frequency of screening test in the next 10 years	10x	5x	0x
Time duration	10 minutes	30 minutes	0 minutes

case. Fortunately, SAS software is capable of generating designs that are highly efficient [15] in such circumstances. Hence, for the labelled DCE a D-efficient design was generated with SAS software (Version 9.1, SAS Institute Inc., Cary, NC, USA), which resulted in 84 choice sets divided over 7 versions of the questionnaire (D-error 0.573). Each choice set contained two CRC screening test alternatives and an opt-out (see Table 9.2b for an example).

The unlabelled as well as the labelled DCE contained a dominant choice set (i.e. a choice set in which one screening test alternative is logically preferable) to assess the understanding of the questionnaire (i.e. rationality test). Testing for internal validity should not automatically lead to deleting responses based on 'irrational' preferences, although it may be 'common' practice (e.g. [17-21]). Deleting ' irrational responses' may lead to removal of valid preferences, induce sample selection bias, and reduce the statistical efficiency and power of the estimated choice models [22]. Therefore further sensitivity analyses were conducted to quantify the effect of including and excluding 'irrational' responses.

All respondents received the same prior information to the questionnaire: an information brochure explaining different current CRC screening tests (FOBT, sigmoidoscopy, and colonoscopy; i.e. how the sample could be obtained) and their characteristics (advantages and disadvantages). Both DCEs were pilot tested to make sure that respondents could manage the length of the questionnaires, and to check for any problems in interpretation and face validity. None of the respondents raised any problems with understanding the questionnaires, so that the pilot test did not result in any changes to the questionnaires.

Table 9.2b An example of labelled choice set

	Sigmoidoscopy **(A)**	**FOBT** **(B)**	**No screening test** **(C)**
Preparation:	Enema, no fasting	None	None
Location:	Hospital	At home	None
Pain:	Mild pain	None	None
Complication risk:	Small	None	None
Mortality risk of colorectal cancer decrease:	from 3% to 0.9%	from 3% to 2.3%	remain 3%
Frequency of screening test in the next 10 years	2x	10x	0x
Time duration	15 minutes	30 minutes	0 minutes

9.3.3 Study sample and elicitation mode

The questionnaires were sent by mail to subjects who had recently participated in a regional call-recall CRC screening program (unlabelled n=212; labelled n=769) and to randomly selected screening naïve subjects of the same region (Groot-Rijnmond) (unlabelled n=500; labelled n=1,498). It was not possible to directly calculate the statistical power to inform the sample size for a choice experiment. Other studies showed that a sample size of 42-208 respondents was sufficient to answer 16 unlabelled choice sets [23-25]. Because the fractional factorial design of the labelled approach was much larger than the unlabelled DCE and more coefficients had to be estimated due to alternative specific parameters, we sent out more labelled than unlabelled DCEs to strive for precise estimation of the parameters. All respondents were between 50-74 years of age. Besides the choice sets, the questionnaires also included background variables of respondents such as age, gender, endoscopy (i.e. sigmoidoscopy or colonoscopy) experience, familiarity with CRC because of cases in family or friends, and included standardized questions (EQ-5D) to measure self-reported health state. A reminder was sent to non-responders four weeks later.

9.3.4 Analyses

Chi-square and Student t-tests were used to assess the differences between the characteristics of respondents of the unlabelled and labelled DCE (for participants in CRC screening and for screening naïve respondents separately).

To assess feasibility we determined the response rate, rationality test outcome, missing values, and the self-rated ease of the task. We used Chi-square tests to compare differences in these aspects of feasibility.

Both DCEs were analyzed by using multinomial logit regression models, in which the unlabelled DCE had generic parameters and the labelled DCE had alternative specific parameters. These models were implemented in SAS software (Version 9.1, SAS Institute Inc., Cary, NC, USA). A priori we expected all attributes to be important, and that all attributes would have a negative effect on utility except for 'mortality reduction'.

To assess the degree of trading behaviour, we tested for dominant preferences (i.e. if respondents based their responses entirely on one specific attribute or label (one specific CRC screening test)). Chi-square tests were used to assess differences between both DCEs for participants in CRC screening and for screening naïve respondents separately.

Finally, relative utility values for different screening test profiles were determined based on the weights that respondents placed on the attributes of alternatives. The total utility value of a screening test profile was equal to the sum of the coefficient weights of its attribute levels [26-29]. The agreement between the labelled and unlabelled DCE outcomes depends strongly on the scale of both DCEs. In DCEs the scale is not identified and everything that depends on the scale is not reliable. Only measures based on correlation are really informative. Therefore, convergent validity between both variants was assessed

by determining the degree of agreement by means of Pearson correlations (r). Noteworthy, perfect *agreement* only exists if the relative utility outcomes between unlabelled and labelled DCE lie along the line of equality, whereas perfect *correlation* (i.e. strength of a relation between the two approaches) exists if the relative utility outcomes lie along any straight line [30].

9.4 Results

The total response rate was 276/712 (39%) and 1,033/2,262 (46%) for unlabelled and labelled DCE respectively (p<0.001). In total, 4 of 276 respondents (1%) and 30 of 1033 respondents (3%), who missed responses to three or more DCE questions, were excluded for further analyses. Forty-four percent of the respondents to the unlabelled DCE came from the CRC screening group and 56% from the screening naïve group; this was 53% and 47% for the labelled DCE. To correct for this imbalance all further analyses were focused on the CRC screening group and screening naïve group separately. Respondents did not differ with respect to mean age, gender, and endoscopy experience (p>0.13) for unlabelled and labelled variants respectively (Table 9.3).

Table 9.3 An example of labelled choice set

Variable	**Unlabelled DCE** (n=272)		**Labelled DCE** (n=1003)		**Unlabelled vs labelled DCE (p-value)**
Group (n-%)					0.01
CRC screening respondents	120	(44.1)	529	(52.7)	
Screening naïve respondents	152	(55.9)	474	(47.3)	
Age (mean-SD)					
CRC screening respondents	62.2	(6.3)	61.0	(7.0)	0.13
Screening naïve respondents	59.9	(5.7)	60.9	(6.6)	0.29
Gender (male; n-%)					
CRC screening respondents	59	(49.2)	277	(52.4)	0.53
Screening naïve respondents	74	(48.7)	238	(50.2)	0.74
Endoscopy experience (n-%)					
CRC screening respondents					0.26
Yes	64	(53.3)	255	(48.2)	
No	54	(45.0)	271	(51.2)	
Unknown	2	(1.7)	3	(0.6)	
Screening naïve respondents					0.43
Yes	33	(21.7)	119	(25.1)	
No	117	(77.0)	353	(74.5)	
Unknown	2	(1.3)	2	(0.4)	

Table 9.4 Differences in several aspects of feasibility

	CRC screening subjects					CRC screening naïve subjects				
	Unlabelled		Labelled		p-value[a]	Unlabelled		Labelled		p-value[a]
	n	(%)	n	(%)		n	(%)	n	(%)	
Response rate					<0.001					0.51
response	121	(57.1)	545	(70.9)		155	(31.0)	488	(32.6)	
no response	91	(42.9)	224	(29.1)		345	(69.0)	1010	(67.4)	
Rationality test					0.96					<0.001
passed	110	(90.9)	496	(91.0)		148	(95.5)	400	(82.0)	
failed	11	(9.1)	49	(9.0)		7	(4.5)	88	(18.0)	
Missing values					0.10					0.28
missing value	15	(0.8)	79	(1.3)		13	(0.5)	37	(0.8)	
no missing value	1777	(99.2)	5873	(98.7)		2371	(99.5)	4763	(99.2)	
Ease of task as perceived by subject					0.28					0.61
very easy	14	(11.7)	91	(17.2)		41	(27.0)	122	(25.7)	
easy	60	(50.0)	258	(48.8)		66	(43.4)	216	(45.6)	
average	31	(25.8)	130	(24.6)		38	(25.0)	102	(21.5)	
difficult	12	(10.0)	32	(6.0)		5	(3.3)	25	(5.3)	
very difficult	0	(0.0)	4	(0.8)		0	(0.0)	3	(0.6)	
missing values	3	(2.5)	14	(2.6)		2	(1.3)	6	(1.3)	

[a] Difference between respondents unlabelled and labelled DCE

9.4.1 Feasibility

The response rate was higher for the labelled DCE then for the unlabelled DCE (Table 9.4). The labelled DCE especially led to a higher response rate for the CRC screening group (71% vs 57%; $p<0.001$, 33% vs 31% for screening naïve group; $p=0.51$). An equal proportion of respondents of the CRC screening group passed the rationality test irrespective of DCE approach (91% vs 91%; $p=0.96$). However, more respondents failed the rationality test with the labelled design in the screening naïve group (18% vs 5% for unlabelled design; $p<0.001$). There was an equal proportion missing values of 1% for both DCEs irrespective of the response group. Most respondents indicated that they had no difficulties in completing the DCE task, and the groups did not perceive the task differently ($p=0.28$ and $p=0.61$ for CRC screening group and screening naïve group respectively) (Table 9.4).

9.4.2 DCE results

The effects (i.e. positive or negative direction) of the coefficients of both DCEs were consistent with a priori expectations (and showed therefore theoretical validity), except for the attribute 'frequency of screening' in the unlabelled approach (details in the Appendix 9.A). The positive coefficient of this attribute in the unlabelled DCE suggests that respondents preferred a higher frequency of screening over a lower frequency of screening per 10 years.

Regarding the unlabelled DCE, all attributes except the attribute 'location of screening' proved to be important for preferences of both groups for CRC screening tests (see Appendix 9.A). The positive constant term suggests that respondents from the CRC screening group preferred 'CRC screening test' over 'no CRC screening test' if all other attributes were set to zero.

Regarding the labelled DCE, all attributes proved to be important for preferences of both groups for CRC screening tests (see Appendix 9.B; note that, five out of seven attributes (i.e. location of screening, preparation for the patient, side effects of screening, complication risk, and screening duration) were attributes that had one alternative specific level; as a result their coefficients were catch up in the coefficient of the alternative label). The positive and significant alternative specific constants suggest that the CRC screening group had a positive attitude towards 'CRC screening test' over 'no CRC screening test' irrespective the utilized screening test (i.e. FOBT, sigmoidoscopy, or colonoscopy). This phenomenon was also seen in the screening naïve group, although the alternative specific constant of FOBT did not significantly differ from the base level 'no CRC screening test'.

The outcomes of the sensitivity analyses, which excluded the respondents who failed the rationality test, were quite similar whether or not these irrational responses were retained (data not shown). To avoid removal of valid preferences, induction of sample selection bias, and unnecessary reduction of the statistical efficiency and power of the estimated choice models, we included the responses of respondents who failed the rationality test in all our further analyses.

Table 9.5 Differences in repsondents' trading-behaviour

	CRC screening subjects				Screening naïve subjects			
	Unlabelled (n=120)		**Labelled** (n=529)		**p-value**[a]	**Unlabelled** (n=152)		**Labelled** (n=474)		**p-value**[a]
	n	**(%)**	**n**	**(%)**		**n**	**(%)**	**n**	**(%)**	
Trading behaviour										
in general					<0.001					0.001
dominant preferences	25	(20.8)	219	(41.4)		37	(24.3)	183	(38.6)	
non-dominant preferences	92	(76.7)	310	(58.6)		108	(73.7)	280	(59.1)	
no test at all	3	(2.5)	2	(0.4)		3	(2.0)	11	(2.3)	
dominant preferences for										
mortality reduction	22	(18.3)	91	(17.2)	0.77	36	(23.7)	81	(17.1)	0.07
frequency	0		3	(0.6)	0.41	0		4	(0.8)	0.26
alternative	n.a.		125	(23.6)		n.a.		98	(20.6)	

[a] Difference between respondents unlabelled and labelled DCE

n.a. = not applicable

9.4.3 Respondents trading behaviour

The labelled DCE led to more dominant preferences (i.e. responses entirely based on one specific attribute or label) (Table 9.5). This difference was significant for both the CRC screening group (41% vs 21% for unlabelled DCE; $p<0.001$) and the screening naïve group (39% vs 24% for unlabelled DCE; $p=0.001$). This difference was caused by the test labels; 24% and 21% of the CRC screening and screening naïve respondents, respectively, had dominant preferences for screening test labels. Table 9.5 also shows that the attributes of both DCEs did not make the difference in the proportion of dominant preferences ($0.07<p<0.77$).

9.4.4 Convergent validity

Based on the coefficients of the multinomial logit regression models of the unlabelled and labelled DCE (Appendices 9.A and 9.B), Figures 9.1a and 9.1b plot the difference in relative utility values for different realistic CRC screening programs for CRC screening naïve and CRC screening respondents respectively (see Table 9.6 for more details about

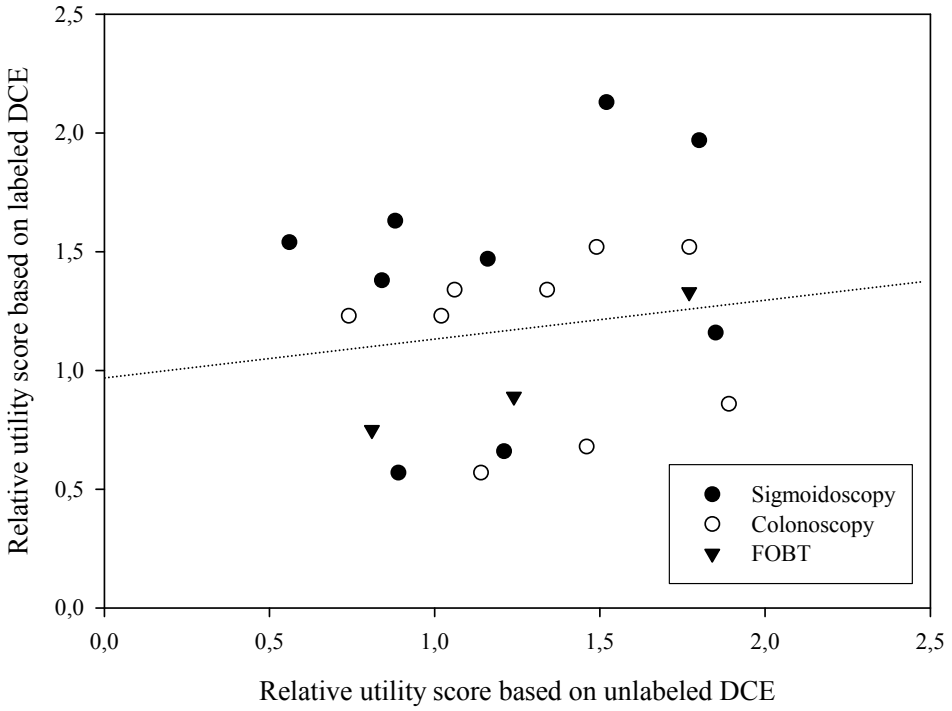


Figure 9.1a Correlation between the relative utility values of screening naïve subjeccts for different screening tests based on the labelled and unlabelled DCE (Pearson correlation 0.17; p=0.45)

the total relative utility scores). Using Pearson correlations the convergent validity between unlabelled and labelled DCEs was low for screening naïve respondents (r=0.17; p=0.45), but modest for respondents with screening experience (r=0.54; p=0.01). The regression comparison between unlabelled (independent variable) and labelled DCEs (dependent variable) showed a scaling as well as a shift phenomenon. The intercept was 0.99 (p<0.01) and 0.90 (p<0.01) and the scaling factor was 0.19 (p=0.45) and 0.51 (p=0.01) for screening naïve respondents and respondents with screening experience respectively. Respondents reacted about 0.19 or half as strong to the labelled attributes. Taking the attribute levels of frequency into account (i.e. ignoring the relative utility values of the attribute 'frequency of screening'), the strength of the relation between both approaches was reasonable good for screening naïve respondents (r=0.71; p=0.03; and r=0.53; p=0.07 for low and high frequency levels respectively), and very good for respondents with screening experience (r=0.93; p<0.001; and r=0.95; p<0.001 for low and high frequency levels respectively).

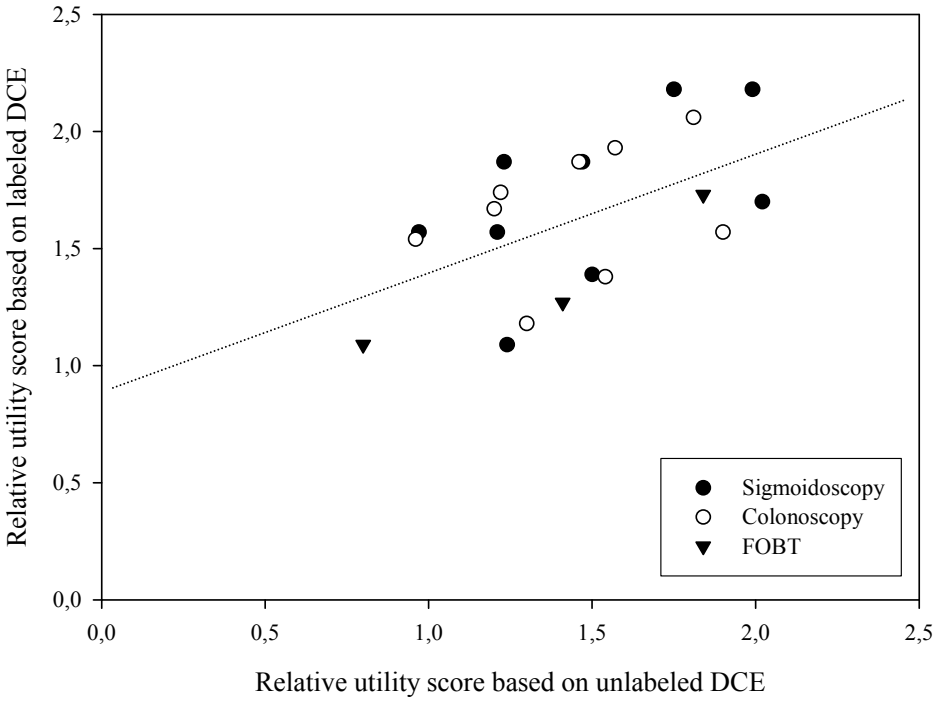


Figure 9.1b Correlation between the relative utility values of CRC screening subjects for different screening tests based on the labelled and unlabelled DCE (Pearson correlation 0.54; p=0.01)

Table 9.6 Relative utility scores for realistic CRC screening programmes

Realistic CRC screening programme (test invasiveness* / mortality risk decreases from 3% to …% / frequency per 10 years)	**Relative utilty score**			
	CRC screening respondents		**CRC screening naïve respondents**	
	unlabelled DCE	**labelled DCE**	**unlabelled DCE**	**labelled DCE**
Sigmo / 1.8% / 1	0.97	1.57	0.56	1.54
Sigmo / 1.5% / 1	1.23	1.87	0.88	1.63
Sigmo / 0.9% / 1	1.75	2.18	1.52	2.13
Sigmo / 1.8% / 2	1.21	1.57	0.84	1.38
Sigmo / 1.5% / 2	1.47	1.87	1.16	1.47
Sigmo / 0.9% / 2	1.99	2.18	1.80	1.97
Sigmo / 1.8% / 10	1.24	1.09	0.89	0.57
Sigmo / 1.5% / 10	1.50	1.39	1.21	0.66
Sigmo / 0.9% / 10	2.02	1.70	1.85	1.16
Colono / 0.8% / 1	0.96	1.54	0.74	1.23
Colono / 0.5% / 1	1.22	1.74	1.06	1.34
Colono / 0.1% / 1	1.57	1.93	1.49	1.52
Colono / 0.8% / 2	1.20	1.67	1.02	1.23
Colono / 0.5% / 2	1.46	1.87	1.34	1.34
Colono / 0.1% / 2	1.81	2.06	1.77	1.52
Colono / 0.8% / 5	1.29	1.18	1.14	0.57
Colono / 0.5% / 5	1.54	1.38	1.46	0.68
Colono / 0.1% / 5	1.90	1.57	1.89	0.86
FOBT / 2.7% / 10	0.80	1.09	0.81	0.75
FOBT / 2.3% / 10	1.41	1.27	1.24	0.89
FOBT / 1.8% / 10	1.84	1.73	1.77	1.33

* each type of test (sigmoidoscopy, colonoscopy, FOBT) had one fixed level for the following five attributes: complication risk, location of screening, screening duration, preparation for the patient, and side effects of screening; see Table 9.1 for more detail information

9.5 Discussion

This study shows that it is feasible to use realistic alternatives in labelled DCEs in a healthcare context. The labelled DCE led to a higher response rate, especially for the CRC screening group who had familiarity with the context. However, more respondents who were not familiar with the context failed the rationality test with the labelled design. The inclusion of labels appeared to play a significant role in individual choices, and increased non-trading behaviour. The convergent validity between both DCE variants was low, but better for respondents with CRC screening experience.

In health economics there are no previous publications directly comparing labelled and unlabelled DCEs empirically. However, a DCE in ecological economics considered the effects of employing a labelled rather than an unlabelled DCE [31]. That study showed that the inclusion of alternative-specific labels reduced the attention which respondents gave to the attributes (i.e. increased non-trading behaviour). This is in line with our study, which showed that 24% and 21% of the CRC screening experienced and screening naïve respondents, respectively, only focused at the screening test labels. The ecological economics study also demonstrated convergent validity between a labelled and an unlabelled DCE contrary to our study.

In line with the focus of this article, the results of the unlabelled and labelled DCE are only described briefly (for further detail information about the practical outcomes of these DCEs for colorectal cancer screening practice see Van Dam et al. [13] and Hol et al. [12]). The respondents in our labelled experiment are actually getting more and partly different information than in the unlabelled experiment, particularly if they have had experience of one of the options. This might be a possible explanation for the differences in our outcomes between the screening naïve and CRC screening respondents. Note that, if the reader wants to compare the beta-coefficients of CRC screening respondents and screening naïve respondents directly (see Appendix 9.A), it should be clear that scale effects might be an issue (for more detail information see Swait and Louviere [32]).

The positive direction (effect) of the attribute 'frequency of screening per 10 years' in the unlabelled approach (Appendix 9.A) seems to be inconsistent with utility theory. However, these 'irrational' responses may be explained by respondents making additional assumptions or bringing additional information to the choice [33, 34]. As Ryan et al. [35] provided evidence that respondents assumed tests with higher costs would be of higher quality. Respondents in our study might associate higher frequency of screening with a more effective test. The differences in preferences for screening frequency between the two DCE approaches demonstrates the importance of continuing research into the biases present across these elicitation methods. Mixed methods may be useful to get more insight into the internal validity of the DCEs. Qualitative techniques, such as the think aloud technique, may show that seemingly 'irrational' choice behaviour may not be so irrational after all [35].

The predominant use of unlabelled experiments in health care may be a result of the perception that labelled experiments are difficult to construct. The design of a labelled

DCE does generally mean a larger sample size is required, because it is assumable that most of time there are interactions between the alternative label and the attributes. Indeed, this may not be feasible in a health care setting (e.g. the target group of patients or medical specialists is too small). However, this is not only the case for labelled DCE, but also for unlabelled DCE in which all (two-way) interactions between attributes are taken into account. Unlabelled DCEs in which all (two-way) interactions between attributes are taken into account may be even much larger then a labelled DCE, because in a labelled DCE a lot of characteristics can be compressed in one label, whereas in an unlabelled DCE all possible interactions should be taken into account.

Another explanation for the predominant use of unlabelled experiments in health care may be that labelled DCEs in health care are not necessary (yet). Although it is not clear why labelled DCEs in health economics are rarely used, it has to be clear that the design should be made to fit the research objectives and not the other way around. If it is to be expected in the experiment that the alternative labels have important differences, then it may be desirable to go for a labelled DCE. Underestimating the role of the alternative labels may lead to worse or even wrong predictions of alternatives people actually prefer. On the other side, if the objective is to estimate attribute values, it may be desirable to use an unlabelled DCE to reduce non-trading behaviour due to alternative labels.

This study had some limitations. First, we conducted two DCEs in two samples. It might have been preferable (from a theoretical point of view) to conduct the two DCEs in the same group of respondents (i.e. all respondents filled in one DCE, and then the other DCE; sequence in random order). However, that was not possible because of the respondent burden. As a result, we cannot directly compare the absolute values of the utility levels for the attributes and tests. Second, the design of both DCEs was not exactly the same. The combination of d-efficiency criteria and the use of alternative specific and generic attribute levels in the labelled and unlabelled DCE respectively, resulted in different choice sets presented to the respondents. We have no reason to believe that this has influenced the results to a large extent. Third, testing the convergent validity between unlabelled and labelled DCE was based on comparison of the total utility of alternatives. The labelled DCE had five attributes with one alternative specific level, and therefore a direct comparison of the coefficients of the attributes (taking scale factor into account) was not possible. Fourth, two attribute levels regarding the alternative specific attribute 'frequency' of FOBT (3 and 30 times screening per 10 years) were not presented in the unlabelled DCE. Therefore, we could only include three total utility scores of (hypothetical) CRC screening programs with FOBT-test in our convergent validity test between both DCE variants. In conclusion, this study provides important insights in the feasibility and difference in results from labelled and unlabelled DCEs. The inclusion of labels appeared to play a significant role in individual choices, but reduced the attention respondents give to the attributes. There was low convergent validity between both DCE variants, largely due to different preferences for screening frequency. The choice for a labelled or unlabelled DCE may depend on the type of respondents and the research question. Unlabelled DCEs may be more suitable to investigate trade-offs between

attributes and for respondents who do not have familiarity with the alternative labels, whereas labelled DCEs may be more suitable to explain real life choices such as uptake of cancer screening.

References

1. Lancsar E, Louviere JJ (2008). Conducting discrete choice experiments to inform healthcare decision making: a user's guide. Pharmacoeconomics; 26(8):661-77.
2. Ryan M, Gerard K, Amaya-Amaya M (2008). Using Discrete Choice Experiments to Value Health and Health Care. (Vol. 11). Dordrecht, The Netherlands: Springer.
3. Bliemer MCJ, Rose JM (2006). Designing Stated Choice Experiments: State-of-the-art. The 11th International Conference on Travel Behaviour Research. Kyoto.
4. Hensher DA, Rose JM, Greene WH (2005). Applied choice analysis: a primer. Cambridge, Cambridge University Press.
5. Louviere JJ, Hensher DA, Swait JD (2000). Stated choice methods: analysis and application. Cambridge, Cambridge University Press.
6. Kruijshaar ME, Essink-Bot ML, Donkers B, Looman CW, Siersema PD, Steyerberg EW (2009). A labelled discrete choice experiment adds realism to the choices presented: preferences for surveillance tests for Barrett esophagus. BMC Med Res Methodol; 19(9):31.
7. Mitchell RC, Carson R (1989). Using Surveys to Value Public Goods: The Contingent Valuation Method. Resources for the future. Washington.
8. Ferlay J, Autier P, Bonial M, et al (2007). Estimates of the cancer incidence and mortality in Europe in 2006. Ann Oncol; 18(3): 581-92.
9. Jemal A, Siegel R, Ward E, et al (2007). Cancer statistics, 2007. CA Cancer J Clinicians; 57(1): 43-66.
10. Health Council of the Netherlands (2008). Screening: between hope and hype. The Hague, Health Council of the Netherlands.
11. Levin B, Lieberman DA, McFarland B, et al (2008). A joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. Cancer J Clinicians; 58(3):130-60.
12. Hol L, de Bekker-Grob EW, van Dam L, Donkers B, Kuipers EJ, Habbema JDF, Steyerberg EW, van Leerdam ME, Essink-Bot ML. Population preferences for different colorectal cancer screening strategies in the Netherlands: a discrete choice experiment. *Submitted.*
13. Van Dam L, Hol L, de Bekker-Grob EW, Steyerberg EW, Kuipers EJ, Habbema JDF, Essink-Bot ML, van Leerdam ME (2009). What determines individuals' preferences for colorectal cancer screening programmes? A discrete choice experiment. Eur J Cancer [*forthcoming*].
14. Sloane NJA (2009). A library of orthogonal arrays. http://www.research.att.com/ ~njas/oadir/. [Accessed 2009 April 12].
15. Street DJ, Burgess L, Louviere J (2005). Quick and easy choice sets: constructing optimal and nearly optimal stated choice experiments. Int J Res Market; 22: 459-70.
16. Street D, Burgess L (2007). Discrete choice experiments [computer software]. Sydney: University of Technology [online]. Available from URL: http://maths.science.uts.edu.au/maths/ wiki/SPExpts [Accessed 2009 Mar 05].
17. Ryan M, Major K, Skâtun D (2005). Using discrete choice experiments to go beyond clinical outcomes when evaluating clinical practice. J Eval Clin Pract; 11(4):328-38.

18. Weston A, Fitzgerald P (2004). Discrete choice experiment to derive willingness to pay for methyl aminolevulinate photodynamic therapy versus simple excision surgery in basal cell carcinoma. Pharmacoeconomics; 22(18):1195-1208.
19. Baltussen R, ten Asbroek AH, Koolman X, Shrestha N, Bhattarai P, Niessen LW (2007). Priority setting using multiple criteria: should a lung health programme be implemented in Nepal? Health Policy Plan; 22(3):178-85.
20. Goto R, Nishimura S, Ida T (2007). Discrete choice experiment of smoking cessation behaviour in Japan. Tob Control; 16(5):336-43.
21. Langenhoff BS, Krabbe PF, Ruers TJ (2007). Computer-based decision making in medicine: A model for surgery of colorectal liver metastases. Eur J Surg Oncol; 33 Suppl 2:S111-7.
22. Lancsar E, Louviere J (2006). Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? Health Econ; 15(8):797-811.
23. Walzer S (2007). What do parents want from their child's asthma treatment? Ther Clin Risk Manag; 3(1):167-75.
24. Lloyd A, Doyle S, Dewilde S, Turk F (2008). Preferences and utilities for the symptoms of moderate to severe allergic asthma. Eur J Health Econ; 9(3):275-84.
25. Watson V, Ryan M, Brown CT, Barnett G, Ellis BW, Emberton M (2004). Eliciting preferences for drug treatment of lower urinary tract symptoms associated with benign prostatic hyperplasia. J Urol; 172(6 Pt 1):2321-2325.
26. Bishop AJ, Marteau TM, Armstrong D, Chitty LS, Longworth L, Buxton MJ, Berlin C (2004). Women and health care professionals' preferences for Down's Syndrome screening tests: a conjoint analysis study. BJOG; 111(8):775-9.
27. Baltussen R, Stolk E, Chisholm D, Aikins M (2006). Towards a multi-criteria approach for priority setting: an application to Ghana. Health Econ; 15(7):689-96.
28. Papanikolaou P, Lyne P, Ratcliffe J (2007). Using the discrete choice experimental design to investigate decision-making about pressure ulcer prevention by community nurses.Health Soc Care Community; 15(6):588-98.
29. De Bekker-Grob EW, Essink-Bot ML, Meerding WJ, Koes BW, Steyerberg EW (2009). Preferences of GPs and patients for preventive osteoporosis drug treatment: a discrete-choice experiment.. Pharmacoeconomics; 27(3):211-9.
30. Bland JM, Altman DG (1986). Statistical methods for assessing agreement between two methods of clinical measurement. Lancet; 1:307-10.
31. Blamey RK, Bennett JW, Louviere JJ, et al (2000). A test of policy labels in environmental choice modeling studies. Ecol Econ; 32: 269-86.
32.. Swait J, Louviere J (1993). The role of the scale parameter in the estimation and comparison of multinomial logit models. Journal of Marketing Research; 30(3): 305-314
33. Ryan M, San Miguel F (2000). Testing for consistency in willingness to pay experiments. J Econ Psychol; 21: 305-17.
34. San Miguel F, Ryan M, Amaya-Amaya M (2005). Irrational stated preferences: a quantitative and qualitative investigation. Health Econ; 14(3): 307-22.
35. Ryan M, Watson V, Entwistle V (2009). Rationalising the 'irrational': a think aloud study of discrete choice experiment responses. Health Econ; 18(3):321-36.

Appendix 9.A

Multinomial logit results for unlabelled DCE

	CRC screening respondents			Screening naïve respondents		
Attribute	**coefficient**	**error**	**p-value**	**coefficient**	**error**	**p-value**
Constant (screening test)	0.62	0.14	<0.001	0.25	0.13	0.05
(base level 'no test')						
Side effects of screening	-0.23	0.06	<0.001	-0.31	0.06	<0.001
(base level 'no pain')						
Complication risk	-0.13	0.06	0.04	-0.16	0.06	<0.01
(base level 'no risk')						
Location of screening	-0.01	0.06	0.83	-0.09	0.06	0.11
(base level 'home')						
Preparation for patient			<0.001			<0.001
(base level 'none')						
enema no fasting	-0.23	0.11		-0.37	0.10	
drinking 0.75l + fasting	-0.22	0.12		-0.51	0.11	
drinking 4l + fasting	-0.88	0.11		-0.98	0.11	
Screening duration (per 10 min)	-0.03	0.01	<0.01	-0.03	0.01	<0.01
Frequency of screening			<0.001			<0.001
(base level 1 time)						
2 times	0.24	0.09		0.28	0.09	
5 times	0.33	0.10		0.40	0.10	
10 times	0.27	0.09		0.33	0.08	
Mortality reduction (per 0.3%)	0.26	0.01	<0.001	0.32	0.01	<0.001

Appendix 9.B

Multinomial logit results for labelled DCE

Attribute	CRC screening respondents			Screening naïve respondents		
	coefficient	error	p-value	coefficient	error	p-value
Label (base level 'no test')						
FOBT	0.49	0.11	<0.001	0.16	0.12	0.16
Sigmoidoscopy	1.09	0.11	<0.001	0.57	0.11	<0.001
Colonoscopy	1.18	0.11	<0.001	0.57	0.11	<0.001
Mortality reduction FOBT			<0.001			<0.001
(base level from 3.0% to 2.7%)						
from 3.0% to 2.3%	0.18	0.09		0.14	0.09	
from 3.0% to 1.8%	0.64	0.10		0.58	0.11	
Mortality reduction Sigmo			<0.001			<0.001
(base level from 3.0% to 1.8%)						
from 3.0% to 1.5%	0.30	0.08		0.09	0.09	
from 3.0% to 0.9%	0.61	0.10		0.59	0.11	
Mortality reduction Colono			<0.001			<0.001
(base level from 3.0% to 0.8%)						
from 3.0% to 0.5%	0.20	0.08		0.11	0.09	
from 3.0% to 0.1%	0.39	0.10		0.29	0.10	
Frequency of screening FOBT			<0.001			<0.001
(base level 30 times)						
10 times	0.60	0.09		0.59	0.09	
3 times	0.62	0.10		0.78	0.11	
Frequency of screening Sigmo			<0.001			<0.001
(base level 10 times)						
2 times	0.48	0.08		0.81	0.09	
1 time	0.48	0.10		0.97	0.10	
Frequency of screening Colono			<0.001			<0.001
(base level 5 times)						
2 times	0.49	0.08		0.66	0.09	
1 time	0.36	0.10		0.66	0.09	

10

General discussion

10 Discrete choice experiments in health care: discussion

This thesis addresses applications and theoretical aspects of DCE in health care. In this chapter, the main findings regarding the objectives of the thesis will be presented first. Subsequently, theoretical aspects of DCE and application possibilities of DCE will be discussed. This chapter ends with conclusions and recommendations for future research.

10.1 Main findings

Objective 1: To provide an overview of current DCE practice in health economics, and to compare and assess progress in methodological advances.

We showed in Chapter 2 that between 2001-2008 114 DCE papers were published covering a range of policy questions. These included valuing patient experience factors, health outcomes, investigating trade-offs between health outcomes and patient experience factors, estimating utility weights within the QALY framework, investigating labour-market choices amongst health care professional, priority setting frameworks at the local/national level and preferences regarding clinical decision making. This means that the applications today are much broader than to value patient experience factors, and shows that a DCE is a potentially useful tool for investigating preferences for all kind of health care issues. The condition is that the choice problem can be broken down into discrete choice options characterised by limited numbers of attributes and levels. Looking at the results of these DCE papers we think evidence is accumulating that DCEs can contribute to benefit measurement for use in economic evaluation. DCEs are useful to measure process attributes and non-health outcomes, and to determine the trade-offs between these attributes and health outcomes.
We are less convinced that these stated preference outcomes are useful to predict actual behaviour (i.e. real preferences). No research has considered external validity (i.e., testing whether individuals behave in reality as they stated in the survey). A good development, regarding the technical quality of DCEs, is that there has been a shift towards using statistically efficient designs. The construction of the experimental design is a crucial part of any DCE. Constructing choices randomly do harm the precision of the parameter estimates or may even bias the results. We also showed in Chapter 2 that richer econometric models are beginning to be employed for analysis. On the one side we think that the application of richer econometric models is a positive development to relax assumptions such as 'errors are independent and identically distributed', 'observed choices are independent', and 'preferences are homogenous' that may bias the results. On the other side we discourage to use these richer models if data are insufficient (e.g. a small number of respondents), because in that case these models will do more harm than good.

Objective 2: To study the usefulness of DCE for real choice problems in health care for measuring benefits beyond health outcomes.

We showed in chapters 3-8 that studies in this thesis illustrated the feasibility of DCE to elicit valuations for health and non-health outcomes of individuals aged 18-89 years for a wide range of health care interventions. Only a small fraction of respondents failed the rationality tests (1-13%). This suggests that adults can handle DCEs. However, the fraction of potential respondents who agreed to participate in the DCEs differed largely (31-86%), and was especially low (31-33%) for the general population who had not experienced the choice problem by themselves. We do not know the reasons of non-response and the characteristics of the non-respondents neither. Therefore we have to make a marginal note that DCEs probably cause some extra selection bias compared with general questionnaire surveys due to the complexity of a DCE task. Given the complexity of a DCE task, selection of respondents by education level or language skill is likely.
The DCE studies conducted in this thesis provided evidence that the DCE approach can be applied successfully in health care to directly compare the determinants of the preferences from different groups (e.g., health professionals, patients, general population). The DCE studies all gave insight into whether or not the given (non-)health attributes were important (according to a significant statistical test), the relative importance of the attributes, the rate at which individuals were willing to trade between attributes, and overall benefit scores for intervention alternatives. Confidence may be larger for the usefulness of results showing the importance of each attribute (i.e. relative values), than for in the usefulness of results as predictors for real choice (i.e. absolute values).

Objective 3: To compare labelled versus unlabelled DCEs in health care in various aspects of feasibility, trading-behaviour, and convergent validity.

We showed in Chapter 9 that it is feasible to use realistic alternatives in labelled DCEs in a healthcare context. The labelled DCE led to a higher response rate, especially for the CRC screening group who had familiarity with the context. However, more respondents who were not familiar with the context failed the rationality test with the labelled design. This suggests that unlabelled DCEs are more suitable for respondents who are not familiar with the alternative labels. The inclusion of labels appeared to play a significant role in individual choices, and increased non-trading behaviour. Labelled DCEs may be more useful to predict subjects' behaviour, whereas unlabelled DCEs may be more useful to investigate the trade-offs between attributes of a health care intervention or service. Consequently, the choice of a labelled or an unlabelled DCE depends on the research question. Practical reasons can also play a role. Labelled DCEs quickly result in large experimental designs, which require a large number of respondents.
The convergent validity between both DCE variants was low, but better for subjects with CRC screening experience. This outcome again stresses that the choice of a labelled or un unlabelled DCE depends on the research question. The current dominance of unlabelled designs in the health economic area may not be defensible.

10.2 Theoretical aspects of DCE

10.2.1 Experimental design

Constructing an experimental design is a crucial part of a DCE. A full factorial design includes all possible combinations of attributes and attribute levels in the choice set. When using a full factorial design all interaction terms between attributes and/or attribute levels can be estimated. However, a full factorial design often results in a large number of alternatives (i.e. profiles). Therefore experimental design methods are used to create a fractional factorial design. Several types of design exist to create a fractional factorial design. In health economics, mostly orthogonal arrays have been used. These arrays consider orthogonality and level balance. Choice sets can then be manually created. We observed a move from constructing choices randomly to constructing choices using the fold-over technique. The fold-over technique minimizes the overlap between attribute levels in a choice set. We demonstrated in various studies (chapters 4-7) that this method resulted in highly optimal designs (95-100% efficient), in which all main effects were uncorrelated.

As demonstrated in Chapter 2, experimental design methods in health economics have progressed in recent years. We observed a move from orthogonal designs to constructing statistically efficient designs. D-efficiency (i.e. precision of parameter estimates) has been a popular criterion to measure statistical efficiency. Here the variance–covariance matrix is minimized and as a result information obtained from observed choices is maximised [1]; (orthogonal designs do not take dominant alternatives into account, whereas statistically efficient designs do). Whilst such designs consider orthogonality, level balance and minimum overlap, they do not necessary minimize these.

We showed in Chapter 8 that a labelled DCE has to take into account all two-way interactions (since the alternative labels may be correlated with the attributes) and implausible combinations of attribute levels and labels. Orthogonal designs are not available in such circumstances. Using a statistically efficient design is more appropriate then [2]. Statistical efficient designs maximise D-efficiency or equivalently minimize D-error. Using specific software as SAS software or Ngene software choice sets can be generate directly [3].

More recently, in other areas than health economics efficient choice sets are being created by making prior assumptions about the parameters to be estimated [4-16]. This Bayesian approach lets go of the principle of orthogonality and constructs designs in a manner that is intended to minimise the expected elements of the asymptomatic variance covariance (AVC) matrix that will be obtained from models estimated on data collected using that design. The primary advantage of this method is that the constructed design is related to the expected outcome of the modelling process. This method has not been used in health care (yet). Therefore research is needed to investigate its usefulness in a health care context. It may be worthwhile for the health economics community to link to other areas. Chapter 2 showed that a main effects-only fractional factorial design (i.e., assuming that

all interactions between attributes were zero) was used in 89% of the DCE studies conducted during 2001-2008 in health care. Also in the DCE studies in this thesis a main effects only fractional factorial design was commonly used. Although these generally account for 70 to 90% of explained variance in a DCE [17], specific combinations of attribute levels may have specific effects that remained unidentified. As a result the main effects-only fractional factorial designs used in our DCE studies were mentioned as a limitation. However, if a researcher intends to estimate interaction terms, more degrees of freedom are required for estimation purposes. More degrees of freedom mean larger designs. We support the recommendation of Lancsar and Louviere [18] to use designs which allow estimation of all two-way interactions to minimise the potential for bias.
In Chapter 9 we advocated that unlabelled DCEs may be more suitable to investigate trade-offs between attributes and for respondents who do not have familiarity with the alternative labels, whereas labelled DCEs may be more suitable to explain real life choices such as uptake of cancer screening. It depends on the choice problem if choice sets should be generated in a labelled or unlabelled form. If in the experiment is expected that the alternatives have important differences, and are wished to be modelled, then a labelled experiment should be used. Therefore, the current dominant use of unlabelled DCEs in the health economic area may not be defensible.

10.2.2 Estimation procedures

After data are collected the analyst chooses a modelling strategy to analyse responses. McFadden's multinomial logit (MNL) model provides a clear link between choice behaviour and utility and associated set of assumptions and properties. The MNL model assumes independent and identically Gumbel distributed error terms to characterise stochastic utility (ε) [19]. The undoubted popularity of the MNL is due to its simplicity and ease of estimation. This model has three assumptions: independence of irrelevant alternatives (IIA) (i.e., for a given individual, the ratio of the choice probabilities of any two alternatives is unaffected by other alternatives); constant error term across observations (therefore not allowing for multiple observations from respondents); and homogeneity of preferences. Recognition of the restrictive nature of these assumptions has led to developments of the MNL model.
Chapter 2 showed that 10% of the DCE studies conducted between 2001-2008 used richer econometric methods (i.e. nested logit, mixed logit or latent logit models) compared with 3% of the DCE studies conducted between 1990-2000. It appears that using richer econometric models for estimation is slowly being taken up, particularly with respect to preference heterogeneity. Wansbeek et al. [20] has singled heterogeneity out as 'the most salient feature of consumer demand at the micro level'. However, caution is needed when choosing a more advanced model for analysis, given the data requirements of these more flexible econometric specifications [21]. Especially in case of relatively small number of cases, such as mentioned in Chapter 4 (n=39), the estimation of a mixed model will not make sense. The estimation might become instable, and the MNL model

might fit better. Therefore we recommend researchers to select the econometric model to analyse responses, which is in balance with the available data.
In the area of econometric modelling outside health economics there have been many advances of late [22-29]. It may be worthwhile for the health economics community to link to these other areas for keeping pace with econometric modelling.

10.2.3 Complexity of DCE

The ability of respondents to fully comprehend valuation tasks has always been the subject of some scepticism. The difficulty of a DCE task and in other valuation tasks (e.g. the time trade-off (TTO) method, or standard gamble (SG) method) is an issue of attention. Craig et al. [30] showed that illiteracy and innumeracy can hinder implementation of complex preference elicitation techniques (TTO and VAS) in diverse settings and populations. Selection bias appears if subjects drop out in a valuation task due to their cognitive ability. This dropping out can take place in different ways. Subjects may not participate in the DCE task at all (unusual questionnaire), or if these subjects participate in the for them difficult DCE task they cause instable parameter estimates. Therefore, we recommend to keep the DCE survey simple: the questionnaire should be as clear as possible (e.g., difficult linguistic use must be avoided), and the DCE task should be as simple as possible (e.g., the number of attributes, and innumeracy and literacy problems must be taken into account). We also recommend to test if respondents can handle the DCE survey. If respondents can not handle the DCE survey (e.g., due to literacy problems or problems with rational decision-making), than do not conduct the DCE survey.
Studies from other areas have found that task complexity affects DCE parameter estimates, increasing the unexplained model variance [31-36]. Furthermore, the concern of design strategies seeking to maximise statistical efficiency may come at the cost of lower respondent efficiency [37-38]. We do not know which specific aspects of a DCE survey are difficult for which group of respondents. Qualitative research methods may be useful in this respect. Huston and Rowan [39] found that qualitative research can provide important insights into questionnaire behaviour. Our DCE studies used qualitative methods as additional information for the selection of attributes and levels, and for piloting of questionnaires. We suggest to extend the role of qualitative research in the field of DCE, to investigate the choice processes in respondents who complete a DCE and to evaluate the precise meanings of their responses, using formal methods such as the Three-Step Test Interview [40]. In quite another research area, such research resulted in very useful insights for the interpretation of seemingly counterintuitive patterns of health-related quality of life scores found in men after treatment for localized prostate cancer [41].

10.2.4 Validity of DCE

Given DCEs, like other health economic valuation techniques, rely on choices in hypothetical situations, investigating external validity of responses has been identified as

important [42]. Evidence of external validity (testing whether individuals behave in reality as they state in the hypothetical survey) is the strongest form of evidence for validity. Given that many applications of DCEs have taken place in countries with a publicly provided health care system, availability of data on actual choices to compare real and stated behaviour is limited [43]. DCE studies from other areas than health economics found evidence of external validity [44-48]. Chapter 2 showed that none of the DCE studies conducted during 2001-2008 in health care has considered external validity. Also none of the DCE studies conducted in this thesis was tested for external validity. Testing for external validity of the DCE approach seems to be difficult in a health care context. As a result, a number of internal validity tests have been applied. For example, testing theoretical validity (i.e. checking if the model coefficients have the sign expected given theory) or 'rationality' of responses (checking if respondents prefer more of a good thing rather than less of it).

The DCEs conducted in this thesis showed theoretical validity. Large fractions of our respondents passed the rationality tests (87-99%), which provided evidence to internal consistency. We tested the "rationality" of responses by including dominance tests (i.e. choice sets where one alternative is clearly superior). These tests have been criticised for being too easy to pass [49]. Empirical research is needed to find out if a transitivity test (i.e. if an individual prefers alternative A over B, and B over C then the individual should prefer A over C) or a Sen's contraction property test (i.e. an individual who chooses alternative A in the first choice set of for example four alternatives (A, B, C, plus an opt-out option) should not choose alternative B or the opt-out alternative in the reduced choice set of three alternatives (A, B, plus an opt-out option) [50]) are more appropriate to test the rationality of responses.

Within study constraints it is important to continue with internal validity tests. Nevertheless, testing for external validity is crucial, an area health economists must get into. Again the health economics area can learn from other areas. For example, Carlsson and Martinsson [45], Chang et al. [46], and Lusk and Schroeder [47] used a laboratory experiment to test the external validity of DCEs. Making use of such external validity tests may be a way forward to test external validity in health care.

10.3 Applications of DCE

The DCEs conducted in this thesis provide valuable insights regarding benefit valuation beyond clinical outcomes of various health care interventions. They all gave insight into the relative importance of the attributes; the rate at which individuals were willing to trade between attributes; and overall benefit scores for intervention alternatives. Therefore our DCEs provide useful insights regarding benefit valuation of various health care interventions' characteristics, which would be 'ignored' in clinical outcomes or QALYs, and might be useful as additional information in a medical-decision process about health care interventions or services.

Chapter 3 showed that the willingness to pay (WTP) for each (non-)health outcome attribute could be indirectly determined, because a cost attribute was included in that DCE. The possibility to estimate a WTP is an additional advantage of DCE. However, to our opinion this additional advantage is limited. Slothuus and Gyrd-Hansen [51] emphasised that interpretation of implicit WTP values should be tackled with caution. They showed that a wider cost range including higher payments is associated with lower parameter weight associated with the payment variable, and thus increased WTP values. Ryan and Watson [52] showed that the WTP estimate derived from the DCE was higher than that derived from the payment card contingent valuation, and Van der Pol et al. [53] showed that the mean WTP derived from an open-ended question was substantially lower than the mean derived from a DCE. As an aside, the inclusion of a cost-attribute has some important disadvantages such as it may be unrealistic for countries with a publicly provided health care system, or price sensitivity may change when the order of the cost attribute is moved in the alternative description [54]. In summary, the possibility to estimate a WTP is an additional option of DCE. Only if relevant in health care we recommend to include a cost attribute. However, the results should be interpreted with caution: the absolute WTP values derived from a DCE should be interpreted relatively (i.e. ranking order).

A disadvantage of carrying out a DCE is that the statistical experimental design of a DCE and the econometric modelling have become quite complex. Practitioners using DCE for the first time are recommended to have expert help from a specialist. We are convinced that it is important to conduct a DCE methodological correctly and to do research on it. However, it should be balanced with the relevance of answering the research question. In other words, doing a lot of research to generate a methodologically perfect DCE does not make sense if respondents can not understand the valuation task anymore or if it will be unpractical in a health care context (e.g. too many patients are required). A DCE should be as good as needed, and does not have to be as good as possible. The preparation required for applying a DCE is another disadvantage. The determination of attributes and attribute levels are crucial, and the number of attributes in a DCE is limited, because increasing numbers of attributes would impact on the random component variability [17]. As a result extensive literature research is needed and qualitative data collection is recommendable (e.g. focus groups, individual interviews with target subjects or experts). A pilot study is recommendable to ascertain that subjects can manage the length of the questionnaire and to examine the intelligibility, acceptability and validity of the questionnaire. A final issue we experienced was that we had to report the DCE approach and outcomes relatively simple to make it acceptable for clinical journals. As a result researchers who want to replicate our DCE studies will face some lack of methodological information.

10.4 Conclusions and recommendations for future research

Conclusions

- The DCEs conducted in this thesis provided valuable insights regarding benefit valuation beyond clinical outcomes of various health care interventions, and are therefore useful as additional information to clinical outcomes or QALYs in a medical decision-making process about health care interventions.
- The DCE approach is a feasible instrument for the measurement of preferences of adults for all kind of health care interventions under the condition that the choice problem can be broken down into discrete choice options characterised by limited numbers of attributes and attribute levels.
- The DCE approach is a useful tool in health care to directly compare the determinants of the preferences for health care interventions from different groups such as health professionals, patients, and general population.
- DCE outcomes may be more appropriate to show the relative importance of the attributes of a health care intervention than to predict real choice.
- Unlabelled DCEs may be more suitable to investigate trade-offs between attributes and for respondents who do not have familiarity with the alternative labels, whereas labelled DCEs may be more suitable to explain real life choices.
- The possibility to estimate a WTP is an additional advantage of DCE, but the results should be interpreted relatively.
- In terms of experimental design and estimation procedure it may be worthwhile for the health economics community to link with areas outside health economics.
- More research to generate a methodologically perfect DCE is not useful if respondents can not understand the valuation task anymore or if it will be unpractical in a health care context.

Recommendations for future research

Future research is needed:

- To define the place of DCEs versus other preference elicitation methods in health care.
- To generate DCEs which are as easy as possible for respondents, but still provide adequate answers to the research question.
- To optimise the design and analysis of DCEs, including issues as the usefulness of non-orthogonal efficient designs and dealing with 'irrational' responses.
- To find opportunities to test the external validity of DCEs in health care.

References

1. Street DJ, Bunch DS, Moore BJ (2001). Optimal designs for 2^k paired comparison experiments. Commun Stat - Theor M 30:2149-2171.
2. Street DJ, Burgess L, Louviere J (2005). Quick and easy choice sets: constructing optimal and nearly optimal stated choice experiments. Intern J of Research in Marketing 22:459-470.
3. Kuhfeld W (2000). Marketing Research Methods in the SAS System, Version 8 Edition. SAS Institute.
4. Bliemer MCJ, Rose JM, Hensher DA (2009). Efficient stated choice experiments allowing for estimating nested logit models. Transp Res Part B 43(1):19-35.
5. Ferrini S, Scarpa R (2007). Designs with a priori information for nonmarket valuation with choice experiments: A Monte Carlo study. J Environ Econ Manage 53(3):342-363.
6. Fowkes AS (2000). Recent developments in stated preference techniques in transport research. Stated Preference Modelling Techniques pp.37-52 (published by PTRC edited by J de D Ortuzar).
7. Huber J, Zwerina K (1996). The importance of utility balance in efficient choice designs. J Marketing Res 33:307-317.
8. Kessels R, Goos P, Vandebroek M (2006). A Comparison of Criteria to Design Efficient Choice Experiments. J Marketing Res 43(8)409-419.
9. Sandor Z, Wedel M (2001). Designing conjoint choice experiment using mangers' prior beliefs. J Marketing Res 38:430-443.
10. Sandor Z, Wedel M (2002). Profile construction in experimental choice designs for mixed logit models. Marketing Sci 21:455-475.
11. Sandor Z, Wedel M (2005). Heterogeneous conjoint choice designs. J Marketing Res 42:210-218.
12. Rose JM, Bliemer MCJ (2008). Stated preference experimental design strategies, in Hensher DA and Button KJ (eds). Handbook of Transport Modelling, Elsevier, Oxford, Ch 8, 151-180.
13. Rose JM, Bliemer MCJ (2005). Sample optimality in the design of stated choice experiments. Working paper ITLS-WP-05-13.
14. Rose JM, Bliemer MCJ, Hensher DA, Collins AT (2008). Designing efficient stated choice experiments in the presence of reference alternatives. Transp Res Part B: Methodological 42(4):395-406.
15. Watson SM (2000). Efficiency properties of orthogonal stated preference designs Stated Preference Modelling Techniques pp.91-101 (published by PTRC edited by J de D Ortuzar).
16. Kanninen B (2002). Optimal designs for multinomial choice experiments. J Marketing Res 39:214-227.
17. Louviere JJ, Hensher DA, Swait JD (2000). Stated choice methods: analysis and application. Cambridge University Press, Cambridge.
18. Lancsar E, Louviere J (2008). Conducting Discrete Choice Experiments to Inform Healthcare Decision Making: A User's Guide. Pharmacoecon 26:661-677.
19. McFadden, D (1974). Conditional logit analysis of qualitative choice behaviour. In P. Zarembka (Ed.), Frontiers in Econometrics (pp. 105-142). New York: Academic Press.
20. Wansbeek T, Meijer E, Wedel M (2001). Comment on microeconomics. J Econometrics 100/101:89-91.
21. Louviere J (2006). What you don't know might hurt you: some unresolved issues in the design and analysis of discrete choice experiments. Environ Resource Econ 34:173-188.
22. Bhat CR (2003). Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences Transp Res B 37(9):837-855.

23. Bhat CR (2001). Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model. Transp Res Part B 35(7):677-693.
24. Fosgerau M (2007). Using nonparametrics to specify a model to measure the value of travel time. Transp Res Part A 41(9):842-856.
25. Fosgerau M, Bierlaire M (2007). A practical test for the choice of mixing distribution in discrete choice models. Transp Res Part B 4 (7)784-794.
26. Greene W, Hensher DA, Rose JM (2006). Accounting for heterogeneity in the variance of the unobserved effects in mixed logit models. Transp Res Part B 40(1):75-92.
27. Hensher DA, Greene W (2003). The mixed logit model: the state of practice. Transp 30:133-176.
28. Hess S, Train KE, Polak JW (2005). On the use of a Modified Latin Hypercube Sampling (MLHS) approach in the estimation of a Mixed Logit model for vehicle choice. Transp Res Part B 40(2):147-163.
29. Sandor A, Train K (2004). Quasi-random simulation of discrete choice models. Transp Res Part B 38:313-327.
30. Craig BM, Busschbach JJV, Salomon JA (2009). Keep it simple: Ranking health states yields values similar to cardinal measurement approaches. J Clin Epidemiol 62(3):296-305.
31. Caussade S, Ortúzar JD, Rizzi LI, Hensher DA (2005). Assessing the influence of design dimensions on stated choice experiment estimates. Transp Res Part B 39(7):621-640.
32. DeShazo JR, Fermo G (2002). Designing choice sets for stated preference methods: the effects of complexity on choice consistency. J Environ Econ Manage 44:123-143.
33. Islam T, Louviere JJ, Burke PF (2007). Modelling the effects of including/excluding attributes in choice experiment on systematic and random components. International Journal of Research Marketing 24:289-300.
34. Mazzotta M, Opaluch J (1995). Decision making when choices are complex: a test of Heiner's hypothesis. Land Econ 7:500-515.
35. Swait J, Adamowicz W (2001). The influence of task complexity on consumer choice: a latent class model of decision strategy switching. J Cons Res 28(1):135-48.
36. Swait J, Adamowicz W (2001). Choice environment, market complexity, and consumer behaviour: a theoretical and empirical approach for incorporating decision complexity into models of consumer choice. Organ Behav Hum Decis Process 86(2):141-67.
37. Louviere JJ, Islam T, Wasi N, Street D, Burgess L (2008). Designing Discrete Choice Experiments: Do Optimal Designs Come at a Price? J Consum Res 35(2):360–375.
38. Severin VC, Burgess L, Louviere J (2004). Comparing statistical efficiency and respondent efficiency in choice experiments. Sydney, Australia: Research report: Department of Mathematical Sciences, University of Technology.
39. Huston P, Rowan M (1998). Qualitative studies. Their role in medical research. Can Fam Physician 44:2453-2458.
40. Hak T, Van der Veer K, Jansen H (2004). The Three-Step Test-Interview (TSTI): An observational instrument for pretesting self-completion questionnaires. (No. ERIM Report ERS-2004-029-ORG.). Rotterdam: Erasmus Research Institute of Management.
41. Korfage IJ, Hak T, de Koning HJ, et al. (2006). Patients' perceptions of the side-effects of prostate cancer treatment--a qualitative interview study. Soc Sci Med 63(4):911-919.
42. Ryan M, Gerard K (2003). Using discrete choice experiments to value health care programmes: current practice and future research reflections. Appl Health Econ Health Pol 2:55-64.
43. Telser H, Zweifel P (2007). Validity of discrete-choice experiments evidence for health risk reduction. Appl Econ 39:69-78.

44. Adamowicz W, Louviere J, Williams M (1994). Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities. J Environ Econ Manage 26(3):271-292.

45. Carlsson F, Martinsson P (2001). Do Hypothetical and Actual Marginal Willingness to Pay Differ in Choice Experiments? Application to the Valuation of the Environment. J Environ Econ Manage 41(2):179-192.

46. Chang JB, Lusk JL, Norwood FB (2008). External Validity of Hypothetical Surveys and Laboratory Experiments. Selected Poster prepared for presentation at the American Agricultural Economics Association Annual Meeting, Orlando, FL, July 27-29, 2008.

47. Lusk JL, Schroeder TC (2004). Are Choice Experiments Incentive Compatible? A Test with Quality Differentiated Beef Steaks. Amer J Agr Econ 86(2):467–482.

48. Lusk JL, Pruitt JR, Norwood FB (2006). External validity of a framed field experiment. Economics Letters 93(2):285-290.

49. San Miguel F, Ryan M, Amaya-Amaya M (2004). Irrational stated preferences: a quantitative and qualitative investigation. Health Econ 14:307-322.

50. Sen A (1993). Internal consistency of choice. Econometrica 61: 495-521.

51. Slothuus SU, Gyrd-Hansen D (2003). Conjoint analysis. The cost variable: an Achilles' heel? Health Econ 12(6):479-491.

52. Ryan M, Watson V (2008). Comparing welfare estimates from payment card contingent valuation and discrete choice experiments. Health Econ [Epub ahead of print].

53. Pol van der M, Shiell A, Au F, Johnston D, Tough S (2008). Convergent validity between a discrete choice experiment and a direct open-ended method: Comparison of preferred attribute levels and willingness to pay estimates. Soc Sci Med 67:2043-2050.

54. Kjaer T, Bech M, Gyrd-Hansen D, Hart-Hansen K (2006). Ordering effect and price sensitivity in discrete choice experiments: need we worry? Health Econ 15(11):1217-1228.

Appendices

Summary

Health economics is concerned with issues related to scarcity in the allocation of health care. The basic tasks of any economic evaluation are to identify, measure, value, and compare the costs and consequences of alternatives being considered. Traditional means of measuring benefits in the delivery of health care have concentrated on improvements in health outcomes using clinical outcomes and Quality Adjusted Life Years (QALY). However, additional to health outcomes valuing process outcomes (e.g. treatment location, route of drug administration) and non-health outcomes (e.g. amount of information) may be worthwhile. These might be relevant for individuals' preferences and acceptability for specific health care interventions or programmes (i.e. demand-led health care), and for some interventions that do not provide reduction in morbidity or mortality (e.g. cosmetic surgery).

The discrete choice experiment (DCE) approach provides opportunities for evaluation of process effects and non-health outcomes additional to traditional QALY analysis. The technique of DCE is an attribute-based measure of benefit, based on the assumptions that, first, alternatives (goods or services) can be described by their characteristics, known as attributes, and second, an individual's valuation (i.e., benefit, utility, satisfaction or preference) depends upon the levels of these attributes. Within a DCE individuals are offered a series of choice sets, and are asked to choose in each choice set between two or more alternatives. The choice observed is assumed to reveal an underlying (latent) utility function. The DCE approach combines random utility theory with consumer theory, experimental design theory and econometric analysis.

In comparison to other stated preferences techniques that require the individual to rank or rate alternatives, a DCE presents a reasonably straightforward task and one which more closely resembles a real world decision (i.e., trading off health outcomes, process attributes and/or non-health outcomes). DCEs have been used widely in marketing, transport economics, and environmental economics. The application of DCE to the field of health care seems promising.

This thesis addresses the following objectives:

1. To provide an overview of current DCE practice in health economics, and to compare and assess progress in methodological advances. (Chapter 2)
2. To study the usefulness of DCE for real choice problems in health care for measuring benefits beyond health outcomes. (Chapters 3-8)
3. To compare labelled versus unlabelled DCEs in health care in various aspects of feasibility, trading-behaviour, and convergent validity. (Chapter 9)

Review of discrete choice experiments in health care

Chapter 2 provides an overview of DCEs applied in health economics between 2001 and 2008, and builds on a review of published papers between 1990-2000. This previous

review recommended to follow methodological developments taking place in other disciplines or areas of economics. Therefore, in our review consideration is given to three methodological issues ('experimental design'; 'methods of analysis'; and 'validity') to investigate if accumulating technical knowledge has been (more) often used in current DCEs.
We identified 114 DCEs published between 2001 and 2008. Our review showed that, whilst the technique was introduced into health economics to value what might be called patient experience factors, the applications today are much broader. These included valuing patient experience factors, health outcomes, investigating trade-offs between health outcomes and patient experience factors, estimating utility weights within the QALY framework, investigating labour-market choices amongst health care professional, priority setting frameworks at the local/national level and preferences regarding clinical decision making.
Regarding the technical quality of DCEs, there has been a shift towards more using statistically efficient designs and richer econometric models. Internal validity tests were largely incorporated in the DCE practice. However, much progress has still to be made towards external validity; i.e., testing whether individuals behave in reality as they stated in the survey. Also consideration needs to be given to incorporation of the results of a DCE into a decision-making frame-work by policy makers.

Practical applications of discrete choice experiments in health care

This thesis describes four applications of DCEs developed for real choice problems in health care to measure benefits beyond health outcomes. These four problems illustrate the potential of DCE to include non-health outcomes and process effects of health care interventions in preference elicitation.

The first DCE focused on preferences for preventive osteoporosis drug treatment. Various practice guidelines recommend a case-finding approach to identify persons with a high risk of osteoporotic fractures. However, the usefulness of this approach depends on whether the identified persons are willing to take preventive osteoporosis drug treatment, and on what conditions.
Chapter 3 showed that treatment attributes as effectiveness, side effects (nausea), total treatment duration, route of drug administration, and out-of-pocket costs all proved to be important for women's choices. A reduction of the relative 10-year risk of hip fracture by 40% or more by the drug was considered to compensate for nausea as a side effect. Women were prepared to pay an out-of-pocket contribution for the currently available drug treatment (bisphosphonate) if the fracture risk reduction was at least 12%.
Differences in opinions on the desirability of treatment initiation may hamper the process and outcome of shared decision-making between physician and patient. Therefore, we compared general practitioners' and patients' preferences for preventive osteoporosis

drug treatment, and evaluated the determinants of the preference differences found **(Chapter 4)**.
We showed that general practitioners had a significantly less favourable attitude towards preventive osteoporosis drug treatment than patients; they placed significantly higher values on effectiveness of drug treatment and short total treatment duration than patients.

The second DCE focused on preferences for idiopathic scoliosis brace treatment. The effectiveness of brace treatment in idiopathic scoliosis patients has not been established in randomized controlled trials (RCTs). Insight into patients' preferences for (characteristics of) brace treatment will be useful for future trials and for the development of braces that may optimize compliance with brace treatment.
We showed that treatment attributes as effectiveness (i.e. risk reduction of a surgical intervention), visibility of the brace, uncomfortable wearing of the brace, and treatment duration, all proved to be important for patients' choices (**Chapter 5)**. Effectiveness and discomfort in wearing a brace played the most important role in their choices. Patients were prepared to undergo treatment with a Boston brace for three years, if the brace would reduce the need for surgery by 32-74%.
These results are important if RCTs would conclusively establish that bracing is effective, and show directions for the further technical development of braces to increase the compliance with brace treatment.

Chapter 6 described the third DCE, which focused on (determinants of) women's preferences for three breast reconstruction modalities. Patients' preferences are important determinants in the decision for a specific type of breast reconstruction. Understanding women's motivational factors can contribute to further improve patient information and to develop demand-led healthcare. Breast reconstructions were characterized by six treatment attributes: 1) material used, 2) number and duration of operations, 3) short- and 4) long-term complication rate, 5) aesthetic result, and 6) waiting time. All these treatment attributes proved important for women's choices. Our results show that (autologous) material and aesthetic result are the most important determinants for women's choice for breast reconstruction. Autologous free flap breast reconstruction fitted in best with women's preferences.

The fourth DCE focused on preferences for colorectal cancer (CRC) screening programmes. Screening can reduce CRC mortality by early detection of CRC and endoscopic removal of premalignant precursors of CRC (adenomas). Uptake of CRC screening has remained low. Therefore, we determined the influence of different screening tests and their characteristics on individual preferences for CRC screening.
Chapter 7 presents the results of a DCE in which the screening strategies were described by various test types (faecal occult blood test (FOBT), flexible sigmoidoscopy, and colonoscopy) accompanied by realistic screening intervals and CRC mortality reductions.
Respondents preferred screening over no screening irrespective of the screening test. Screening test, interval and the risk on CRC mortality influenced subject's preferences for

CRC screening. Respondents were willing to sacrify 25% risk reduction on CRC mortality to obtain a five-yearly FS instead of a ten-yearly colonoscopy. Subjects in the target population preferred endoscopy screening to FOBT screening if realistic levels of the test characteristics were applied.

Chapter 8 presents the results of a DCE in which the screening strategies were described by treatment characteristics as pain, risk of complications, screening location, preparation, duration of procedure, screening interval and risk reduction of CRC related death. All aspects significantly influenced the respondents' preferences. Respondents required an additional relative risk reduction of CRC related death of 32% to utilise an extensive bowel preparation instead of no bowel preparation. Screening intervals shorter than 10 years were significantly preferred to a 10-year screening interval. Especially type of bowel preparation, length of screening interval and mortality reduction influenced individuals' trade-offs. These results provide insight in the decision-making process regarding the decision to participate in a CRC screening programme, thereby contributing to the improvement of information provided to CRC screening invitees, and they identify targets for increasing participation rates.

Theoretical aspects of discrete choice experiments in health care

A fundamental question that arises in the application of DCE is whether to present the choice sets in a labelled or unlabelled form. The unlabelled form involves assigning unlabelled alternatives in the choice set, such as 'alternative A', 'alternative B' and so on. The labelled form involves assigning labels that communicate information regarding the alternative. In marketing applications, labels tend to consist of brand names and logos, which consumers have learnt to associate with different product characteristics and feelings. In the context of health economics, labels tend to consist of generic or brand-name medications, specific screening tests (e.g. colonoscopy, sigmoidoscopy), specific treatments (surgery vs conservative), or other descriptors. An advantage of assigning labels is that alternatives will be more realistic and the choice task less abstract, which adds to the validity of the results (i.e., responses reflect better the real preference structure) and hence, the results may be better suitable to support decision-making at policy level. However, by far most commonly applied DCEs in health economics used unlabelled alternatives.

Chapter 9 empirically compared a labelled and an unlabelled DCE for CRC screening programs. The labelled version presented screening test alternatives as 'FOBT', 'sigmoidoscopy', and 'colonoscopy', whereas the unlabelled version presented them as 'test A' and 'test B'. The labels played a significant role in individual choices; approximately 22% of subjects had dominant preferences for screening test labels. The convergent validity was modest to low (participants in CRC screening: $r=0.54$; $p=0.01$; non-participants: $r=0.17$; $p=0.45$).

Conclusion

In **Chapter 10** the findings of this thesis are summarised and discussed. All DCE studies in this thesis gave insight into whether or not the given (non-)health attributes is important (according to a significant statistical test), the relative importance of the attributes, the rate at which individuals were willing to trade between attributes, and overall benefit scores for intervention alternatives. However, it is questionable whether the results really present serious trade-offs that agree with their actual choice. As a result, we stress that DCE outcomes may be more appropriate to show the relative importance of the attributes of a health care intervention, than to predict real choice.

The possibility to estimate a WTP is an additional option of DCE. However, to our opinion this additional option is limited. As a result, we recommend to include a cost attribute only if this is relevant in health care, and to interpret the WTP values derived from a DCE relatively.

We showed that labelled DCE led to a higher response rate, especially for the group who had familiarity with the context. However, more respondents who were not familiar with the context failed the rationality test with the labelled design. The inclusion of labels appeared to play a significant role in individual choices, and increased non-trading behaviour. From our point of view therefore unlabelled DCEs may be more suitable to investigate trade-offs between attributes and for respondents who are not familiar with the alternative labels, whereas labelled DCEs may be more suitable to explain real life choices such as uptake of cancer screening.

It may be worthwhile for the health economics community to link with areas outside health economics in terms of experimental design and estimation procedure. However, it should be realised that doing a lot of research to generate a methodologically perfect DCE is not useful if respondents can not understand the valuation task anymore or if it will be unpractical in a health care context.

Further research is needed to 1) define the place of DCEs versus other preference elicitation methods in health care; 2) generate DCEs which are as easy as possible for respondents, but still provide adequate answers to the research question; 3) optimise the design and analysis of DCEs, including issues as the usefulness of non-orthogonal efficient designs and dealing with 'irrational' responses; 4) find opportunities to test the external validity of DCEs in health care.

We conclude that the DCEs conducted in this thesis provided valuable insights regarding benefit valuation beyond clinical outcomes of various health care interventions, and are therefore useful as additional information to clinical outcomes or QALYs in a medical decision-making process about health care interventions. The DCE approach is a feasible instrument for the measurement of preferences of adults for all kind of health care interventions under the condition that the choice problem can be broken down into discrete choice options characterised by limited number of attributes and levels. It is a useful tool in health care to directly compare the determinants of the preferences for

health care interventions from different groups, such as health professionals, patients, and general population.

Samenvatting

Gezondheidseconomie is een deelgebied van de economische wetenschap waarin onderzoek wordt gedaan naar schaarste en de toewijzing van gezondheidszorg. De belangrijke taken van iedere economische evaluatie zijn het identificeren, het meten, het waarderen en het vergelijken van de kosten en consequenties van verschillende alternatieven. In de gezondheidseconomie wordt traditiegetrouw vooral gekeken naar de gezondheidswinst die een interventie oplevert door gebruik te maken van klinische uitkomsten en het aantal zogenoemde QALYs (Quality Adjusted Life Years; een maat waarmee voor kwaliteit gecorrigeerde levensjaren uitgedrukt worden). Echter, het kan waardevol zijn naast de gemeten gezondheidswinst ook procesuitkomsten (bv. de locatie van de behandeling of de wijze van medicatietoediening) en niet-gezondheidsuitkomsten (bv. de hoeveelheid verkregen informatie) te meten. Deze uitkomsten kunnen relevant zijn voor:

1. de preferenties van individuen en hun bereidheid om specifieke zorginterventies of programma's te aanvaarden (de zogenoemde vraaggestuurde zorg) en
2. voor medische interventies, die geen gezondheidswinst opleveren, maar wel geprefereerd worden (e.g. plastische chirurgie).

Het discrete keuze experiment (DCE) kan proceseffecten en niet-gezondheidsuitkomsten evalueren naast de traditionele QALY analyse. De DCE techniek meet het nut van alternatieven (goederen of diensten) onder de aannames dat:

1. de alternatieven beschreven kunnen worden door hun karakteristieken (ook wel attributen genoemd), en
2. dat de waardering van een individu (ofwel het nut, utiliteit, tevredenheid of preferentie) afhangt van de niveaus (levels) van die karakteristieken.

In een discrete keuze experiment krijgen individuen een aantal keuzesets gepresenteerd. In elke keuzeset van twee of meer alternatieven wordt de individuen gevraagd een keuze te maken. Er wordt aangenomen dat de geobserveerde keuze een onderliggende (verborgen) preferentiefunctie reflecteert. De DCE techniek combineert nutstheorie met consumententheorie, experimenteel ontwerptheorie en econometrische analyse.
In vergelijking met andere preferentietechnieken die het individu vraagt om alternatieven een rangordening te geven of te waarderen, presenteert een discrete keuze experiment een redelijke recht-toe-recht-aan taak. De DCE-taak komt vrijwel overeen met een beslissing in de werkelijke praktijk; het gaat om het afwegen van gezondheidsuitkomsten, proceseffecten en/of niet-gezondheidsuitkomsten. DCEs zijn al veel toegepast in de marketing, in de transporteconomie en de milieueconomie. De toepassing van DCE op het gebied van gezondheidszorg lijkt veelbelovend.

Dit proefschrift heeft drie doelen:

1. Een overzicht geven van huidige toegepaste discrete keuze experimenten (DCEs) in de gezondheidszorg en de vooruitgang beoordelen in de gebruikte werkwijze. (Hoofdstuk 2)
2. Het nut van DCEs bestuderen voor enkele realistische keuzeproblemen in de gezondheidszorg. (Hoofdstukken 3-8)
3. Een vergelijking maken van een DCE met gelabelde alternatieven met een DCE waarin de alternatieven een algemene benaming hebben op aspecten als uitvoerbaarheid, mate van afweging en convergente validiteit. (Hoofdstuk 9).

Overzicht van discrete keuze experimenten in de gezondheidszorg

Hoofdstuk 2 geeft een overzicht van toegepaste DCEs in de gezondheidseconomie tussen 2001 en 2008. Deze sluit aan op een overzichtsstudie van gepubliceerde DCE-artikelen tussen 1990-2000, waarin werd aanbevolen om methodologische ontwikkelingen op andere economische terreinen te volgen. Daarom is in onze overzichtsstudie aandacht gegeven aan drie methodologische aspecten: 1) het ontwerp van een discrete keuze experiment, 2) de analysemethoden, en 3) de validiteit (validiteit is de mate waarin een meting, bewering of uitspraak juist is).
Wij identificeerden 114 gepubliceerde DCEs tussen 2001 en 2008. Onze overzichtsstudie liet zien dat DCEs werden toegepast voor uiteenlopende doeleinden zoals: om gezondheidsuitkomsten te waarderen, om afwegingen tussen gezondheidsuitkomsten en/of procesuitkomsten en/of niet-gezondheidsuitkomsten in kaart te brengen, om gewichten te schatten binnen een QALY (kwaliteit gecorrigeerde levensjaar), en om te achterhalen welke behandelingen medisch personeel prefereert voor hun patiënten.
Kijkend naar de technische kwaliteit van DCEs kunnen we constateren dat er een verschuiving is geweest; in de huidige discrete keuze experimenten werden meer statistisch efficiënte ontwerpen en vooruitstrevende econometrische modellen toegepast. Het toetsen van interne validiteit was grotendeels geïntegreerd in de DCE praktijk. Echter, bij geen enkele toegepaste DCE was de externe validiteit getoetst; ofwel toetsen of individuen zich werkelijk gedragen als ze zeiden dat ze zouden doen in het onderzoek.

Praktische toepassingen van discrete keuze experimenten in de gezondheidszorg

Dit proefschrift beschrijft vier DCEs, die ontwikkeld zijn voor realistische keuzeproblemen in de gezondheidszorg waarin (ook) effecten anders dan gezondheids-winsten in acht werden genomen. Deze vier problemen illustreren de mogelijkheid van DCEs om proceseffecten en niet-gezondheidsuitkomsten van medische interventies mee te nemen bij het bepalen van voorkeuren voor die interventies.

De eerste DCE werd uitgevoerd om preferenties van patiënten en artsen te achterhalen voor medicatie om osteoporose (botontkalking) te voorkomen. In verschillende richtlijnen wordt aanbevolen om mensen met een verhoogde kans op osteoporotische facturen (meestal vrouwen) op te sporen door middel van een zogenoemde 'case-finding' benadering (case-finding is het zoeken door de hulpverlener of arts naar risicofactoren of beginnende afwijkingen bij mensen die om andere redenen een hulpverlener of arts bezoeken; het doel is vroegtijdig kunnen behandelen of erger voorkomen). Deze benadering is alleen succesvol als de opgespoorde personen ook bereid zijn om medicatie te gebruiken om osteoporose te voorkomen en op welke voorwaarden.
Hoofdstuk 3 liet zien dat behandelingskarakteristieken als effectiviteit, bijwerkingen (misselijkheid), totale behandelingsduur, de wijze van medicatietoediening en de eigen financiële bijdrage een relevante invloed hadden op de keuzen. Misselijkheid als bijwerking van de medicatie werd geaccepteerd als de preventieve medicatie ervoor zorgde dat het relatieve 10-jaars risico op een heupfractuur met 40% gereduceerd werd. Vrouwen waren bereid een eigen bijdrage te betalen voor de huidige beschikbare preventieve medicatie (bisfosfonaat) als de kans op een fractuur met minstens 12% afnam.
Meningsverschillen over het beginnen met een (preventieve) behandeling kan het nemen van een gezamenlijke beslissing van de arts en de patiënt in de weg staan. Daarom vergeleken wij de voorkeuren van huisartsen en patiënten met elkaar wat betreft preventieve medicatie voor osteoporose en evalueerden de determinanten (factoren die een toestand of een ontwikkeling (mede)bepalen) van de gevonden preferentieverschillen **(Hoofdstuk 4)**.
Wij toonden aan dat huisartsen een duidelijk minder positieve houding hadden tegenover preventieve medicatie dan patiënten; huisartsen hechtten meer waarde aan de effectiviteit van de medicatie en aan een korte totale behandelingsduur dan patiënten.

De tweede DCE **(Hoofdstuk 5)** werd uitgevoerd om voorkeuren voor brace-behandeling bij idiopatische scoliose (zijwaartse verkromming van de wervelkolom zonder duidelijke oorzaak) te onderzoeken. De effectiviteit van een brace-behandeling bij idiopatische scoliose patiënten is nog niet vastgesteld.
Wij toonden aan dat behandelingskenmerken als effectiviteit (ofwel risico reductie op een chirurgische interventie), zichtbaarheid van de brace, het draagcomfort van de brace en de behandelingsduur allen belangrijk waren voor de keuzen van de patiënt. Effectiviteit en het draagcomfort van de brace waren het belangrijkst. Patiënten waren bereid een brace-behandeling van drie jaar te ondergaan met een Boston brace (een korset gemaakt van lichtgewicht plastic, dat ontwikkeld is in de Amerikaanse stad Boston), als dat de kans op een operatie met één tot tweederde zou reduceren.
Deze resultaten zijn belangrijk 1) als gerandomiseerde gecontroleerde onderzoeken (effectonderzoeken waarbij twee groepen met elkaar worden vergeleken die door toeval zijn samengesteld) zouden aantonen dat brace-behandeling effectief is en 2) voor verdere

technische brace-ontwikkeling om het correct opvolgen van een brace-behandeling door de patiënt te verhogen.

Hoofdstuk 6 beschreef een derde DCE, die de (determinanten van) voorkeuren van vrouwen voor drie manieren van borstreconstructie onderzocht. De preferenties van vrouwen zijn belangrijke factoren in de beslissing voor een specifiek type borstreconstructie. Het begrijpen van de motivaties van vrouwen kan bijdragen aan het verder verbeteren van informatie voor de patiënt en om vraaggestuurde zorg te ontwikkelen. De borstreconstructies werden gekarakteriseerd door zes behandelingskenmerken: 1) gebruikte materiaal, 2) aantal operaties en operatieduur, 3) korte- en 4) lange termijn complicatiekansen, 5) esthetisch resultaat, and 6) wachttijd. Al deze behandelingskenmerken waren belangrijk voor de keuzen van vrouwen. Onze resultaten toonden aan dat (lichaamseigen weefsel) materiaal en esthetisch resultaat de belangrijkste factoren waren voor de keuze van vrouwen om een borstreconstructie te ondergaan. De huidige borstreconstructie met eigen weefsel kwam het best overeen met de voorkeuren van vrouwen.

De vierde DCE **(Hoofdstukken 7 en 8)** onderzocht de voorkeuren voor dikke darmkanker screening. Screening (bevolkingsonderzoek) kan naar verwachting sterfte door dikke darmkanker reduceren door dikke darmkanker in een vroeg stadium op te sporen en kwaadaardige voorlopers van dikke darmkanker (adenomen) endoscopisch (ofwel met een inwendig kijkonderzoek) te verwijderen. Het deelnamepercentage aan dikke darmkanker screening is lager dan aan borstkanker screening. Daarom onderzochten we de invloed van (kenmerken) van verschillende screeningstesten op de voorkeuren van individuen voor dikke darmkanker screening.
Hoofdstuk 7 geeft de resultaten weer van een DCE waarin de screeningsstrategieën werden beschreven door verschillende type testen (ontlastingstest, flexibele sigmoidoscopie (inwendig kijkonderzoek van het laatste deel van de dikke darm) en colonoscopie (inwendig kijkonderzoek van de gehele dikke darm)) vergezeld door realistische sterftereducties en screeningsintervallen (screeningstussenpozen; hoe vaak een individu in 10 jaar zou moeten komen voor het bevolkingsonderzoek).
De respondenten prefereerden screening boven geen screening ongeacht welke screeningstest gebruikt werd. Zowel de screeningstest, het screeningsinterval als het risico op sterfte door dikke darmkanker beïnvloedden de voorkeuren van individuen voor dikke darmkanker screening. De respondenten leken bereid 25 procent risicoreductie op sterfte door dikke darmkanker op te offeren als ze daarvoor eens per vijf jaar een flexibele sigmoidoscopie kregen in plaats van eens per tien jaar een colonoscopie. De doelgroep prefereerde screening met een endoscoop boven screening met een ontlastingstest als realistische niveaus van testkenmerken en uitkomsten golden.
Hoofdstuk 8 geeft de resultaten weer van een DCE waarin de screeningsstrategieën beschreven werden door testkenmerken als pijn, complicatierisico, screeningslocatie,

voorbereiding, screeningsduur, screeningsinterval en risicoreductie op sterfte door dikke darmkanker screening.
Alle testkenmerken hadden een significante invloed op de preferenties van respondenten. Respondenten eisten een 32 procent additionele relatieve risicoreductie op mortaliteit door dikke darmkanker om een intensieve dikke darm voorbereiding te ondergaan in tegenstelling tot geen dikke darm voorbereiding. Screeningsintervallen korter dan tien jaar werden geprefereerd boven een screeningsinterval van eens per tien jaar. Vooral het type dikke darm voorbereiding, de tijdsinterval van screening en de mortaliteitreductie beïnvloedden de afwegingen van individuen. Deze resultaten leveren inzichten in het besliskundig proces omtrent de beslissing om aan dikke darmkanker screening deel te nemen en draagt bij om de informatie verstrekt aan de screeningsdoelgroep te verbeteren en om punten te identificeren om de deelname aan dikke darmkanker screening te verhogen.

Theoretische aspecten van discrete keuze experimenten in de gezondheidszorg.

Een belangrijke vraag in het uitvoeren van een discrete keuze experiment (DCE) is of de DCE gepresenteerd moet worden waarin de keuze opties (alternatieven) een specifiek label hebben of niet, respectievelijk een gelabelde DCE en ongelabelde DCE genoemd. Een ongelabelde DCE bevat keuze opties met een algemene benaming in een keuze set, zoals ‘ optie A’, ‘optie B’ enzovoort. Een gelabelde DCE bevat keuze opties met een specifiek etiket dat informatie geeft over die bepaalde keuze optie. Denk bijvoorbeeld in de marketing aan merken en logo’s, die voor consumenten bepaalde productkarakteristieken en gevoelens oproepen. Op gezondheidseconomisch gebied kunnen we denken aan de naam van een geneesmiddel, een specifieke screeningstest (colonoscopie, sigmoidoscopie), een specifieke behandeling (operatie, bestraling) enzovoort. Een voordeel van het toewijzen van specifieke benamingen aan de keuzeopties is dat de opties realistischer worden en de keuzetaak minder abstract, wat zou kunnen bijdragen aan de validiteit van de resultaten en dus geschikter is om beslissingen op beleidsniveau te ondersteunen. Echter, de meeste toegepaste discrete keuze experimenten in de gezondheidszorg gebruikten keuze opties met algemene (ofwel ongelabelde) benamingen.
In **Hoofdstuk 9** werden een gelabelde en een ongelabelde DCE op het gebied van dikke darmkanker screening empirisch vergeleken. De ongelabelde DCE presenteerde de screeningstest opties als ‘ontlastingstest’, ‘sigmoidoscopie’ en ‘colonoscopie’, terwijl de ongelabelde DCE de screeningstesten presenteerden als ‘screeningstest A’ en ‘ screeningstest B’. De specifieke labels van de keuze opties speelden een duidelijke rol in de keuzen van individuen; ongeveer 22 procent van de respondenten had een dominante voorkeur voor het specifieke label van een screeningstest. De uitkomsten van de gelabelde en ongelabelde DCE kwamen gemiddeld tot slecht overeen.

Conclusie

In **Hoofdstuk 10** worden de bevindingen van dit proefschrift samengevat en besproken. Alle DCE studies in dit proefschrift gaven inzicht of de gegeven (niet)gezondheidskenmerken belangrijk waren (volgens een significante statistische test) voor de keuzen, in het relatieve belang van de kenmerken, in de mate waarin individuen bereid waren kenmerken met elkaar af te wegen en in de totale score van tevredenheid voor een interventie. Het is echter twijfelachtig of de resultaten werkelijk afwegingen representeren die overeenkomen met de uiteindelijke keuze. Wij stellen dat DCE uitkomsten geschikter lijken te zijn om de relatieve invloed van de kenmerken van gezondheidsinterventies op het keuzegedrag weer te geven dan om werkelijk gedrag te voorspellen.
De mogelijkheid om een betalingsbereidheid te schatten is een extra voordeel van discrete keuze experimenten, maar naar ons inziens is dit voordeel beperkt. Wij bevelen aan om alleen een kostenkenmerk in een DCE mee te nemen als dit daadwerkelijk relevant is. Tevens bevelen wij aan om de geschatte betalingsbereidheid voor een medische interventie relatief te interpreteren en niet absoluut.
Wij lieten zien dat een gelabelde DCE tot een hogere respons leidde, vooral voor de groep die bekend was of ervaring had met de gelabelde opties. Echter meer individuen die niet bekend waren met of geen ervaring hadden met de gelabelde opties zakten voor de rationaliteitstest (een test om te kijken of de individu een rationele keuze maakt) in het gelabelde DCE design. Het toevoegen van specifieke labels voor de keuze opties in een DCE bleek een belangrijke rol te spelen in de keuzen van individuen, als gevolg dat individuen minder bereidheid waren om verschillende kenmerken van een medische interventie tegen elkaar af te wegen. Daarom zijn vanuit ons oogpunt DCEs met ongelabelde keuzeopties geschikter om de afwegingen tussen kenmerken van medische interventies te onderzoeken en voor individuen waarvoor de gelabelde keuzeopties geen betekenis hebben, terwijl DCEs met gelabelde keuze opties geschikter zijn om werkelijk gedrag te verklaren zoals deelname aan kankerscreening.
Voor de gezondheidseconomische onderzoeksgemeenschap kan het waardevol zijn om aansluiting te zoeken met andere gebieden buiten de gezondheidseconomie als het gaat om het ontwerp van een DCE en om de analysemethoden. Echter veel onderzoek doen om een methodologische perfecte DCE te genereren is weinig zinvol als de respondenten daardoor de waarderingstaak niet meer begrijpen.

Verder onderzoek is nodig om:

1. de plaats van DCE te definiëren ten opzichte van andere methoden om preferenties te onderzoeken;
2. discrete keuze experimenten te genereren die zo makkelijk mogelijk zijn voor respondenten, maar toch adequate antwoorden leveren op de onderzoeksvraag;
3. het design en de analyse van DCEs te optimaliseren;
4. de externe validiteit van DCE in de gezondheidszorg te testen.

Wij concludeerden dat de discrete keuze experimenten (DCEs) uitgevoerd in dit proefschrift waardevolle inzichten opleverden bij de waardering van gezondheidsinterventies door rekening te houden met meer dan alleen klinische uitkomsten. Hierdoor zijn de resultaten aanvullend op de klinische uitkomsten of QALYs (kwaliteit gecorrigeerde levensjaren) in een besluitvormingsproces over gezondheidsinterventies. De DCE is een hanteerbaar instrument om preferenties van volwassenen te meten voor alle vormen van gezondheidsinterventies onder de voorwaarde dat het keuzeprobleem opgedeeld kan worden in discrete keuze opties die gekarakteriseerd worden door een aantal kenmerken en kenmerkniveaus. De DCE is een nuttig instrument in de gezondheidszorg om de factoren die van invloed zijn op de preferenties van verschillende groepen, zoals medische beroepsbeoefenaren, patiënten en algemene bevolking, voor gezondheidsinterventies direct met elkaar te vergelijken.

Co-authors

Name	Affiliation
E.M. Bunge, PhD	Department of Public Health, Erasmus MC, Rotterdam
F.C. van Biezen, MD	Department of Orthopaedics, Erasmus MC, Rotterdam
L. van Dam, MSc	Department of Gastroenterology and Hepatology, Erasmus MC, Rotterdam
T.H.C. Damen, MD	Department of Plastic and Reconstructive Surgery, Erasmus MC, Rotterdam
B. Donkers, PhD	Department of Business Economics, Erasmus University, Rotterdam
M.L. Essink-Bot, MD, PhD	Department of Social Medicine, Academic Medical Centre, University of Amsterdam, Amsterdam; Department of Public Health, Erasmus MC, Rotterdam
K.M. Gerard, PhD	School of Health Sciences, University of Southampton, United Kingdom
J.D.F. Habbema, PhD	Department of Public Health, Erasmus MC, Rotterdam
S.O.P. Hofer, MD, PhD, FRCS(C)	Department of Plastic and Reconstructive Surgery, Erasmus MC, Rotterdam; Division of Plastic Surgery, University Health Network, University of Toronto, Toronto, Ontario, Canada
L. Hol, MD	Department of Gastroenterology and Hepatology, Erasmus MC, Rotterdam
B.W. Koes, PhD	Department of General Practice, Erasmus MC, Rotterdam

H.J. de Koning, PhD	Department of Public Health, Erasmus MC, Rotterdam
E.J. Kuipers, MD, PhD	Department of Gastroenterology and Hepatology, Erasmus MC, Rotterdam; Department of Internal Medicine, Erasmus MC, Rotterdam
M.E. van Leerdam, MD, PhD	Department of Gastroenterology and Hepatology, Erasmus MC, Rotterdam
W.J. Meerding, PhD	Department of Public Health, Erasmus MC, Rotterdam
M.B. Menke-Pluijmers, MD, PhD	Erasmus University Medical Centre/ Daniel den Hoed Cancer Centre, Family Cancer Clinic, Department of Surgical Oncology, Rotterdam
M.A.M. Mureau, MD, PhD	Department of Plastic and Reconstructive Surgery, Erasmus MC, Rotterdam
H.A.P. Pols, MD, PhD	Department of Internal Medicine, Erasmus MC, Rotterdam
M. Ryan, PhD, FRSE	Health Economics Research Unit, University of Aberdeen, United Kingdom
C. Seynaeve, MD, PhD	Erasmus University Medical Centre/ Daniel den Hoed Cancer Centre, Family Cancer Clinic, Department of Medical Oncology, Rotterdam
E.W. Steyerberg, PhD	Department of Public Health, Erasmus MC, Rotterdam

Dankwoord

Met veel plezier heb ik de afgelopen vier jaar aan dit proefschrift gewerkt. Bij de totstandkoming van dit proefschrift zijn velen betrokken geweest. Ik wil deze ruimte dan ook benutten om allen hiervoor te bedanken, enkelen van hen in het bijzonder.

Allereerst gaat mijn dank uit naar mijn promotor en co-promotor: Ewout en Marie-Louise. Gedurende de jaren heb ik veel van jullie mogen leren. Ik bewonder jullie voor jullie kennis, opbouwende en snelle commentaar, creatieve denken, en bovenal jullie coachende werkwijze. Dat ik jullie in één adem noem is bewust. Jullie samenwerking is van bijzondere klasse: efficiënt en doelgericht, gecombineerd met de nodige humor.
Ewout, jouw laagdrempeligheid en openheid zijn karaktereigenschappen die ik waardeer. Ik wil je bedanken voor de kansen die jij mij gegeven hebt en voor je inzet om mijn doel te verwezenlijken. Marie-Louise, jouw input op verschillende vlakken is voor mij van onschatbare waarde geweest. Het vertrouwen dat je in mij hebt, heeft mij de vrijheid gegeven mijzelf te kunnen ontplooien.

Mijn dank gaat ook uit naar alle co-auteurs, met in het bijzonder Eveline, Tim, Lieke en Leonie. Dank voor jullie bijdrage en voor het delen van jullie klinische kennis. I also want to give special thanks to Mandy and Karen; writing a review paper is more challenging than writing up an empirical study. Thank you for teaching these important academic skills. En natuurlijk Bas; mijn helpdesk op DCE-gebied. Dank voor het delen van je kennis en praktische tips.

Dit brengt mij bij de leden van de kleine commissie. Professor Brouwer, professor Hunink en professor Kuipers, bedankt voor het lezen en beoordelen van mijn proefschrift.

Ook al mijn (oud) MGZ-collega's wil ik hierbij bedanken voor de gezellige tijd en inspirerende werkomgeving. Tinneke, Eveline, Elin, Mirjam, Hester, Esther, Judith en Hilde, dank voor alle leuke (niet-)werkgerelateerde gesprekken. Kees en Sonja, bedankt voor jullie computer- en administratieve ondersteuning. Ida, Robine, Inge, Marjolein, Saskia en René, dank voor de prettige samenwerking aan projecten buiten dit proefschrift om. Dit geldt ook voor mijn RIVM-collega's, met in het bijzonder Lany, Johan, Nikkie en Matthijs.

Dan natuurlijk mijn paranimfen, Suzanne en Iris. Ik vind het een eer dat jullie aan mijn zijde willen staan tijdens de promotieplechtigheid. Lieve Suzanne, je was een geweldige kamergenote de afgelopen 4 jaar. We hebben de nodige serieuze gesprekken gevoerd, maar bovenal veel gelachen. Lieve Iris, wat ben ik blij dat jij in Bleiswijk bent komen wonen. Ik heb genoten van onze gezamenlijke fietstochtjes naar huis waar menig onderwerp besproken werd.

Mijn proefschrift had niet nu al af kunnen zijn zonder de leidsters en directie van Kinderdagverblijf de Petteflat te Bergschenhoek. Lieve Alke, Leonie, Esther, Esmeralda,

Jessica, Marsha, Ellen, Lieke, Jamila, Belinda, Claudia, Anja, en alle anderen, jullie enthousiasme en verantwoordelijkheid hebben enorm bijgedragen aan dit proefschrift.

Beste vrienden, hoewel ik probeerde om het proefschriftonderwerp te vermijden, bleven jullie interesse tonen. Bedankt voor jullie afleiding en betrokkenheid. Liebi Fründi us dr Schwiz. I ha öies schöne Land verla, um die Doktrarbeit z'mache. Obwou i froh bi, dä Schritt gemacht dsha, ha ig nie vergässe wie schöns bi öich gsi isch. Merci viumau.

En dan natuurlijk mijn familie. Bea, Sjax, Grete, Bart, George, Hanneke, Rachèl, en Bas, dank voor de steeds weer getoonde interesse. Lieve Kiki, Dominique, Collin en Mats, bedankt voor jullie hartverwarmende afleiding.

Lieve pa en ma, bedankt voor jullie onvoorwaardelijke vertrouwen en steun. Altijd waren jullie er voor mij wanneer dat nodig was. Bedankt voor alle mogelijkheden die jullie mij gegeven hebben om datgene te doen wat ik ambieer.

Lieve Chayenne en Kenai, jullie zijn mijn drive geweest om het proefschrift zo snel mogelijk af te ronden. Jullie waren echte slaapkopjes wat een grote bijdrage is geweest aan de voortgang van mijn proefschrift. Mijn leven is zoveel mooier met jullie om mij heen.
Lieve Lars, je hebt altijd voor honderd procent achter dit proefschrift gestaan. Stilzwijgend luisterde je naar al mijn ups en downs. Dank voor je onvoorwaardelijke steun. Zonder zo'n achterban was deze marathon een stuk vermoeiender verlopen. En ja, ook hier geldt: alles komt terecht.

Esther

PhD portfolio summary

Summary of PhD training and teaching activities

Name PhD student: Esther de Bekker-Grob
Erasmus MC Department: Public Health

PhD period 2005-2009
Promotor: Prof. dr. E.W. Steyerberg
Supervisor: Dr. M.L. Essink-Bot

PhD Training	Year	Workload
Research skills		
Erasmus Summer Programme, Erasmus MC Rotterdam		
• Biostatistics	2005	20 hours
Institute of Medical Technology Assessment, Rotterdam		
• Basic principles of discrete choice experiments	2005	20 hours
Centre for microdata methods and practice (CEMMAP) – Institute for Fiscal Studies (IFS) – University College London		
• Discrete choice modelling	2006	40 hours
Nihes, Erasmus MC Rotterdam		
• Planning and evaluation of screening	2006	40 hours
Institute of Medical Technology Assessment, Rotterdam		
• Designing and modelling discrete choice experiments	2007	32 hours
University of Glasgow, Public Health and Health Policy Section, Division of Community Based Science		
• Advanced modelling methods for health economic evaluation	2009	24 hours
Presentations		
Landelijk Forum Medische Besliskunde, Rotterdam		
• Patient's preferences for osteoporosis drug treatment: a discrete choice experiment	2007	40 hours
Seminar Public Health, Rotterdam		
• Patient's preferences for osteoporosis drug treatment: a discrete choice experiment	2007	20 hours
International Society for Pharmacoeconomics and Outcomes Research (ISPOR), Dublin, UK		
• Patient's preferences for osteoporosis drug treatment: a discrete choice experiment	2007	20 hours

	Year	Workload
Presentations (continued)		
Society of Medical Decision Making, Pittsburgh, US		
• GPs and patient's preferences for osteoporosis drug treatment: a discrete choice experiment	2007	40 hours
Conjoint analysis in health Care, Delray Beach, US		
• Labelled versus unlabelled DCEs in health care: an application to colorectal cancer screening	2009	40 hours
Seminars and workshops		
Attending seminars of the department of Public Health	2005-2009	80 hours
Workshops on designing DCEs, Lund, Sweden	2007	20 hours
Workshops on economic evaluation, Pittsburgh, US	2007	20 hours
Workshops on DCEs, Delray Beach, US	2007	8 hours
International conferences		
Society of Medical Decision Making, Pittsburgh, US	2007	24 hours
International Health Economics Association, Copenhagen, Denmark	2007	20 hours
Conjoint analysis in health care, Delray Beach, US	2009	24 hours

List of publications

by September 1, 2009

First authorships

BEKKER-GROB EW de, Aa MNM van der, Zwarthoff EC, Eijkemans MJC, Rhijn BW van, Kwast TH van der, Steyerberg EW. Non-muscle invasive bladder cancer surveillance in which cystoscopy is partly replaced by microsatellite analysis on urine: a cost-effective alternative? (CEFUB-trial). BJU International 2008; 104:41-47.

BEKKER-GROB EW de, Bunge EM, Biezen FC van, Essink-Bot ML, Koning HJ de. Patients' preferences for scoliosis brace treatment: a discrete choice experiment. Spine. I*n press.*

BEKKER-GROB EW de, Essink-Bot ML, Meerding WJ, Koes BW, Steyerberg EW. General practitioners' and patients' preferences for osteoporosis drug treatment: a discrete choice experiment. PharmacoEconomics 2009; 27(3):211-209.

BEKKER-GROB EW de, Essink-Bot ML, Meerding WJ, Pols HAP, Koes BW, Steyerberg EW. Patients' preferences for osteoporosis drug treatment: a discrete choice experiment. Osteoporosis International 2008; 19(7):1029-1037.

BEKKER-GROB EW de, Hol L, Donkers B, Dam L van, Habbema JDF, Leerdam ME van, Kuipers EJ, Essink-Bot ML, Steyerberg EW. Labelled versus unlabelled discrete choice experiments in health economics: an application to colorectal cancer screening. *Submitted.*

BEKKER-GROB EW de, Polder JJ, Mackenbach JP, Meerding WJ. Towards a comprehensive estimate of national spending on prevention. BMC Public Health 2007; 7:252.

BEKKER-GROB EW de, Polder JJ, Witte KE, Mackenbach JP, Meerding WJ. Kosten van preventie in Nederland 2003. Bilthoven, 2006: RIVM report 270751011

BEKKER-GROB EW de, Ryan M, Gerard KM, Steyerberg EW, Essink-Bot ML. Applying discrete choice experiments to value health and health care: a review of the literature. *Submitted.*

Co-authorships

Casteren NJ van, Jong J de, Stoop H, Steyerberg EW, BEKKER-GROB EW de, Dohle GR, Oosterhuis JW, Looijenga LH. Evaluation of testicular biopsies for carcinoma in situ: immunohistochemistry is mandatory. International Journal of Andrology. *In press.*

Dam L van, Hol L, BEKKER-GROB EW de, Steyerberg EW, Kuipers EJ, Habbema JDF, Essink-Bot ML, Leerdam ME van. What influences individuals' preferences for colorectal cancer screening tests? An application of the discrete choice experiment. Eur J Cancer. *In press.*

Damen THC, BEKKER-GROB EW de, Mureau MAM, Menke-Pluijmers MB, Seynaeve C, Hofer SOP, Essink-Bot ML. Patients' Preferences for Breast Reconstruction: a Discrete Choice Experiment. *Submitted.*

Hol L, BEKKER-GROB EW de, Dam L van, Donkers B, Kuipers EJ, Habbema JDF, Steyerberg EW, Leerdam ME van, Essink-Bot ML. Population preferences for different screening strategies for colorectal cancer in the Netherlands; a discrete choice experiment. *Submitted.*

Mello NM van, Mol F, Opmeer BC, BEKKER-GROB EW de, Essink-Bot ML, Ankum WM, Mol BW, Veen F van der, Hajenius PJ. Salpingostomy or salpingectomy in tubal ectopic pregnancy: what do women prefer? A discrete choice experiment. *Submitted.*

Curriculum vitae

Esther Wilhelmina Grob werd geboren op 15 december 1978 te Zevenaar. Ze behaalde in 1997 haar Gymnasium diploma aan het Liemers College te Zevenaar. Na driemaal te zijn uitgeloot voor de studie Geneeskunde rondde zij in 2001 de opleiding HBO - Medische Beeldvormende en Radiotherapeutische Technieken af aan de Hogeschool Haarlem. In het zelfde jaar voltooide zij ook de opleiding Stralingsdeskundigheid Niveau 3 aan het Interuniversitair Onderzoeksinstituut voor Radiopathologie en Stralingsbescherming te Leiden. Gedurende 3 jaar werkte zij als radiotherapeutisch laborant en opleidings-coördinator op de afdeling radiotherapie in het Inselspital te Bern, Zwitserland. In 2003 behaalde ze haar managementdiploma aan de Management-Fachschule te Bern. Haar doctoraal diploma Public Health – Health Policy, Economics, and Management behaalde zij in 2005 aan de Universiteit Maastricht. Eveneens in 2005 werd zij aangesteld als junior onderzoeker op de afdeling Maatschappelijke Gezondheidszorg van het Erasmus Medisch Centrum te Rotterdam. Hier is zij ook nu nog werkzaam. Ze was als gezondheidseconoom betrokken bij diverse kosten(effectiviteit)studies en preferentie-studies, waarover werd gerapporteerd op verschillende nationale en internationale congressen en in publicaties in medisch-wetenschappelijke tijdschriften. Tevens was zij als co-auteur betrokken bij de 'Volksgezondheid Toekomst Verkenning 2006' en bij het leerboek 'Van kosten tot effecten – een handleiding voor evaluatiestudies in de gezondheidszorg'.
Esther is getrouwd met Lars de Bekker en samen hebben zij een dochter Chayenne (2006) en een zoon Kenai (2008).