DISCRETE CHOICE EXPERIMENTS IN PUBLIC HEALTH

Jorien Veldwijk

Discrete Choice Experiments in Public Health

Thesis Utrecht University - with a summary in Dutch
Proefschrift Universiteit Utrecht - met een samenvatting in het Nederlands

ISBN: 978-90-393-6306-5
Author: Jorien Veldwijk
Cover illustration: Antwan van Delft
Lay-out: Concept Design
Printing: Gildeprint

Printing of this thesis was financially supported by the Julius Center for Health Sciences and Primary Care and The National Institute for Public Health and the Environment.

The research reported in this thesis was financially supported by the National Institute for Public Health and the Environment (RIVM, program 'From Knowledge to Action', grant number S/260216/01/FW).

DISCRETE CHOICE EXPERIMENTS IN PUBLIC HEALTH

Discrete Keuze Experimenten in de Volksgezondheid

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus,

prof. dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op

dinsdag 12 mei 2015 des middags te 12.45 uur

door

Jorien Veldwijk

geboren op 7 januari 1987 te Heino

Promotor: Prof. dr. H.A. Smit

Copromotoren: Dr. G.A. de Wit
Dr. M.S. Lambooij

CONTENTS

GENERAL INTRODUCTION

Public Health interventions and individuals' preferences

One approach to improve public health is to implement preventive programs that have been proven effective and cost-effective at the population level. For any preventive program to be successful, it is of paramount importance that a large majority of the target population participates; otherwise, the anticipated effectiveness and cost effectiveness of the program will be hampered and consequently public health might not improve. Unfortunately, it is not self-evident that high participation rates will be attained once preventive programs become available. Even additions to existing successful preventive programs do not automatically ensure high participation rates. For example, the Netherlands has an effective National Immunization Program (NIP), through which about 96% of all children in the Netherlands are being vaccinated [1]. However, in the recently introduced human papilloma virus (HPV) vaccination, only 59% of all eligible girls participated [1]. In cancer screening as well, the success of one program does not automatically ensure similar participation rates in other screening programs. The Netherlands has extensive breast and cervical cancer-screening programs. While 83% of the eligible women participate in breast cancer screening [2], only 65% participate in the cervical cancer-screening program [3], and 68% participated during the first six months of the recently implemented screening program for colorectal cancer (CRC) [4]. Besides these nationwide-implemented vaccination and screening programs, several other initiatives are undertaken to improve the health of the population. For example, lifestyle programs are recommended in standardized care protocols for the treatment of type 2 Diabetes Mellitus (T2DM) patients in the Netherlands but participation rates remain unsatisfactory, ranging from 20%-70% [5-7]. Thus, HPV-vaccination, CRC screening and lifestyle programs are all proven to be (highly) effective at the individual level [6-12]. Nevertheless, these programs have not (yet) attained high participation rates and, consequently, the effects of these programs on public health will probably be lower than anticipated.

Characteristics that influence individuals' participation decisions have to be identified to be able to understand why they decide not to participate in effective preventive programs. Besides the well-investigated demographic and psychosocial characteristics and environmental variables, certain program characteristics such as costs or side effects, might also contribute to the low uptake of preventive programs. Insight into preferences of the target population for specific program characteristics is crucial for the development of new, attractive, and broadly used preventive programs, and for improving preventive programs that already exist. Until recently, preferences were mainly measured using traditional Likert-scale questions, asking respondents to indicate to what extent they agree or disagree with various statements or theorems. This method suffers from a range of drawbacks [13, 14]. Likert-scale questions increase respondents' tendency to provide socially desirable answers, to agree with the provided statements (i.e. acquiescence) and to make extreme choices (i.e. extreme response bias) [13]. Discrete Choice Experiments (DCEs) do not suffer from these limitations because respondents are required to choose between different sets of program characteristics rather than just ranking or rating a single characteristic. This method, which is relatively new in the field of public health [15-17], enables researchers to determine preferred program characteristics of both existing as well as non-existing programs and to quantify the importance of the measured program characteristics relative to each other. Therefore, DCEs appear to be a promising method to determine individuals' preferences concerning characteristics of specific preventive programs.

Discrete Choice Experiments

A DCE is a method to elicit individuals' preferences, to quantify the relative importance of product or service characteristics, and to determine potential uptake rates [18-22]. DCEs stem from mathematical psychology [23] and are widely used in the marketing and transportation research setting [24-27]. This method is now increasingly being used in health care and public health as well [18-22].

The DCE methodology is a stated preference method. Respondents are asked to complete several 'choice tasks' (see figure 1 for an example of a choice task). Each choice task consists of two or more scenarios that describe the program at hand. The description of the program is based on its characteristics or 'attributes'. It is assumed that the individual's preference for a scenario is determined based on the values of the levels of the included attributes. Based on the choices that individuals make, their preferences are elicited.

The DCE methodology is based on several normative theoretical assumptions that describe how people are assumed to make decisions when they are completing the choice tasks of a DCE. First, respondents are assumed to be actively involved in answering the choice tasks [18, 28]. Second, respondents are expected to choose rationally by means of complex decision strategies, which results in the most accurate response possible [18, 28]. Third, respondents are assumed to understand and correctly interpret the information provided to them [18, 28]. Fourth, respondents are expected to act in accordance with four axioms [28-31]: (1) they prefer attractive over less attractive attribute levels (monotonicity axiom), (2) they include all attributes in their decision and decide by trading off those attributes against each other (continuity axiom), (3) they have stable preferences, thus answering the same choice task twice should result in the same decision (completeness axiom), and (4) they have consistent preferences, if they prefer scenario 1 over scenario 2, and scenario 2 over scenario 3, then they also prefer scenario 1 over scenario 3 (transitivity axiom).

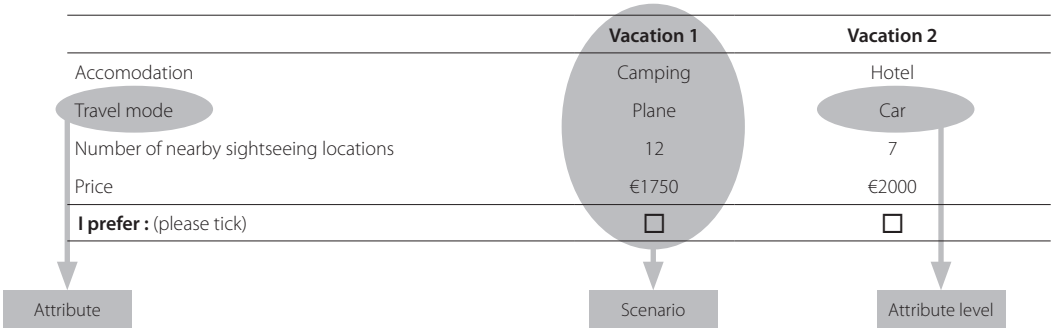

Figure 1. Example of a choice task measuring preferences for a two-week sightseeing vacation in Italy.

Theoretical basis

DCEs are based on the long-standing Random Utility Theory (RUT), which Thurstone first described in 1927 [31]. McFadden, Nobel Prize winner in 2000, extended this theory [32, 33], which still forms the basis of most of the DCEs that are currently conducted in the field of public health.

According to RUT, an individual derives a certain 'utility' for every possible scenario he or she is confronted with in a choice task [18, 20]. Since researchers cannot observe these utilities, they are called 'latent'. The latent utility consists of two elements: a measurable systematic element, and a random element that cannot be measured [18, 20]. The systematic element includes all attributes of the DCE together with all other measured covariates that determine individual decisions (e.g., demographic characteristics). The equation below specifies this further. The latent utility 'U' of individual 'n' concerning scenario 'j', can be estimated by taking the sum of the systematic utility element 'V' (i.e. the utility of individual 'n' concerning scenario 'j' calculated based on all attribute levels and covariates) and the random error term 'ε' (i.e. all unobserved and unobservable factors that influence the utility of person 'n' concerning scenario 'j') [18, 20].

$$U_{nj} = V_{nj} + \varepsilon_{nj}$$

The systematic element, which is expected to approximate an individual's true (but latent) utility, can be measured by means of a DCE. The systematic utility is calculated by taking the sum of all the attribute level estimates [18, 20, 22]. The estimates of the attribute levels are called 'part-worth utilities'. By comparing the part-worth utilities, conclusions can be drawn about how important the attributes are relative to each other [18, 20, 22]. The attribute of which the levels represent the largest range in part-worth utilities has the largest impact on utility, and is therefore relatively the most important attribute in the decision-making process at hand. Additionally, the trade-offs that people make between attributes when they are deciding between scenarios can be determined. This represents the amount of one attribute that an individual is willing to trade for another attribute [18, 20, 22]. If the price of a program is one of the included attributes, the trade-off reflects how much an individual is willing to pay for a certain increase or decrease in another attribute [18, 20, 22]. In light of the example choice task in Figure 1, one could for example calculate how many sightseeing locations people are willing to trade to stay in a hotel instead of a camping site or how much more people are willing to pay to travel by plane instead of by car. Finally, based on the attribute level estimates, researchers can estimate the potential participation rate for any specified hypothetical program [18, 20, 22].

The validity of the preferences that are measured with DCEs in public health

The DCE method has acquired an established position for studying decision-making behavior in marketing and transportation [18, 20]. Currently, DCEs are being used more and more frequently to measure both health care and public health related preferences, and results increasingly serve as input for policy making [15]. However, deciding about health-related issues is quite different and by definition more complex than other, more everyday decisions such as how one will commute to work, or what one will have for dinner. In part, this is due to the difficult concepts that people

are often required to evaluate when making health-related decisions. Risk information, such as the probability of experiencing side effects, is usually included in DCEs about public health interventions. In general, risks are difficult for respondents to interpret [34-36], especially the very small probabilities, as often reported for public health interventions. In order for respondents to be able to make accurate choices in a DCE that reflect their true preferences, it is critical that they understand all the information that is included in the DCE. The validity of the preferences is jeopardized if respondents do not fully understand how to complete choice tasks because they do not understand the attribute levels in the DCE. Altogether, it is uncertain whether people are able to complete health-related DCEs in a way that is in accordance with the theory underlying DCEs. In particular, specific subgroups in the population (e.g., older people, people with a lower educational level or less literate individuals) might experience difficulties. To date, little research has been conducted to test whether all of the theoretical assumptions about how respondents make decisions actually hold when respondents complete a DCE questionnaire. Some research, mainly focusing on the rationality assumption as well as the monotonicity and continuity axiom, suggests this might not always be the case [37-42].

Additionally, researchers stress the importance of studying the external validity of DCEs [43-45]. There is little to no evidence about the extent to which the stated preferences for a preventive program, which are measured by means of a DCE, resemble revealed preferences (actual behavior) of respondents [46]. Increased insight into the predictive value of DCEs in public health will provide more 'credibility and confidence' in this method [45].

The presentation of choice tasks in public health related DCEs

In social sciences, there is a significant amount of literature describing how human behavior (among which decision-making behavior) is influenced by the way information is presented [47-52]. However, other than the common notion that choice tasks should reflect the actual real-life choice situation as closely as possible, current DCE guidelines do not include detailed descriptions on how to present choice tasks to participants [19, 53]. It is probably for this reason that the presentation of choice tasks is poorly described in DCE literature [28, 54, 55].

Besides the lack of guidelines, there is very limited to no empirical evidence on the effects of different ways of presenting choice tasks on respondents' decision-making behavior and on the outcomes of the DCE in the public health setting. Nevertheless, the way choice tasks in a DCE are presented probably plays a crucial role in communicating clear and understandable information. This is vital because, as described previously, DCEs about public health topics are often perceived to be difficult. Moreover, researchers have to make decisions concerning various presentation options every time they conduct a DCE study, for example, on the issue of how to include the opt-out option and how to frame and depict the attribute levels. Currently, those decisions cannot be based on evidence of good practice. To ensure the quality of the data, which are retrieved by DCEs, a combination of knowledge from the social sciences and additional research regarding the effects of choice task presentation on respondents' decision-making behavior and on the outcomes of a DCE is necessary.

The aim and objectives of this thesis

This thesis aims to determine whether DCEs can be used to measure individuals' preferences for public health interventions. This aim has been further specified in three detailed objectives for three different parts of this thesis:

1. To measure individuals' preferences for specific public health interventions with a DCE;
2. To study the validity of the preferences that are measured with DCEs in public health;
3. To assess how the presentation of choice tasks influences respondents' decision-making behavior and the outcomes of a DCE.

Structure of this thesis

In part 1 of this thesis, three applications of DCEs in the public health research setting will be described. In **Chapter 1**, preferences of type 2 Diabetes Mellitus patients with respect to a lifestyle intervention will be discussed. **Chapter 2** describes parental preferences regarding a rotavirus vaccine for young children and **Chapter 3** explores the preferences of the general population aged 55-65 regarding genetic screening for colorectal cancer within the population-based colorectal cancer-screening program.

In part 2 of this thesis, three aspects of the validity of the preferences that are measured with DCEs in the public health setting will be studied. **Chapter 4** describes how respondents complete a DCE questionnaire and to what extent the normative theoretical assumptions of the DCE methodology hold when respondents complete a DCE. **Chapter 5** investigates whether health literacy is associated with parental preferences for rotavirus vaccination. **Chapter 6** describes an exploratory application in which the predictive value of a DCE was tested by means of comparing the measured (i.e. stated) preferences of type 2 Diabetes Mellitus patients regarding a lifestyle intervention with their actual participation decision.

In part 3 of this thesis, it is investigated to what extent including an opt-out option **(Chapter 7),** using graphics to depict attribute levels **(Chapter 8)** and using positive or negative framing to describe a risk attribute **(Chapter 9)** influences respondents' decision-making behavior and the outcomes of a DCE.

This thesis is finalized with a discussion and summary of our findings.

References

1. van 't Schurink-van 't Klooster, T.M. and H.E. De Melker, *The National Immunisation Programme in the Netherlands: Surveillance and developments in 2013-2014*. 2014, National Institute for Public Health and the Environment: Bilthoven.
2. National Evaluation Team Breastcancer Screening, *National Evaluation of the Population Based Breastcancer Screening Progran in the Netherland 1990-2011/2012*. 2012, Erasmus Medical Center: Rotterdam.
3. National Evaluation Team Cervicalcancer Screening, *National Evaluation of the Population Based Cervicalcancer Screening Progran in the Netherland 1990-2011/2012*. 2012, Erasmus Medical Center: Rotterdam.
4. Penning, C., Lansdorp-Vogelaar, I., van Leerdam, M.E., van der Meulen, M.P., van Vuuren, A.J., Kuipers, E.J., Bonfrer, J.M., van Kemenade, F.J., Biermann, K., and M.G.J. Thomeer, Spaander, V.M.C.W., Kroep, S., van Ballegooijen, M., de Koning, H.J., *Landelijke monitoring van het Bevolkingsonderzoek Darmkanker: resultaten eerste half jaar 2014 [National surveillance of the colorectal cancer screening program: results of the first 6 months in 2014]*. 2014, ErasmusMC: Rotterdam.
5. James, D.V.B., et al., *Factors associated with physical activity referral uptake and participation.* Journal of Sports Sciences, 2007. 26(2): p. 217-24.
6. Knowler, W.C., et al., *Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin.* N Engl J Med, 2002. 346(6): p. 393-403.
7. Wing, R.R., *Long-term effects of a lifestyle intervention on weight and cardiovascular risk factors in individuals with type 2 diabetes mellitus: four-year results of the Look AHEAD trial.* Arch Intern Med, 2010. 170(17): p. 1566-75.
8. Koutsky, L.A., et al., *A cohort study of the risk of cervical intraepithelial neoplasia grade 2 or 3 in relation to papillomavirus infection.* N Engl J Med, 1992. 327(18): p. 1272-8.
9. Diabetes Prevention Program Research Group, *The Diabetes Prevention Program (DPP): description of lifestyle intervention.* Diabetes Care, 2002. 25(12): p. 2165-71.
10. Lindstrom, J., et al., *Prevention of diabetes mellitus in subjects with impaired glucose tolerance in the Finnish Diabetes Prevention Study: results from a randomized clinical trial.* J Am Soc Nephrol, 2003. 14(7 Suppl 2): p. S108-13.
11. Pignone, M., et al., *Screening for colorectal cancer in adults at average risk: a summary of the evidence for the U.S. Preventive Services Task Force.* Ann Intern Med., 2002. 137(2): p. 132-41.
12. Walsh, J.M. and J.P. Terdiman, *Colorectal cancer screening: scientific review.* JAMA., 2003. 289(10): p. 1288-96.
13. Pauhus, D.L., *Measurement and conrol of response bias*, in *MEasures of personality and social psychological attritudes*, J.P. Robinson, P.R. Shaver, and L.S. Wrighsman, Editors. 1991, Academic Press: San Diego CA. p. 17-59.
14. Hawkins, G.E., et al., *Integrating cognitive process and descriptive models of attitudes and preferences.* Cogn Sci., 2014. 38(4): p. 701-35.
15. Clark, M.D., et al., *Discrete Choice Experiments in Health Economics: A Review of the Literature.* Pharmacoeconomics, 2014. 9: p. 9.
16. Viney, R., E. Lancsar, and J. Louviere, *Discrete choice experiments to measure consumer preferences for health and healthcare.* Expert Rev Pharmacoecon Outcomes Res, 2002. 2(4): p. 319-26.
17. Ryan, M. and K. Gerard, *Using discrete choice experiments to value health care programmes: current practice and future research reflections.* Appl Health Econ Health Policy, 2003. 2(1): p. 55-64.
18. Hensher, D.A., J.M. Rose, and W.H. Greene, *Applied Choice Analysis: A Primer*. 2005, New York: Cambridge University Press.
19. Lancsar, E. and J. Louviere, *Conducting discrete choice experiments to inform healthcare decision making: a user's guide*. Pharmacoeconomics, 2008. 26(8): p. 661-77.
20. Louviere, J.J., D.A. Hensher, and J.D. Swait, *Stated Choice Methods; analysis and apllication*. 2000, Cambridge: Cambridge University Press.
21. Ryan, M. and S. Farrar, *Using conjoint analysis to elicit preferences for health care.* Bmj, 2000. 320(7248): p. 1530-3.
22. Ryan, M., K. Gerard, and M. Amaya-Amaya, *Using Discrete Choice Experiments to Value Health and Health Care*. The Economics of Non-Market Goods and Resources, ed. I.J. Bateman. 2008, Dordrecht: Springer.
23. Luce, R.D. and J.W. Tukey, *Simulataneous conjoint measurement: a new type of fundamental measurement.* J Math Psychol, 1964. 1: p. 1-27.

24. Wittink, D.R. and P. Cattin, *Commercial Use of Conjoint Analysis: An Update.* J Market, 1989. 53(3): p. 91-96.
25. Green, P.E. and V. Srinivasan, *Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice.* J Market, 1990. 54(4): p. 3-19.
26. Louviere, J.J., *Conjoint Analysis Modelling of Stated Preferences: A Review of Theory, Methods, Recent Developments and External Validity.* J Trans Econ Policy, 1988. 22(1): p. 93-119.
27. Hensher, D., *Stated preference analysis of travel choices: the state of practice.* Transportation, 1994. 21: p. 107-133.
28. Lloyd, A.J., *Threats to the estimation of benefit: are preference elicitation methods accurate?* Health Econ, 2003. 12(5): p. 393-402.
29. Cairns, J. and M. van der Pol, *Repeated follow-up as a method for reducing non-trading behaviour in discrete choice experiments.* Soc Sci Med, 2004. 58(11): p. 2211-8.
30. Lancsar, E. and J. Louviere, *Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences?* Health Econ, 2006. 15(8): p. 797-811.
31. Thurstone, L.L., *The Method of Paired Comparisons for Social Values.* J Abnorm Soc Psychol, 1927. 21: p. 384-400.
32. McFadden, D., *The Choice Theory Approach to Market Research.* Marketing Science, 1986. 5(4): p. 275–297.
33. McFadden, D., *Conditional Logit Analysis of Qualitative Choice Behavior*, in *Frondtiers in Econometrics*, P. Zarembka, Editor. 1974, Academic Press: New York.
34. Galesic, M. and R. Garcia-Retamero, *Statistical numeracy for health: a cross-cultural comparison with probabilistic national samples.* Arch Intern Med., 2010. 170(5): p. 462-8.
35. McCaffery, K.J., et al., *The influence of graphic display format on the interpretations of quantitative risk information among adults with lower education and literacy: a randomized experimental study.* Med Decis Making., 2012. 32(4): p. 532-44.
36. Waters, E.A., et al., *Formats for improving risk communication in medical tradeoff decisions.* J Health Commun., 2006. 11(2): p. 167-82.
37. Ryan, M., V. Watson, and V. Entwistle, *Rationalising the 'irrational': a think aloud study of discrete choice experiment responses.* Health Econ, 2009. 18(3): p. 321-36.
38. Miguel, F.S., M. Ryan, and M. Amaya-Amaya, *'Irrational' stated preferences: a quantitative and qualitative investigation.* Health Econ, 2005. 14(3): p. 307-22.
39. Alemu, M.H., et al., *Attending to the reasons for attribute non-attendance in choice experiments.* Environ Resour Econ, 2013. 54(3): p. 333-359.
40. Lagarde, M., *Investigating attribute non-attendance and its consequences in choice experiments with latent class models.* Health Econ, 2013. 22(5): p. 554-67.
41. Ryan, M. and A. Bate, *Testing the assumptions of rationality, continuity and symmetry when applying discrete choice experiments in health care.* Appl Econom Lett, 2001. 8(1): p. 59-63.
42. Whitty, J.A., et al., *A Think Aloud Sutdy Comparing the Validity and Acceptability of Discrete Choice and Best Worst Scaling Methods.* PLoS ONE, 2014. 9(4): p. e90635.
43. de Bekker-Grob, E.W., M. Ryan, and K. Gerard, *Discrete choice experiments in health economics: a review of the literature.* Health Econ., 2012. 21(2): p. 145-72.
44. Lancsar, E. and J. Swait, *Reconceptualising the External Validity of Discrete Choice Experiments.* Pharmacoeconomics, 2014. 12: p. 12.
45. Louviere, J.J. and E. Lancsar, *Choice experiments in health: the good, the bad, the ugly and toward a brighter future.* Health Econ Policy Law., 2009. 4(Pt 4): p. 527-46.
46. Krucien, N., A. Gafni, and N. Pelletier-Fleury, *Empirical Testing of the External Validity of a Discrete Choice Experiment to Determine Preferred Treatment Option: The Case of Sleep Apnea.* Health Econ, 2014. 1(10).
47. Kahneman, D., J.L. Knetsch, and R.H. Thaler, *The endowment effect, loss aversion and status quo bias.* J of Econ Perspect, 1991. 5(1): p. 193-206.
48. Lipkus, I.M., *Numeric, verbal, and visual formats of conveying health risks: suggested best practices and future recommendations.* Med Decis Making., 2007. 27(5): p. 696-713.
49. Luce, M.F., J.W. Payne, and J.R. Bettman, *Emotional trade-off difficulty and choice.* J Market Res, 1999. 36(2): p. 143-159.

50. Timmermans, D.R.M., C.F. Ockhuysen-Vermey, and L. Henneman, *Presenting health risk information in different formats: The effect on participants' cognitive and emotional evaluation and decisions.* Patient Educ Couns, 2008. 73: p. 443-447.

51. Tversky, A. and D. Kahneman, *The framing of decisions and the psychology of choice.* Science, 1981. 211(4481): p. 453-8.

52. Tversky, A. and D. Kahneman, *Judgment under uncertainty: Heuristics and biases.* Science, 1974. 185(4157): p. 1124-1131.

53. Bridges, J.F., et al., *Conjoint analysis applications in health--a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force.* Value Health., 2011. 14(4): p. 403-13.

54. Bryan, S. and P. Dolan, *Discrete choice experiments in health economics. For better or for worse?* Eur J Health Econ, 2004. 5(3): p. 199-202.

55. Marshall, D., et al., *Conjoint Analysis Applications in Health - How are Studies being Designed and Reported?: An Update on Current Practice in the Published Literature between 2005 and 2008.* Patient., 2010. 3(4): p. 249-56.

PART 1

MEASURING INDIVIDUALS' PREFERENCES FOR PUBLIC HEALTH INTERVENTIONS WITH A DISCRETE CHOICE EXPERIMENT

CHAPTER 1

TYPE 2 DIABETES PATIENTS' PREFERENCES AND WILLINGNESS TO PAY FOR LIFESTYLE PROGRAMS

J. Veldwijk, M.S. Lambooij, P.F. van Gils,
J.N. Struijs, H.A. Smit, G.A. de Wit

BMC: Public Health (2013), 13: 1099.

ABSTRACT

Background: Participation rates of lifestyle programs among type 2 diabetes mellitus (T2DM) patients are less than optimal, even though research shows notable delays in the development of the disease among lifestyle program participants. Very little is known about the relative importance of barriers for participation as well as T2DM patients' willingness to pay for participation in such programs. The aim of this study was to identify the preferences of T2DM patients with regard to lifestyle programs, to calculate participants' willingness-to-pay (WTP) as well as to estimate the potential participation rates of lifestyle programs.

Methods: A Discrete Choice Experiment (DCE) questionnaire assessing five different lifestyle programs was distributed among 1250 Dutch adults aged 35-65 years with T2DM, 391 questionnaires (31%) were returned and included in the analysis. The relative importance of the program attributes (i.e. meal plan, physical activity (PA) schedule, consultation structure, expected program outcome and out-of-pocket costs) was determined using panel-mixed logit models. Based on the retrieved attribute estimates, patients' WTP and potential participation rates were determined.

Results: The out-of-pocket costs (β=-0.75, P<.001), consultation structure (β=-0.46, P<.001) and expected outcome (β=0.72, P<.001) were most important for respondents when deciding whether to participate in a lifestyle program. Respondents were willing to pay €128 per year for individual instead of group consultation and €97 per year for 10 kilograms anticipated weight loss. Potential participation rates for different lifestyle-program scenarios ranged between 48.5% and 62.4%.

Conclusions: When deciding whether to participate in a lifestyle program, T2DM patients are mostly driven by low levels of out-of-pocket costs. Thereafter, they prefer individual consultation and high levels of anticipated outcomes with respect to weight loss.

INTRODUCTION

Participation rates of lifestyle programs among type 2 diabetes mellitus (T2DM) patients are less than optimal [1, 2]. Yet, there appear to be notable delays in the development of the disease and the onset of diabetes-related complications among participants in different lifestyle programs [1-5]. Driven by these results, combined with the increasing prevalence of T2DM and high disease-specific mortality rates [6, 7], lifestyle interventions have been included in standardized care protocols for the treatment of T2DM patients in several countries [8-10]. However, in general, the participation rates of such lifestyle programs among T2DM patients are unsatisfactory, ranging from 20-70% of those eligible for such programs [1, 2, 11]. Suboptimal participation rates are especially worrisome among the growing population of T2DM patients aged 35-65 years. Lifestyle changes among patients within this specific age group are expected to have a greater and more long-term impact on disease progress than among older T2DM patients [12]. Moreover, these relatively young T2DM patients suffer less from serious diabetes-related physical or medical restrictions [13, 14] or from other chronic diseases or disabilities [15, 16]. Therefore, younger T2DM patients are relatively suitable candidates for participation in a lifestyle program.

In order to obtain a better insight into the motives of T2DM patients for participating in lifestyle interventions, various studies were conducted that resulted in an extensive list of barriers for participation in lifestyle interventions as reported by T2DM patients [17-21]. However, very little is known about the relative importance of such factors for these patients, while it seems reasonable to assume that not all of the factors are of equal importance in the decision-making process regarding the participation in a lifestyle program. Previous research has shown this among other target populations [22-25]. A second omission in many studies on the willingness to participate in lifestyle programs is that the costs of lifestyle programs are often not included. If lifestyle interventions were to be implemented in 'real life', participants in such programs would have to pay at least part of the program costs out-of-pocket. Whereas participation in programs within a research setting tends to be free of charge. Taking into consideration that these programs need to be (partly) financed by the participants, it is worthwhile to examine the amount of money that potential participants would be willing to pay. Previous research among a non-diabetic population showed that the 582 individuals at high risk of developing T2DM were willing to pay out-of-pocket for a lifestyle intervention [23]. These rates varied between $63 and $5 per month for a three-year course depending on the diet restrictions, hours of exercise, hour of counseling, use of medication, the goal that was set with respect to weight loss and the percentage expected reduction in T2DM risk [23].

Once we have obtained a better insight into program-related factors that are crucial for the participation of T2DM patients in lifestyle programs as well as for their willingness to pay (WTP), recommendations can be made as to what type of program would most likely be preferred by its potential users. These recommendations can be taken into account when developing new lifestyle programs, thus increasing their reach, and hence their public health benefit.

The aim of this study was to identify the preferences of T2DM patients aged 35-65 years for different characteristics of lifestyle programs. Based on these preferences, participants' willingness-to-pay (WTP) as well as the potential participation rates of different lifestyle programs have been assessed.

MATERIALS AND METHODS

Participants and recruitment

Within the Netherlands, generic diabetes care is arranged in care groups. A care group is a legal entity formed by multiple health care providers (however, these are often exclusively general practitioners (GPs)) [26]. A random selection of care groups (per province of the Netherlands) was contacted to distribute the questionnaire among T2DM patients aged 35-65 years who were not suffering from any serious diabetes-related complications (i.e. cardiovascular diseases, nephropathy, retinopathy, neuropathy) and who were registered with that care group. In total five care groups, located within five different provinces of the Netherlands, participated. These care groups distributed a total number of 1250 questionnaires to all the eligible patients within their care groups, 391 (31.3%) of which were completed and included in the analysis. Due to confidentiality agreements with the care groups that distributed the questionnaires, no reminder letters could be sent. As a result, there was no non-response information available to empirically test whether responders differed from non-responders with respect to their demographic and disease-specific characteristics. The Dutch National Ethics Board (Central Committee on Research involving Human Subjects) concluded that formal testing by a medical ethical committee was not necessary, as T2DM patients were only required to complete an anonymous questionnaire once, which is in accordance with the guidelines laid down in the Declaration of Helsinki.

The mean age of the final study population was 57.2 years, more than half of the respondents were male (57.4%), almost all respondents were Dutch nationals (94.1%), and about half of the population (49.6%) completed higher secondary education or lower general professional education (Coded 'medium' in Table 1). On average, the participants were diagnosed with T2DM 6.1 years ago and they had a mean HbA_{1c} of 49.9 mmol/mol. Of all the respondents, 37.8% was overweight (indicated by a BMI between 25 and 30 kg/m^2) and 41.1% was obese (indicated by a BMI higher than 30 kg/m^2). The majority of the respondents reported no complications (78.3%) and no other chronic conditions (99.7%). 19% of the respondents reported that they did not use any form of medication, 67.2% reported that they used glucose lowering medication in the form of pills, 4.6% reported that they injected insulin, and 9.2% reported that they used both. Almost all respondents (94.1 %) reported that their primary health care contact was their GP and not a specialist in secondary care (e.g., an internist). Self-management measures were applied by approximately half of this population; 45.8% reported that they monitored their HbA_{1c} at home and 50.3% reported that they keep a T2DM diary.

In total 47.2% of this population reported that they considered lifestyle programs to be useful and 66.7% thought that their partner, family or friends would support their participation in a lifestyle program, 33.5% felt that they would be able to complete such a lifestyle program and 22.6% would actually like to participate in such a program (Table 1).

Table 1. General characteristics and psychosocial determinants of the study population (n=391).

		Mean (SD)	Percentage
Age (n = 385)		57.2 (6.4)	
Gender (n = 390)	Male		57.4
Educational level (n = 379)	Low		31.1
	Medium		49.6
	High		19.2
Ethnicity (n = 390)	Dutch		94.1
Duration of diabetes (years) (n = 382)		6.1 (5.5)	
HbA1c (mmol/mol) (n = 101)		49.9 (16.6)	
Primary health care contact (n = 187)	GP		94.1
Medication (n = 390)	None		19.0
	Oral glucose lowering medication		67.2
	Insulin		4.6
	Both		9.2
Complications present (n = 303)			21.7
Chronic condition present (n = 390)			0.3
Weight category* (n = 364)	Underweight		0.2
	Normal weight		19.5
	Overweight		37.8
	Obese		41.1
EQ5d score (n = 391)		0.91 (0.19)	
Self-management (n = 386)	monitoring HbA1c at home		45.8
	Keeping a T2DM diary		50.3
What is your opinion concerning lifestyle programs in general? (n = 388)	Very useful or useful		47.2
	Not useful at all		2.6
Do you think you are able to complete a lifestyle program that endures 1 year, without dropping out? (n = 391)	Certainly or probably		33.5
	Certainly not		18.2
Would your partner, friends and/or family support you if you would participate in a lifestyle program? (n = 391)	Certainly or probably		66.7
	Certainly not		4.9
Would you like to participate in a lifestyle program? (n = 391)	Certainly or probably		22.6
	Certainly not		19.4

* Respondents were categorized as underweight if their BMI < 20 kg/m^2, normal weight if BMI 20-25 kg/m^2, overweight if BMI 25-30kg/m^2 and obese if BMI >30kg/m^2.

Discrete Choice Experiment

DCEs are becoming a frequently used tool in public health research to estimate the potential participation rates of interventions or medical treatments and to provide knowledge about which of the program components determine these participation rates [27, 28]. The DCE methodology is based

on the Random Utility Theory and assumes that any intervention or treatment can be described by its characteristics (i.e. attributes; such as frequency of consultation). Individuals' preference for an intervention or treatment is dependent on the levels (e.g., weekly or monthly consultation) of those attributes [27, 28]. By varying the levels of the attributes, different scenarios are constructed. Respondents are provided with at least two scenarios (i.e. choice tasks) simultaneously, they then have to choose the scenario that they prefer most. Each respondent is asked to complete a series of such choice tasks. In the end, conclusions can be drawn regarding the attributes that constitute an intervention that is most preferred by its potential users.

Attributes, levels and design

The attributes and levels included in the current study were determined in a stepwise manner, which subsequently included a literature review, expert interviews and focus group interviews with T2DM patients. First, a list of barriers for participating in a lifestyle intervention by T2DM patients was compiled based on previously published literature [17-21]. Second, the list of barriers was discussed during expert interviews with a physician, a dietician and a scientist with a specific interest in diabetes care. These expert interviews were conducted in order to (1) shorten the list of potential attributes and (2) to ensure that the attributes and levels were consistent with current practice. As a third step, four focus group interviews were conducted with 24 T2DM patients in order to ensure that (1) the most important attributes for the decision-making process of T2DM patients were included in the DCE and (2) proper levels were appointed to each of the attributes. Focus group interviews were conducted according to Krueger and colleagues [29]. This process led to the inclusion of five attributes (meal plan, physical activity (PA) schedule, consultation structure, expected outcomes and out-of-pocket costs) with three levels (Table 2).

Table 2. The attributes and levels that were included in this discrete choice experiment

Attributes	Level 1	Level 2	Level 3
Meal plan			
A plan, which describes the aims of the participants with respect to improvements in their diet, developed by the participants of the program together with a coach			
	Flexible: primarily based on the participants' own initiatives and ideas	General: includes general information on a healthy diet and provides example recipes	Elaborate: a patient tailored schedule that is completely prepared by the lifestyle coach
Physical activity (PA) schedule			
A plan, which describes the aims of the participants with respect to improvements in their PA behavior, developed by the participants of the program together with a coach			
	Flexible: primarily based on the participants' own initiatives and ideas	General: includes general information on PA, and provides example exercises	Elaborate: a patient tailored schedule that is completely prepared by the lifestyle coach
Consultation structure			
The composition of the consults with the coach			
	Individually	Groups with 5 other T2DM patients	Groups with 10 other T2DM patients
Expected outcomes			
The results, in terms of weight loss & physical fitness expected by the respondent after completion of a lifestyle program			
	No weight loss but feeling more healthy	5 kilograms of weight lost and feeling more healthy	10 kilograms of weight lost and feeling more healthy
Out-of-pocket costs			
Patients may have to pay (part) of the program costs out-of-pocket			
	75 euro per year	150 euro per year	225 euro per year

Based on the selected attributes and levels, NGene 1.0 (ChoiceMetrics, 2011) software was used to develop a D-efficient design with 18 unique choice tasks [30, 31]. To limit the burden of respondents, NGene divided these 18 choice tasks over two sets of nine choice tasks, each set of nine choice tasks was disseminated among half of the study population. Besides choosing one of the lifestyle programs presented, participants could also choose an opt-out solution. This opt-out option was included because, in real life, people can also choose not to participate in a lifestyle program. Table 3 presents an example of a choice task as included in the questionnaire of this study. Before completing these choice tasks, respondents were provided with an extensive explanation of the meaning of all attributes and levels as well as an explanation about how to deal with a choice task, accompanied by an example. Every choice task starts with the question: 'Imagine that your GP or nurse practitioner advises that you participate in a lifestyle program for a period of one year. In which scenario would you prefer to participate, scenario 1 or scenario 2? If you do not wish to participate in either of the scenarios, you can tick the box "none". A questionnaire containing additional questions was added to the DCE, for further details on this questionnaire see additional file 1.

Table 3. Example of a choice task

	Scenario 1	**Scenario 2**	**None**
Meal plan	Flexible	General	None
Physical activity schedule	General	Elaborate	None
Consultation structure	Individual	In groups of 5 patients	None
Expected outcome	5 kg weight loss and feeling more healthy	10 kg weight loss and feeling more healthy	None
Out-of-pocket costs	150 euro per year	150 euro per year	0 euro per year
Tick the box of the scenario that you prefer:	☐	☐	☐

The draft questionnaire was pilot tested among a subgroup (n=20) of the study population to ensure that the wording used in the questionnaire was correct and understood by the target population. Most of the pilot tests were performed by means of a postal questionnaire, respondents were requested to mark every question or answering category that they did not understand or found hard to grasp and they were asked to provide suggestions for improvement. Moreover, three think aloud pilot tests were conducted to obtain more insight into how the respondents answered the choice tasks. No changes in the attributes and/or levels were deemed necessary based on the results of this pilot study. Moreover, based on the pilot-test data, sample size calculations were performed to ensure that significant differences for each attribute could be detected at a 5% level [30, 32].

Statistical analyses

NLogit 4.0 (Econometric Software, 2007) was used to construct the panel-mixed-logit (Panel-MIXL) models that were estimated within this study. When using such models, results are adjusted for the panel structure (i.e. multilevel structure) of the data. As every respondent completed nine choice

tasks, their answers may be correlated, which is then accounted for. Equation 1.1 was tested using these models.

[Eq. 1.1]

$U = V + \varepsilon = \beta_0 + \beta_1 *$ flexible meal plan $+ \beta_2 *$ elaborate meal plan $+ \beta_3 *$ flexible PA schedule $+ \beta_4 *$ elaborate PA schedule $+ \beta_5 *$ consultation in groups of 5 $+ \beta_6 *$ consultation in groups of 10 $+ \beta_7 *$ expected outcome $+ \beta_8 *$ out-of-pocket costs $+ \varepsilon$

V describes the measurable utility of a specific lifestyle program based on the attributes that were included in the DCE. β_0 represents the alternative specific constant and $\beta_1 - \beta_8$ are the attribute level estimates that indicate the relative importance of each attribute. Based on the model fit (AIC and Chi-square tests), the constant and the expected outcome attributes were set as random parameters (both with a normal distribution). All nonlinear parameters were recoded into effect codes. In contrast to dummy coding, this coding procedure codes the reference category -1, so the sum of the effect coded attributes is always 0. The coefficient for the reference category is therefore $-1*(\beta_{\text{effect code 1}} + \beta_{\text{effect code 2}})$ [33, 34].

In order to calculate patients' marginal willingness-to-pay (WTP), the negative of the out-of-pocket attribute was used as a measure of the marginal utility of money. The ratio of either attribute estimate to this negative of the out-of-pocket cost attribute provides an estimation of patients' WTP concerning that specific attribute [28, 35]. Moreover, the potential participation rates (choice probabilities) of a program that consists of a specific set of attributes was estimated. Since both the constant and the expected outcome attribute were included as random parameters in the analyses, choice probabilities could not be calculated directly, therefore a simulation was used [28, 33]. The mean participation rates of all simulations (n=1000) was estimated by taking the average of all simulated participation rate probabilities, which were calculated as $1/(1+\exp^{-V})$.

For a more detailed description of the statistical methods used in this paper see additional file 2.

RESULTS

Patient preferences

Most of the attribute level estimates were significant, indicating that they were important for T2DM patients when choosing whether to participate in a lifestyle program (Table 4). Participants did not have any distinct preferences with regard to the meal plan. However, they did prefer a general PA schedule above a flexible PA schedule. Participants reported a preference for individual consultation, as compared to consultation in groups of 10 patients. The greater the expected outcome in terms of weight loss, the more willing participants were to participate and higher out-of-pocket costs led to a decrease in their willingness to participate.

Table 4. T2DM patients' preferences for a lifestyle program: the attribute estimates of the Panel MIXL model

Attribute		Estimate	SE
Constant	Mean	0.11	0.13
	Standard deviation	2.61*	0.71
Meal plan	Flexible	0.11	0.07
	General (ref)	-0.04	0.06
	Elaborate	-0.06	0.06
PA schedule	Flexible	-0.13*	0.06
	General (ref)	0.02	0.01
	Elaborate	0.11	0.06
Consultation structure	Individual (ref)	0.50	0.08
	Groups of 5	-0.04	0.06
	Groups of 10	-0.46***	0.08
Expected outcome (10kg)	Mean	0.72***	0.16
	Standard deviation	1.53*	0.51
Out-of-pocket costs (€100)		-0.75***	0.08

* significant at p<.05; *** significant at p<.001.

Since the magnitude of the beta values depends highly on the coding of the attributes, attributes were recoded into the same coding scale (all attribute levels were coded between -1 and 1) to enable the assessment of their relative importance. The results of the recoded analysis are not shown because they show a high degree of overlap with the results presented in Table 4 and provide betas that are difficult to interpret especially concerning the outcome and cost attributes. The results reveal that, based on the value of its coefficient, the out-of-pocket costs were the most decisive factor for respondents in determining whether they wanted to participate in a lifestyle program. This attribute was followed by the consultation structure and expected outcome. The least important factor in the decision-making process was the operationalization of the PA schedule.

The significant coefficient of the standard deviation of the expected outcome attribute indicates that there is indeed a high preference heterogeneity among respondents concerning the amount of weight loss they anticipate before starting the intervention.

Finally, the opt-out option was chosen in 46.3% of the choice tasks and 23.5% of the respondents chose to opt-out in every choice task. Patient preferences did not change when these latter responders were excluded from the analysis.

Willingness to pay

The WTP was calculated for the significant attributes only. The results show that respondents were willing to pay €21.0 (95% CI: €11.3; €30.7) for a switch from a flexible to a general PA schedule.

Respondents were willing to pay an extra €127.8 (95% CI: €106.0; €149.7) per year for a lifestyle program organized via individual consultation instead of consultation with a group of 10 other patients. Respondents were willing to pay €96.8 (95% CI: €85.2; €108.4) per year for 10 kg anticipated weight loss.

Potential participation rate

A lifestyle program that consists of all of the least preferred attribute levels (flexible meal plan, flexible PA schedule, consultation in groups of 10, no weight loss) was set as the 'base' model. This program showed the lowest participation rate (48.5%) (Table 5). The potential participation rate increased if this 'base' model was adapted to the identified patient preferences (change to 51.5%-57.4%). The most preferred program, which includes a flexible meal plan, general PA schedule, individual consultation and a 10 kg weight loss, resulted in an estimated potential participation rate of 62.4%.

Table 5. Expected participation rates for different lifestyle programs based on the attribute estimates of the Panel-MIXL model*

	Participation rates (%)	Explanation
Base model	48.5	A program with a general meal plan, a flexible PA schedule, consultation in groups of 10 and no weight loss
PA schedule	51.5	Base model with a general PA schedule
Consultation structure	54.8	Base model with individual consultation
Expected outcome	57.4	Base model with a 10 kg weight loss
Preferred program	62.4	A program with a flexible meal plan, a general PA schedule, individual consultation and a 10kg weight loss

*Out-of-pocket costs were held constant at €0

DISCUSSION

Our research is the first to demonstrate the relative importance of the factors that affect T2DM patients' preferences for a lifestyle intervention program. Results showed that the out-of-pocket costs were the most crucial factor for T2DM patients when deciding whether to participate in a lifestyle program (i.e. patients preferred lowest costs). Moreover, a lifestyle program with a general physical activity component, individual consultation and large expected outcomes in terms of weight loss was preferred. T2DM patients were willing to pay €21, €128 and €97 per year respectively for a lifestyle program with these desired levels of the attributes (i.e. a general PA schedule, individual consultation, and 10 kg anticipated weight loss). Additionally, it was estimated that approximately 62% of the T2DM patients aged 35-65 years would participate in a lifestyle program with these preferred levels.

Though there is limited evidence regarding T2DM patients' preferences for a lifestyle program, the current findings are comparable to those of previously conducted DCEs among other target

populations describing different lifestyle interventions. The study by Johnson [23] found that individuals at a high risk of developing T2DM preferred a lifestyle program that specified anticipated weight loss over a program that did not describe any anticipated weight loss. They also found that programs with restrictive diets were disliked, that some sort of physical activity component was preferred, and that the respondents were willing to pay up to approximately $63 per month for a total of 36 months for participation in the lifestyle program they preferred most [23]. Roux and colleagues [25], as well as Owen and colleagues [24], reported that participants in a lifestyle program preferred personally oriented programs that included both a diet and some sort of PA component. The diet component in the program should not be too restrictive [24], and participants were willing to pay for participation [25].

Patients in the current study reported to be willing to pay up to €97 per year for every 10 kilograms of anticipated weight loss. Though this seems a promising argument to boost participation without significantly increasing costs, one could argue that, despite the linearity of the initial costs and expected outcome attributes, beyond a certain point, participants are no longer willing to pay an additional €97 for an increase of 10 kilograms of anticipated weight loss. This was also shown by Johnson and colleagues who demonstrated nonlinearity of the WTP [36].

Finally, results showed that a potential participation rate of 62% can be expected when attributes are operationalized in accordance with patient preferences. Current programs can be improved by organizing individual consultation and communicating clearly about the anticipated outcomes of the program (in terms of weight loss and degree of physical fitness of the participant). The participation rates found in this study show that T2DM patients are willing to participate if programs meet certain criteria. However, the participation rates do not exceed the most optimistic participation rates currently observed in lifestyle programs [1, 2, 11]. This suggests that T2DM patients have additional motives for not participating in a lifestyle program other than the tested characteristics of a lifestyle program. These motives may differ considerably between individuals; previous research already suggested tailored lifestyle programs to enhance patient commitment [37, 38]. Tailoring such programs to individual patients can be costly and less feasible compared to generic programs. It should be explored which factors, other than consultation structure and clear communication with respect to expected outcomes, can be maintained over the total target population in order to limit the variation between programs and to keep costs as low as possible while at the same time increasing patient commitment.

Our conclusions are restricted by a number of limitations. The usable response rate was 31.3%. As the current study was questionnaire-based and participation was on a voluntary basis, selective non-response seems plausible. For instance, the number of non-Dutch patients was relatively low. This may be due to language difficulties, as a good command of the Dutch language is needed to complete the questionnaire and especially the DCE tasks. Therefore, generalizability with regard to preferences of non-Dutch patient groups remains limited. Besides, it could be the case that patients who already perceive their lifestyle as being healthy chose not to participate in this study. However,

in real life, it is not likely that these patients would participate in a lifestyle program and therefore these patients are of limited interest for this specific study. The current study included T2DM patients in the age group of 35-65 years, who do not suffer from severe diabetes-related complications. There is limited information on the representativeness of the current study population compared to the target population. Additional analyses of T2DM patients aged 35-65 in a large Dutch cohort study (EPIC-NL [39]) showed the same mean BMI values, but other characteristics could not be compared because of the specific inclusion and exclusion criteria of the current study (in particular, the exclusion of patients with severe diabetes-related complications in this study). Future research should be conducted among T2DM patients aged > 65years without any restrictions with respect to the presence of diabetes-related complications to obtain more insight into the preferences of this subpopulation. Researchers should then take into account that the attributes of the current DCE might not be applicable within this new target population. Conducting such research would contribute to the insights into (the differences and similarities in) preferences of the entire T2DM patient population with respect to lifestyle programs.

In conclusion, when deciding to participate in a lifestyle program, T2DM patients in the age group of 35-65 years are mostly driven by the out-of-pocket costs of a lifestyle program, the structure of the consultation and the expected outcome of the program. We therefore advise that lifestyle programs directed at T2DM patients should be set up based on individual consultation, while communicating about the expected outcomes of the program (in terms of weight loss) and keeping out-of-pocket costs as limited as possible.

References

1. Knowler, W.C., et al., *Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin.* N Engl J Med, 2002. 346(6): p. 393-403.
2. Wing, R.R., *Long-term effects of a lifestyle intervention on weight and cardiovascular risk factors in individuals with type 2 diabetes mellitus: four-year results of the Look AHEAD trial.* Arch Intern Med, 2010. 170(17): p. 1566-75.
3. Diabetes Prevention Program Research Group, *The Diabetes Prevention Program (DPP): description of lifestyle intervention.* Diabetes Care, 2002. 25(12): p. 2165-71.
4. Lindstrom, J., et al., *Prevention of diabetes mellitus in subjects with impaired glucose tolerance in the Finnish Diabetes Prevention Study: results from a randomized clinical trial.* J Am Soc Nephrol, 2003. 14(7 Suppl 2): p. S108-13.
5. Aucott, L., et al., *Effects of lifestyle interventions and long-term weight loss on lipid outcomes - a systematic review.* Obes Rev, 2011. 12(5): p. 412-425.
6. Shaw, J.E., R.A. Sicree, and P.Z. Zimmet, *Global estimates of the prevalence of diabetes for 2010 and 2030.* Diabetes Res Clin Pract, 2010. 87(1): p. 4-14.
7. Wild, S., et al., *Global prevalence of diabetes: estimates for the year 2000 and projections for 2030.* Diabetes Care, 2004. 27(5): p. 1047-53.
8. American Diabetes Association, *Standards of medical care in diabetes: 2010.* Diabetes Care, 2010. 33 Suppl 1: p. S11-61.
9. Paulweber, B., et al., *A European evidence-based guideline for the prevention of type 2 diabetes.* Horm Metab Res, 2010. 42 Suppl 1: p. S3-36.
10. IDF Clinical Guidelines Task Force, *Global guideline for Type 2 diabetes*. 2005, Brussels: International diabetes Federation.
11. James, D.V.B., et al., *Factors associated with physical activity referral uptake and participation.* Journal of Sports Sciences, 2007. 26(2): p. 217-24.
12. Bouchard, D.R., et al., *Age differences in expectations and readiness regarding lifestyle modifications in individuals at high risk of diabetes.* Arch Phys Med Rehabil, 2012. 93(6): p. 1059-64.
13. Clarke, P.M., et al., *A model to estimate the lifestime health outcomes of patients with Type 2 diabetes: the United Kingdom Prospective Diabetes Study (UKPDS) Outcomes Model (UKPDS no. 68).* Daibetologica, 2004. 47: p. 1747-59.
14. Davies, T.M.E., et al., *U.K. Prospective Diabetes Study 22; Effect of age at diagnosis on diabetes tissue damage during the first 6 years of NIDDM.* Diabetes Care, 1997. 20(9): p. 1435-41.
15. van den Akker, M., et al., *Multimorbidity in general practice: prevalence, incidence, and determinants of co-occurring chronic and recurrent diseases.* J Clin Epidemiol, 1998. 51(5): p. 367-375.
16. van Oostrom, S.H., et al., *Multimorbidity and comorbidity in the Dutch population - data from general practices.* BMC Public Health., 2012. 12(715).
17. Dube, M.C., et al., *Physical activity barriers in diabetes: development and validation of a new scale.* Diabetes Res Clin Pract, 2006. 72(1): p. 20-7.
18. Forbes, C.C., et al., *Physical activity preferences and type 2 diabetes: exploring demographic, cognitive, and behavioral differences.* Diabetes Educ, 2010. 36(5): p. 801-15.
19. Lakerveld, J., et al., *Motives for (not) participating in a lifestyle intervention trial.* BMC Med Res Methodol, 2008. 8: p. 17.
20. Thomas, N., E. Alder, and G.P. Leese, *Barriers to physical activity in patients with diabetes.* Postgrad Med J, 2004. 80(943): p. 287-91.
21. Vijan, S., et al., *Barriers to following dietary recommendations in Type 2 diabetes.* Diabet Med, 2005. 22(1): p. 32-8.
22. Jendle, J., et al., *Willingness to pay for health improvements associated with anti-diabetes treatments for people with type 2 diabetes.* Curr Med Res Opin, 2010. 26(4): p. 917-23.
23. Johnson, F.R., et al., *High-risk individuals' willingness to pay for diabetes risk-reduction programs.* Diabetes Care, 2006. 29(6): p. 1351-6.
24. Owen, K., et al., *Individual preferences for diet and exercise programmes: changes over a lifestyle intervention and their link with outcomes.* Public Health Nutr, 2010. 13(2): p. 245-52.

25. Roux, L., et al., *Valuing the benefits of weight loss programs: an application of the discrete choice experiment*. Obes Res, 2004. 12(8): p. 1342-51.
26. Struijs, J.N. and C.A. Baan, *Integrating care through bundled payments--lessons from The Netherlands.* N Engl J Med, 2011. 364(11): p. 990-1.
27. Lancsar, E. and J. Louviere, *Conducting discrete choice experiments to inform healthcare decision making: a user's guide.* Pharmacoeconomics, 2008. 26(8): p. 661-77.
28. Ryan, M., K. Gerard, and M. Amaya-Amaya, *Using Discrete Choice Experiments to Value Health and Health Care*. The Economics of Non-Market Goods and Resources, ed. I.J. Bateman. 2008, Dordrecht: Springer.
29. Krueger, R.A. and M.A. Casey, *Focus groups; A practical guide for applied research*. 2000, Thousand Oaks: Sage publications, Inc.
30. Bliemer, M.C.J. and J.M. Rose, *Efficiency and sample size requirements for stated choice experiments*, in *Transportaion Research Broad Annual Meeting*. 2009: Washington DC.
31. Huber, J. and K. Zwerina, *The Importance of Utility Balance in Efficient Choice Designs.* Journal of marketing research, 1996. 33(3): p. 307-317.
32. Marshall, D., et al., *Conjoint Analysis Applications in Health - How are Studies being Designed and Reported?: An Update on Current Practice in the Published Literature between 2005 and 2008.* Patient., 2010. 3(4): p. 249-56.
33. Louviere, J.J., D.A. Hensher, and J.D. Swait, *Stated Choice Methods; analysis and apllication*. 2000, Cambridge: Cambridge University Press.
34. Bech, M. and D. Gyrd-Hansen, *Effects coding in discrete choice experiments.* Health Econ, 2005. 14(10): p. 1079-83.
35. Revelt, D. and K.E. Train, *Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level.* The Review of Economics and Statistics, 1998. 80(4): p. 647-57.
36. Johnson, F.R., et al., *How does cost matter in health-care discrete-choice experiments?* Health Econ, 2011. 20(3): p. 323-30.
37. Resnicow, K., et al., *Tailoring a fruit and vegetable intervention on novel motivational constructs: results of a randomized study.* Ann Behav Med., 2008. 35(2): p. 159-69.
38. Williams, D.M., et al., *Does tailoring on additional theoretical constructs enhance the efficacy of a print-based physical activity promotion intervention?* Health Psychol., 2011. 30(4): p. 432-41.
39. Beulens, J.W., et al., *Cohort profile: the EPIC-NL study.* Int J Epidemiol., 2010. 39(5): p. 1170-8.

ADDITIONAL FILE 1:

Detailed description of the additional questionnaire

The questionnaire consisted of two parts. The first section of the questionnaire comprised 28 questions starting with questions concerning the participant's demographics, such as gender, age and educational level. Educational levels were categorized into low (primary school or lower general secondary education), medium (higher secondary education or lower general professional education) and high (i.e., bachelor or master's degree) educational level. Thereafter questions pertained to the patient's disease status: type of diabetes, disease duration since diagnosis, primary provider of care, use of medication, latest measured HbA1c level, body height and weight, presence of complications or other chronic illnesses and self-management activities (i.e., self-measurement of HbA1c levels and keeping a diabetes diary). This section of the questionnaire ends with the EuroQol-5D health status questionnaire (EQ-5D) [1] and the patient's attitude, social norm, self-efficacy and intention with respect to lifestyle programs in general (answers were given on a 5-point Likert scale). The second part of the questionnaire consists of the actual DCE as explained in the paper.

References

1. Lamers LM, Stalmeier PF, McDonnell J, Krabbe PF, van Busschbach JJ. *[Measuring the quality of life in economic evaluations: the Dutch EQ-5D tariff]*. Ned Tijdschr Geneeskd 2005; 149: 1574-8.

ADDITIONAL FILE 2:

Detailed description of the statistical methods

NLogit 4.0 (Econometric Software, 2007) was used to construct the panel-mixed-logit (Panel-MIXL) models that were estimated within this study. When using such models, results are adjusted for the panel structure (i.e., multilevel structure) of the data. As every respondent completed nine choice tasks, their answers may be correlated, which is then accounted for. Equation 1.1 was tested using these models.

[Eq. 1.1]

$V = \beta_0 + \beta_1$ * flexible meal plan + β_2 * elaborate meal plan + β_3 * flexible PA schedule + β_4 * elaborate PA schedule + β_5 * consultation in groups of 5 + β_6 * consultation in groups of 10 + β_7 * expected outcome + β_8 * out-of-pocket costs

V describes the utility of a specific lifestyle program based on the attributes that were included in the DCE. β_0 represents the alternative specific constant and $\beta_1 - \beta_8$ are the attribute level estimates that indicate the relative importance of each attribute. Interactions between the attribute estimates and demographic and background characteristics were examined to determine whether there were specific subgroups that reported different preferences. As no such subgroups could be identified, the analyses were not adjusted for demographic or other background characteristics.

All attributes were tested for linearity, all parameters that were found to be non-linear were recoded into effect codes. In contrast to dummy coding, this coding procedure codes the reference category -1, so the sum of the effect coded attributes is always 0. The coefficient for the reference category is therefore $-1*(\beta_{\text{effect code 1}} + \beta_{\text{effect code 2}})$. This makes it possible to compare the estimates of all attributes despite their categorization into effect codes [1,2] as opposed to dummy coding, which normally allows for comparison between the dummies only.

Based on the model fit (AIC and Chi-square tests), the constant of the model was set as a random parameter and the expected outcome attribute (both with a normal distribution) was also set as a random parameter. The expected outcome of the program may possibly show high preference heterogeneity among the participants because of large differences in BMI between the respondents. The presence of preference heterogeneity does not necessarily indicate subgroups within a population (e.g., T2DM preference heterogeneity for the outcome attribute, does not automatically mean that patients with a high versus a low BMI choose substantially different preferences for all attributes).

In order to calculate the patient's marginal willingness- to-pay (WTP), the negative of the out-of-pocket attribute was used as a measure of the marginal utility of money. The ratio of either attribute estimate to this negative of the out-of-pocket cost attribute provided an estimation of the patient's WTP concerning that specific attribute [3,4]. Moreover, the potential participation rates (choice probabilities) of a program that consist of a specific set of attributes

were estimated. Out-of-pocket costs were not included in these analyses, as the influence of out-of-pocket payments was already covered by calculating the WTP ratios. As both the constant and the expected outcome attributes were included as random parameters in the analyses, choice probabilities could not be calculated directly. A simulation was used to estimate the choice probabilities, given a certain set of attributes [1,4]. The mean participation rates of all simulations (n=1000) was estimated by taking the average of all simulated participation rate probabilities, which were calculated as $1/(1+\exp^{-v})$.

References

1. Louviere JJ, Hensher DA, Swait JD: *Stated Choice Methods; analysis and application*. Cambridge: Cambridge University Press; 2000.
2. Bech M, Gyrd-Hansen D. *Effects coding in discrete choice experiments*. Health Econ 2005; 14: 1079-83.
3. Revelt D, Train KE. *Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level.* Rev Econ Statistic 1998; 80: 647-57.
4. Ryan M, Gerard K, Amaya-Amaya M: *Using Discrete Choice Experiments to Value Health and Health Care*. Dordrecht: Springer; 2008.

CHAPTER 2

PARENTAL PREFERENCES FOR ROTAVIRUS VACCINATION IN YOUNG CHILDREN

J. Veldwijk, M.S. Lambooij, P. Bruijning-Verhagen,
H.A. Smit, G.A. de Wit

Vaccine (2014), 32: 6277-83.

ABSTRACT

Objective: This study aimed to identify characteristics that affect parental decisions about rotavirus vaccination, to determine the relative importance of those characteristics and subsequently to estimate vaccination coverage for different implementation strategies.

Methods: A discrete choice experiment (DCE) questionnaire was sent to the parents of 1,250 newborns aged 6 weeks (response rate 37.3%). Mixed-logit models were used to estimate the relative importance of the five included rotavirus vaccine and implementation characteristics; vaccine effectiveness, frequency of severe side effects, protection duration, the healthcare facility that administrates vaccination and out-of-pocket costs. Based on the utility functions of the mixed-logit model, the potential vaccination coverage was estimated for different vaccine scenarios and implementation strategies.

Results: All characteristics, except for healthcare facility that administrates vaccination, influenced parental willingness to vaccinate their newborn against rotavirus. Parents were willing to trade 20.2 percentage points vaccine effectiveness for the lowest frequency of severe side effects (i.e. 1 in 1,000,000) or 20.8 percentage points for a higher protection duration. Potential vaccination coverage ranged between 22.7 and 86.2%, depending on the vaccine scenario (i.e. vaccine effectiveness and protection duration) and the implementation strategy (i.e. out-of-pocket costs and healthcare facility that administrates vaccination).

Conclusions: When deciding about vaccination against rotavirus, parents are mostly driven by the out-of-pocket costs, vaccine effectiveness, protection duration and frequency of severe side effects. The highest vaccination coverage is expected for a vaccine with high effectiveness and protection duration that is implemented within the current National Immunization Program context. Implementation of the same rotavirus vaccine in the free market will result in lowest coverage.

INTRODUCTION

Although two rotavirus vaccines, RotaTeq and Rotarix, have been readily available since 2006, vaccination of newborns against rotavirus is not common practice in many countries, including the Netherlands. Rotavirus is a highly contagious pathogen that is the leading cause of gastroenteritis worldwide [1]. Especially in young children (under the age of two), rotavirus causes high morbidity [2, 3]. Worldwide about 95% of all unvaccinated children will experience at least one infection before the age of five [4]. Peak incidence is reported in children aged 4-26 months [4, 5]. Based on the results of a Dutch prospective cohort study it is estimated that one in every two unvaccinated Dutch children will suffer from a symptomatic rotavirus related gastroenteritis before the age of five [6, 7]. The currently licensed live-attenuated oral vaccines confer up to 94-96% protection against severe rotavirus related complaints and up to 60-80% protection against rotavirus disease of any severity [8-11]. Both vaccines are recommended for use in infants by the WHO and have been implemented in several countries (e.g., US, Australia, Belgium, Finland and recently the UK), where a substantial decrease in rotavirus related GP-visits and hospitalizations was observed [12].

In several countries, among which the Netherlands [13], debate is ongoing whether to implement this vaccine in the National Immunization Program (NIP). This is partly because of inconsistent results of research studying the cost-effectiveness of introducing this vaccine [14-17]. In countries that have not yet decided to include rotavirus vaccination in the NIP, the vaccines are available on the private market. Countries that consider adding rotavirus vaccination to their NIP in the (near) future, will benefit from greater knowledge on parental preferences with regard to vaccine safety and efficacy profiles as well as different implementation strategies (i.e. how the vaccine will be administered to their child), and how these influence the willingness of parents to vaccinate their newborn against rotavirus. Such information can guide their decisions concerning the strategy they might use for the introduction of the vaccine, and may serve as a starting point for efficient communication when the vaccine is introduced. Moreover, in countries that already implemented rotavirus vaccination, this information might provide guidance for policy makers to optimize the vaccination regime.

The aim of this study was to determine parental preferences with regard to rotavirus vaccination for their newborn. A secondary aim was to determine the potential vaccination coverage for different vaccine scenarios (based on vaccine characteristics) and different ways to implement the vaccine (based on implementation characteristics such as vaccine costs).

MATERIALS AND METHODS

Discrete choice experiments (DCEs) are increasingly being used to determine the relative importance of different characteristics of interventions or medical treatment. DCEs may also be used to estimate participants' willingness to pay as well as to estimate potential participation rates (e.g., potential vaccination coverage) [18-20]. The random utility theory is the basis of this method, which assumes that any intervention or treatment can be described by its characteristics or 'attributes' (such as vaccine effectiveness). Individuals' preference for an intervention or treatment

is determined based on the levels (e.g., effectiveness of 50% versus 80% versus 95%) of those attributes [18-20]. Scenarios are constructed by varying the levels of the attributes. Respondents are provided with a series of 'choice tasks' that consist of at least two scenarios. They have to choose the scenario they prefer most within every choice task.

Attributes, levels and DCE design

To construct the current DCE, possible attributes and levels were identified from previously published literature [21-28], expert interviews (i.e. pediatrician with a specific interest in rotavirus infections and a scientist with a specific interest in vaccination behavior), and four group interviews with a total of 25 parents of newborns. These group interviews were conducted using the nominal group technique (NGT) [29], which entails a ranking system as opposed to regular focus group techniques. Finally, five attributes were selected for this DCE (Table 1).

Table 1. Attributes and levels that were included in this DCE

Attributes	Level 1	Level 2	Level 3
Vaccine effectiveness			
The percentage of children that will be protected against a rotavirus infection when vaccinated			
	55%	75%	95%
Frequency of severe side effects			
The number of vaccinated children that will suffer from intussusception due to vaccination. Intussusception is an acute condition in which part of the bowel telescopes into another adjacent part of the bowel, resulting in obstruction [1].			
	1 in 10,000	1 in 100,000	1 in 1,000,000
Protection duration			
The number of years that the vaccine protects against a rotavirus infection			
	1 year	3 years	6 years
Healthcare facility of vaccine administration			
Within the Netherlands all vaccines in the NIP are administrated at a child welfare center, The GP office was included because the rotavirus vaccine may not become part of the NIP; in that case it is likely that this vaccine is administrated here.			
	Child welfare center	General practitioner	
Out-of-pocket costs			
Parents may have to pay (part) of the vaccine costs out-of-pocket			
	€0	€30	€140

1. Murphy TV, Gargiullo PM, Massoudi MS, et al. (2001) Intussusception among infants given an oral rotavirus vaccine. N Engl J Med 344: 564-72.

Using Ngene 1.0 (ChoiceMetrics, 2011) software, a D-efficient design was developed [30, 31], which minimizes the sample size and the number of choice tasks every respondent is asked to complete based on optimizing the variance-covariance matrix. The DCE consisted of 18 unique choice tasks. These were divided into two sets of nine choice tasks, and each set of nine choice tasks

was randomly distributed among half of the study population. Before participants were asked to complete these choice tasks, they received detailed information on the meaning of all attributes and levels as well as an explanation on how to complete a choice task. Here it was explained very clearly to parents that protection against a rotavirus infection is of specific importance when children are younger than two years of age. An example of a choice task of the current study is displayed in Table 2.

Table 2. Example choice task

Imagine that there is a vaccine available against rotavirus infection. In which scenario would you prefer to vaccinate your newborn, scenario 1 or scenario 2?

	Scenario 1	**Scenario 2**
Vaccine effectiveness	95%	75%
Frequency of severe side effects	1 in 1.000.000	1 in 100.000
Protection duration	3 years	6 years
Healthcare facility	General Practitioner	General Practitioner
Out-of-pocket costs	€ 30	€ 0
Tick the box of the scenario that you prefer:	☐	☐

Would you make this same choice in real life?

☐ Yes, if the scenario I chose above would become available, I would vaccinate my child against the rotavirus

☐ No, if the scenario I chose above would become available, I would not vaccinate my child against the rotavirus

The draft questionnaire was pilot tested among a convenience subgroup (n=48) of our study population. Four of these pilot tests were 'think aloud' tests, during which a researcher was present when the participant completed the questionnaire, reading out loud. This pilot test showed that correct wording was used, the target population understood the attributes, levels and choice tasks and provided attribute level estimates that served as priors for the design of the final DCE questionnaire. Additionally, based on these pilot-test data, sample size calculations were performed to ensure that significant differences for each attribute could be detected at a 5% level.

Questionnaire

The questionnaire consisted of two parts. The first section of the questionnaire comprised of 30 questions on demographics such as gender, age, educational level, and ethnicity. Thereafter, questions pertained to information on children's siblings, health status, parental view on the Dutch NIP, but also to their attitude, social norm, self-efficacy, perceived severity of and perceived susceptibility fr rotavirus infection (by presenting parents with different theorems). The second part of the questionnaire consisted of the actual DCE as explained above.

Study population

It is advised to administer the first dose of the rotavirus vaccine when newborns 6 weeks of age,

which is around the same age as during which parents have to decide whether to participate in the regular Dutch National Immunization Program (NIP). Since these parents are in the decision-making phase about vaccination according to the NIP, they were selected as the target population for this study. The target population was identified via Praeventis, which is a national vaccination register that registers the vaccination status of all Dutch newborns. A random sample of the parents of 1250 newborn babies aged 6 weeks was selected to receive a questionnaire. Due to confidentiality agreements with Praeventis, no reminder letters could be sent. For this reason no non-response information was available. The Institutional Review Board of the University Medical Center Utrecht advised that formal testing by a medical ethical committee was not necessary, as parents were only required to complete an anonymous questionnaire once, which is in accordance with the guidelines laid down in the Declaration of Helsinki.

Statistical analysis

Data were analyzed using panel mixed-logit (MIXL) models to account for preference heterogeneity and to adjust for the multilevel structure of the data. Respondents with >10% missing answers on their choice tasks were excluded from the analysis (n=12). All attributes were tested for linearity, afterwards non-linear attributes were recoded using effect codes [32], which resulted in Equation 1.

[eq. 1]

$$U = V + \varepsilon = \beta_{0i} + \beta_1 * \text{vaccine effectiveness} + \beta_2 * \text{frequency of severe side effects}_{1 \text{ in } 10{,}000} + \beta_3 * \text{frequency of severe side effects}_{1 \text{ in } 100{,}000} + \beta_4 * \text{protection duration}_{3 \text{ years}} + \beta_5 * \text{protection duration}_{6 \text{ years}} + \beta_6 * \text{healthcare facility}_{\text{child welfare center}} + \beta_{7i} * \text{out-of-pocket cost}_{€30} + \beta_{8i} * \text{out-of-pocket cost}_{€140} + \varepsilon$$

U describes the utility of a rotavirus vaccine for the respondents. V can be calculated as the observed utility which is the sum of β_0 (the constant) and $\beta_1 - \beta_8$, which are the attribute level estimates that indicate the relative importance of each attribute. The ε-term describes the unmeasured variation in respondents' preferences [18, 20, 33]. Based on model fit tests (AIC and Chi-square), β_0, β_7 and β_8 were included as random parameters. By including random parameters, the model accounts for any heterogeneity in the preferences of the respondents concerning those attributes. Parents probably differ with respect to their initial preferences for vaccination compared to no vaccination and with respect to their willingness to pay for a vaccine. The presence of preference heterogeneity does not indicate subgroups within a population per se (e.g., heterogeneity for the cost attribute does not automatically mean that parents with different incomes choose substantially different for all attributes).

The potential vaccination coverage was determined for different realistic vaccine scenarios and implementation strategies, which consisted of varying levels of the included attributes as described in Table 1. Vaccine scenarios were varied with respect to vaccine effectiveness and protection duration, based on current evidence with respect to the actual effectiveness and protection duration of the available vaccines [10, 11]. All vaccine scenarios included that 1 in 100,000 children

suffer from severe side effects, because this is most realistic for both types of vaccines available [34, 35]. Four different implementation strategies were defined based on the healthcare facility that administrates vaccination and the amount of out-of-pocket costs. The 'NIP' implementation strategy automatically implies no out-of-pocket costs and administration at a child welfare center. A 'NIP+' strategy entails an implementation strategy where the vaccine is part of the NIP but requires an additional out-of-pocket payment of €30. The 'healthcare insurance' implementation strategy entails that the vaccine will not be part of the NIP, but healthcare insurance will pay (part) of the vaccine costs when parents decide to vaccinate their newborn against the rotavirus (i.e. leaving an additional out-of-pocket payment of €30). In this strategy, parents will have to go to their GP office to have their newborn vaccinated. The 'private market' implementation strategy implies the necessity to visit the GP office for administration and an out-of-pocket payment of €140. The potential coverage rates can be calculated as $1/(1+\exp^{-V})$. Since V includes random parameters, the standard deviation of those parameters should be accounted for [18, 33]. The value of the random parameters was determined by taking 10,000 draws from a normal distribution with a mean and standard deviation (SD) for that particular random parameter (i.e. the mean and SD values were retrieved from the MIXL model). For every draw of the random parameter value, the observed utility 'V' as well as the potential coverage rate was calculated. The average of the 10,000 calculated potential vaccination coverage rates was reported for every given vaccination scenario.

The willingness of parents to trade off vaccine effectiveness against a more preferred level of another attribute was estimated by calculating the ratio between the inverse estimate of the attribute concerned and the attribute estimate of vaccine effectiveness. Dividing $-(\beta_5-\beta_4)$ by β_1 (see Equation 1) results in the percentage of vaccine effectiveness that parents are willing to trade off for a switch from a protection duration of 6 years to a protection duration of 3 years.

Finally, by including covariates and interaction terms with all attributes, it was tested whether parental preferences differed based on perceived severity and susceptibility, and vaccination intention.

RESULTS

Participants' characteristics

In total 466 out of 1250 questionnaires were returned and included in the analyzes (response rate of 37.3%). The mean age of the respondents was 31 years and approximately 82% was female (Table 3). About 58% reported to have received higher education, 8% was of a non-Dutch origin, and 14% reported to adhere to a religion or belief system that influenced their thoughts about vaccination. In total, 58.6% of the parents had their first child, 29.5% had their second child, and 11.7% had three children or more. In total, 8% reported that they knew someone who had suffered from a rotavirus infection in the past. Of all parents with other (older) children, 90% reported that these children were vaccinated according to the Dutch NIP and 97% reported that they had good experiences with these vaccinations. Furthermore, the majority of the parents agreed with the theorems they were presented with (Table 3). Approximately 90% of all parents thought that vaccination

in general is useful and self-evident (Table 3). After having read the general information about a rotavirus infection, 62% of the parents reported that they considered a possible rotavirus infection of their child to be "very serious", while 64% thought their newborn could become seriously ill from such an infection. Of all parents, 24% thought their newborn would have a high chance to become infected with rotavirus. Finally, 77% considered vaccination a good way to protect a newborn against a rotavirus infection and 79% reported that they would vaccinate their newborn baby against the rotavirus if a vaccine would become available.

Table 3. Demographic characteristics of the study population and percentage of parents that agreed with different (rotavirus) vaccination-related theorems

		Mean (SD)	**Percentage (%)**
Age (n=466)		31.2 (5.1)	
Gender (n=466)	Female		81.5
Highest attained educational level (n=460)	Low		8.3
	Medium		33.7
	High		58.0
Ethnicity (n=466)	Dutch		91.8
Belief system (n=459)	None		86.1
	Religion		9.6
	Other		4.3
Number of children (n=464)	One		58.6
	Two		29.7
	Three or more		11.7
Knows someone with rotavirus infection (n=463)			8.4
General health of the newborn (n=466)	Good		97.9
Parents with another child older than the newborn (n=228)			
Older child gets the NIP vaccinations			89.9
Opinion about the Dutch NIP based on experience	Good		96.9
Theorems			
I think vaccination is useful			90.8
I think vaccination is self-evident			89.5
I think vaccinating my newborn is easy			88.4
I think it is serious if my child gets infected with rotavirus			61.8
I think that my child gets seriously ill if infected with the rotavirus			63.5
I think the chance that my child gets infected with the rotavirus is large			23.8
I think vaccination is a good way to protect my child against a rotavirus infection			76.6
I think my family and friends would vaccinate their children against rotavirus			72.6
I think my family and friends think it is important to vaccinate my child against rotavirus			48.9
I think I would get my child vaccinated against rotavirus if a vaccine would be available			79.4

Parental preferences

All attributes, except for the healthcare facility that administrates vaccination, showed a significant estimate, indicating they individually contributed to the decision process of parents about whether to vaccinate their newborn against the rotavirus (Table 4; Basic model). Parents preferred high vaccine effectiveness and a frequency of 1 in 1,000,000 children who suffer from severe side effects over any other frequency of severe side effects (i.e. 1 in 10,000 or 1 in 100,000). A protection duration of three and six years was preferred over one year, however parents preferred three years of protection over a protection duration of six years. Finally, the higher the out-of-pocket costs the lower was the willingness to vaccinate.

With respect to the relative importance of these attributes, out-of-pocket costs were most decisive for parents in their decision about vaccination. This was followed by vaccine effectiveness, protection duration and frequency of severe side effects.

Table 4. Parental preferences for rotavirus vaccination

			Basic model		**Model 1±**		**Model 2±**	
Attribute			**Estimate**	**SE**	**Estimate**	**SE**	**Estimate**	**SE**
Constant		Mean	-3.84***	0.24	-3.91***	0.25	-8.50***	0.33
		SD	2.91***	0.13	2.88***	0.12	2.37***	0.12
Vaccine effectiveness			0.66***	0.03	0.64***	0.03	0.67***	0.03
Frequency of severe side effects	1 in 1,000,000 (ref)		0.71***	0.06	0.71***	0.06	0.73***	0.06
	1 in 100,000		-0.08*	0.05	-0.08*	0.05	-0.08*	0.05
	1 in 10,000		-0.63***	0.06	-0.63***	0.06	-0.65***	0.06
Protection duration	1 year (ref)		-0.84***	0.05	-0.84***	0.05	-0.85***	0.05
	3 years		0.53***	0.06	0.54***	0.06	0.54***	0.06
	6 years		0.31***	0.05	0.30***	0.05	0.31***	0.05
Healthcare facility	General practitioner (ref)		-0.02	0.03	-0.02	0.03	-0.02	0.03
	Child Welfare Center		0.02	0.03	0.02	0.03	0.02	0.03
Out-of-pocket costs	0 Euro (ref)	Mean	0.89***	0.07	0.89***	0.07	0.89***	0.06
		SD	0.88	0.89	1.18*	0.97	0.88	1.01
	30 Euro	Mean	0.57***	0.06	0.57***	0.06	0.60***	0.06
		SD	0.21**	0.10	0.16	0.11	0.10	0.12
	140 Euro	Mean	-1.46***	0.07	-1.46***	0.07	-1.49***	0.07
		SD	0.85***	0.06	0.85***	0.06	0.87***	0.06
Severe & susceptible					0.49	0.54		
Severe & susceptible * vaccine effectiveness					0.10**	0.05		
Intention							5.42***	0.31

± only the significant interaction terms were included in this table * p <0.10; ** p <0.05; *** p <0.001

Potential vaccination coverage

Table 5 shows the estimates of potential vaccination coverage for different vaccine scenarios and implementation strategies. If rotavirus vaccination would be implemented in an NIP context, the uptake probabilities ranged from 48.3% to 86.2%, depending on the effectiveness of the vaccine, protection duration and frequency of severe side effects. These estimated coverage rates were higher than the coverage rates for the same vaccination scenarios within any of the other three implementation strategies. The private market implementation strategy would, in any of the vaccine scenarios, results in the lowest coverage rates (difference with the NIP strategy for the most optimal scenario is 18.4 percentage points). Within all implementation strategies, similar trends were observed; estimates of vaccination coverage increased with increasing effectiveness and protection duration.

Table 5. Potential vaccination coverage for different vaccination scenarios stratified by implementation strategy

	Part of the NIP - €0 -Administered at CWC[#]	Part of NIP+ - €30 -Administered at CWC[#]	Health Care Insurance - €30 -Administered at GP[##]	Private Market - €140 -Administered at GP[##]
Vaccine effectiveness 55%				
1 year protection Severe side effects frequency of 1 in 100,000	48.3	45.2	43.3	22.7
3 years protection Severe side effects frequency of 1 in 100,000	64.2	60.3	60.4	37.3
Vaccine effectiveness 75%				
1 year protection Severe side effects frequency of 1 in 100,000	63.8	59.5	59.3	36.2
3 years protection Severe side effects frequency of 1 in 100,000	75.7	74.8	74.2	52.0
Vaccine effectiveness 95%				
1 year protection Severe side effects frequency of 1 in 100,000	75.1	74.8	74.3	52.2
3 years protection Severe side effects frequency of 1 in 100,000	86.2	85.2	84.7	67.8

[#] Child Welfare Center; [##] General Practitioner

Trade-offs

Parents were willing to trade 20.2 percentage points (95% CI: 18.1 percentage points; 22.3 percentage points) vaccine effectiveness to decrease the probability of serious side effects from a frequency of 1 in 10,000 children to a frequency of 1 in 1,000,000 children. Moreover, parents were willing to trade approximately 20.8 percentage points vaccine effectiveness (95% CI: 18.97 percentage points; 22.7 percentage points) to prolong the protection duration from one year to three years.

Stratifications by perceived severity, susceptibility and intention

Parents who considered rotavirus to be a severe disease, and found their newborn to be susceptible to a rotavirus infection, had a significantly higher preference for vaccine effectiveness compared to parents who were less sure about the severity of an infection and their child's susceptibility to such an infection (Table 4; Model 1). Parents who stated that they intended to vaccinate their newborn against rotavirus when the vaccine would become available, also were more willing to vaccinate their newborn according to the DCE analyses. However, their preferences for the vaccine characteristics were no different from those of parents who stated that they did not intend to vaccinate their newborn (i.e. no significant interaction terms were identified) (Table 4; Model 2).

DISCUSSION

This study showed that high vaccine effectiveness, low frequency of severe side effects, a protection duration of three years, and lower out-of-pocket costs increased parents willingness to have their newborn vaccinated against rotavirus. Parents were willing to accept lower vaccine effectiveness if this would imply a lower frequency of severe side effects (1 in 1,000,000 instead of 1 in 10,000) or longer protection duration (i.e. three years instead of one year).

To our knowledge, no DCEs have been conducted before that investigated parental preferences with regard to rotavirus vaccination in newborns. Our findings were in line however, with results of the study by Morin (2012), who studied maternal attitudes and beliefs regarding the rotavirus vaccine in Canada using a Health Belief Model-based questionnaire [36]. They showed that higher perceived susceptibility to a severe rotavirus infection, lower out-of-pocket costs (operationalized as reimbursements) and higher vaccine effectiveness increased the intention of pregnant women to vaccinate their newborn [36]. Moreover, DCEs that were conducted to investigate preferences for other vaccines than a rotavirus vaccine also indicated that the willingness to vaccinate increased with increasing vaccine effectiveness, decreasing risk of (mild and severe) side effects, increasing protection duration and decreasing out-of-pocket costs [21-23, 25-28]. Although the topics of these studies differed slightly from the current study, these findings about parental preferences for vaccination of newborns do provide face validity for the current study results.

Our study shows that choosing a specific implementation strategy (i.e. either include in the current NIP, a NIP+ scenario, a healthcare insurance scenario or a private market scenario) directly influences

the expected vaccination coverage. Implementing the vaccine in the existing NIP is expected to result in the highest coverage as compared to the other implementation strategies. The actual coverage depends on characteristics of the type of vaccine that eventually will be introduced. Within all implementation strategies, the highest coverage rates are to be expected with the highest vaccine effectiveness, a protection duration of three years and the lowest frequency of severe side effects. The findings above are in line with current findings of coverage rates in European countries. High coverage rates (>90%) are only observed in countries that included the rotavirus vaccine in their NIP and (completely or largely) reimburse vaccination (i.e. Belgium, Finland, Austria) [37]. Coverage rates drop below 40% within all other countries where the rotavirus vaccine is not part of the NIP and/or vaccination is not funded but requires full out-of-pocket payments [37].

A recent paper by Weintraup and colleagues showed that the frequency of severe side effects could be as high as 1 in 20,000 [35]. This reported frequency of severe side effects is higher compared to the previously expected number of children that will suffer from intussusception due to rotavirus vaccination [34]. Since vaccination decisions depend heavily on the frequency of expected severe side effects, it is of crucial importance that such risks are clearly communicated to parents. Policy makers should make sure that parents understand the risks, and do not over- or underestimate them. Our coverage estimations combined with the trade-off results showed that, when vaccine effectiveness and protection duration levels were high, parents were still willing to vaccinate their newborn even if the frequency of severe side effects was 1 in 10,000. Therefore, it is recommended to combine the information about the frequency of severe side effects with clear messages about the effectiveness of the vaccine and the protection duration after vaccination.

Interestingly, the highest coverage rates that were estimated for the NIP implementation strategy are lower than the coverage rates as currently observed with existing vaccines administered as part of the Dutch NIP (approximately 95% [38]). This could be due to a more critical view of parents concerning this vaccine compared to the vaccines already available. Rotavirus infections are less known in the general population and may therefore be perceived as less invasive, life threatening or damaging compared to an infection with pathogens of the diseases for which a vaccine already exist within the NIP (e.g., polio, diphtheria, tetanus, pertussis, hepatitis B). A Dutch survey among 491 parents with at least one child aged 0-4 years, confirmed this hypothesis. Parents were asked to rate the perceived severity of different NIP diseases as well as a rotavirus infection on a scale of 1 (i.e. not serious at all) to 10 (i.e. very serious). While a rotavirus infection scored 7.2, other diseases were perceived far more severe; polio (9.1), hepatitis B (8.5), tetanus (8.2) diphtheria (8.2), pertussis (7.6) (unpublished results). The current study showed that parents who perceived rotavirus as a severe disease and found their child to be susceptible to an infection were more likely to participate in a vaccination program, specifically if the effectiveness of the vaccine increases. We recommend communicating clearly about the severity of and susceptibility to a rotavirus infection if such a vaccine would be implemented, because this is crucial for an optimal vaccination coverage rate.

This study was subject to some limitations. First, although a response rate of 37.3% is relatively good for a postal questionnaire without reminders, selective non-response seems plausible. Compared to Dutch national population figures, the number of non-Dutch parents in our study population is relatively low. Generalizability of our results, especially the coverage rate estimates, to non-Dutch parents therefore may be limited. Additionally, the current study sample was slightly higher educated compared to the entire target population. However, two recent studies revealed that educational level was not associated with vaccination intention of Dutch parents [39,40]. We therefore have no reason to assume that the results would be substantially different if lower educated parents were included. Finally, parents with a newborn who suffers from a serious disease or other health problems early in life were underrepresented in this study. Their preference for vaccination might differ from the preference of the average parent. Future research specifically investigating the preferences for vaccination in this high risk group would be worthwhile, since those children are more susceptible to the severe consequences of a rotavirus infection [6].

When deciding whether to vaccinate their newborn baby against rotavirus, parents are mostly driven by out-of-pocket costs, the effectiveness of the vaccine, protection duration and the frequency of severe side effects. For a constant frequency of severe side effects, the highest vaccination coverage was expected for a vaccine with high effectiveness and protection duration that is implemented within the current NIP context (implying no out-of-pocket costs and administration at a child welfare center). Implementation of a rotavirus vaccine in the free market will result in lowest coverage rates, with a maximal difference between this and the NIP implementation strategy of 18.4 percentage points.

References

1. Parashar, U.D., et al., *Global illness and deaths caused by rotavirus disease in children.* Emerg Infect Dis., 2003. 9(5): p. 565-72.
2. Parashar, U.D., et al., *Global mortality associated with rotavirus disease among children in 2004.* J Infect Dis, 2009. 200 Suppl 1: p. S9-S15.
3. Tate, J.E., et al., *2008 estimate of worldwide rotavirus-associated mortality in children younger than 5 years before the introduction of universal rotavirus vaccination programmes: a systematic review and meta-analysis.* Lancet Infect Dis, 2012. 12(2): p. 136-41.
4. Velazquez, F.R., et al., *Rotavirus infections in infants as protection against subsequent infections.* N Engl J Med, 1996. 335(14): p. 1022-8.
5. Cortese, M.M. and U.D. Parashar, *Prevention of rotavirus gastroenteritis among infants and children: recommendations of the Advisory Committee on Immunization Practices (ACIP).* MMWR Recomm Rep, 2009. 58(RR-2): p. 1-25.
6. Bruijning-Verhagen, P., et al., *Targeted rotavirus vaccination of high-risk infants; a low cost and highly cost-effective alternative to universal vaccination.* BMC Med., 2013. 11:112.
7. Friesema, I.H., et al., *Etiology of acute gastroenteritis in children requiring hospitalization in the Netherlands.* Eur J Clin Microbiol Infect Dis., 2012. 31(4): p. 405-15.
8. Ruiz-Palacios, G.M., et al., *Safety and efficacy of an attenuated vaccine against severe rotavirus gastroenteritis.* N Engl J Med, 2006. 354(1): p. 11-22.
9. Vesikari, T., et al., *Safety and efficacy of a pentavalent human-bovine (WC3) reassortant rotavirus vaccine.* N Engl J Med, 2006. 354(1): p. 23-33.
10. Soares-Weiser, K., et al., *Vaccines for preventing rotavirus diarrhoea: vaccines in use.* Cochrane Database Syst Rev, 2012. 11: p. CD008521.
11. Rha, B., et al., *Effectiveness and impact of rotavirus vaccines in the United States - 2006-2012.* Expert Rev Vaccines, 2014.
12. Patel, M.M., et al., *Fulfilling the promise of rotavirus vaccines: how far have we come since licensure?* Lancet Infect Dis, 2012. 12(7): p. 561-70.
13. Health Council of the Netherlands, *The individual, collective and public importance of vaccination.* 2013, Health Council of the Netherlands: The Hague.
14. Bilcke, J. and P. Beutels, *Reviewing the cost effectiveness of rotavirus vaccination: the importance of uncertainty in the choice of data sources.* PharmacoEconomics, 2009. 27(4): p. 281-97.
15. Jit, M., et al., *An update to "The cost-effectiveness of rotavirus vaccination: comparative analyses for five European countries and transferability in Europe".* Vaccine, 2010. 28(47): p. 7457-9.
16. Tu, H.A., et al., *An update of "Cost-effectiveness of rotavirus vaccination in the Netherlands: the results of a Consensus Rotavirus Vaccine model".* BMC Infect Dis., 2013. 13(54).
17. Widdowson, M.A., et al., *Cost-effectiveness and potential impact of rotavirus vaccination in the United States.* Pediatrics, 2007. 119(4): p. 684-97.
18. Ryan, M., K. Gerard, and M. Amaya-Amaya, *Using Discrete Choice Experiments to Value Health and Health Care.* The Economics of Non-Market Goods and Resources, ed. I.J. Bateman. 2008, Dordrecht: Springer.
19. Lancsar, E. and J. Louviere, *Conducting discrete choice experiments to inform healthcare decision making: a user's guide.* Pharmacoeconomics, 2008. 26(8): p. 661-77.
20. de Bekker-Grob, E.W., M. Ryan, and K. Gerard, *Discrete choice experiments in health economics: a review of the literature.* Health Econ., 2012. 21(2): p. 145-72.
21. Bishai, D., et al., *Conjoint analysis of French and German parents' willingness to pay for meningococcal vaccine.* PharmacoEconomics, 2007. 25(2): p. 143-54.
22. Brown, D.S., et al., *Mothers' preferences and willingness to pay for vaccinating daughters against human papillomavirus.* Vaccine., 2010. 28(7): p. 1702-8.
23. de Bekker-Grob, E.W., et al., *Girls' preferences for HPV vaccination: a discrete choice experiment.* Vaccine, 2010. 28(41): p. 6692-7.
24. Flood, E.M., et al., *A survey of children's preferences for influenza vaccine attributes.* Vaccine., 2011. 29(26): p. 4334-40.

25. Flood, E.M., et al., *Parent preferences for pediatric influenza vaccine attributes.* Clin Pediatr (Phila). 2011. 50(4): p. 338-47.
26. Hall, J., et al., *Using stated preference discrete choice modelling to evaluate the introduction of varicella vaccination.* Health Econ, 2002. 11(5): p. 457-65.
27. Poulos, C., et al., *Mothers' preferences and willingness to pay for HPV vaccines in Vinh Long Province, Vietnam.* Soc Sci Med., 2011. 73(2): p. 226-34.
28. Stockwell, M.S., et al., *The effects of vaccine characteristics on adult women's attitudes about vaccination: a conjoint analysis study.* Vaccine., 2011. 29(27): p. 4507-11.
29. Hiligsmann, M., et al., *Nominal group technique to select attributes for discrete choice experiments: an example for drug treatment choice in osteoporosis.* Patient Prefer Adherence, 2013. 7:133-9.
30. Bliemer, M.C.J. and J.M. Rose, *Efficiency and sample size requirements for stated choice experiments, in Transportaion Research Broad Annual Meeting.* 2009: Washington DC.
31. Huber, J. and K. Zwerina, *The Importance of Utility Balance in Efficient Choice Designs.* Journal of marketing research, 1996. 33(3): p. 307-317.
32. Bech, M. and D. Gyrd-Hansen, *Effects coding in discrete choice experiments.* Health Econ, 2005. 14(10): p. 1079-83.
33. Louviere, J.J., D.A. Hensher, and J.D. Swait, *Stated Choice Methods; analysis and apllication.* 2000, Cambridge: Cambridge University Press.
34. Tate, J.E., et al., *Research priorities regarding rotavirus vaccine and intussusception: a meeting summary.* Vaccine, 2012. 30 Suppl 1: p. A179-84.
35. Weintraub, E.S., et al., *Risk of intussusception after monovalent rotavirus vaccination.* N Engl J Med., 2014. 370(6): p. 513-9.
36. Morin, A., et al., *Maternal knowledge, attitudes and beliefs regarding gastroenteritis and rotavirus vaccine before implementing vaccination program: Which key messages in light of a new immunization program?* Vaccine, 2012. 30: p. 5921-5927.
37. Parez, N., et al., *Rotavirus vaccination in Europe: drivers and barriers.* Lancet Infect Dis., 2014. 14(5): p. 416-25.
38. van Lier, E.A., et al., *Vaccinatiegraad Rijksvaccinatieprogramma Nederland [Vaccination coverage National Immunization Program the Netherlands].* 2011, National Institute for Public Health and the Environment (RIVM). Report no.:210021014/2011: Bilthoven.
39. Harmsen, I.A., et al., *Psychosocial determinants of parents' intention to vaccinate their newborn child against hepatitis B.* Vaccine, 2012. 30: p. 4771-4777.
40. Paulussen, T.G.W., et al., *Determinants of Dutch parents' decisions to vaccinate their chils.* Vaccine, 2006. 24: p. 644-651.

CHAPTER 3

PREFERENCES FOR GENETIC SCREENING FOR COLORECTAL CANCER WITHIN A POPULATION-BASED SCREENING PROGRAM

J. Veldwijk, M.S. Lambooij, F.G.J. Kallenberg,
H. van Kranen, A.L. Bredenoord, E. Dekker,
H.A. Smit, G.A. de Wit

Submitted.

ABSTRACT

Objective: This study explored individuals' preferences for genetic screening for colorectal cancer (CRC) and their willingness to participate in genetic screening for Lynch syndrome, familial adenomatous polyposis (FAP), and familial colorectal cancer (FCC).

Methods: 532 respondents aged 55-65 completed a Discrete Choice Experiment (DCE). Using panel latent class models the preferences for two screening situation characteristics (the probability of being genetically predisposed and the probability of developing CRC) and screening test characteristics (the frequency of preventive colonoscopies and CRC survival) were estimated. Based on these preferences, respondents' willingness to participate in the three screening initiatives was estimated.

Results: Lower educated respondents and respondents who express serious anxiety and worries, found colonoscopy frequency and the probability of developing CRC relatively more important and survival relatively less important compared to higher educated respondents and respondents who express no anxiety and worries. These differences in preferences resulted in opposite preferences for participation in FCC and FAP screening.

Conclusions: The general population is willing to participate in genetic screening for CRC. If individuals are suspected of genetic or familial CRC, they should at least be informed about their increased risk of being genetically predisposed, and about the importance of participating in all preventive follow-up colonoscopies in order to maximize survival.

INTRODUCTION

Although genetic screening, in addition to population-based colorectal cancer (CRC) screening, may be beneficial for those who run a higher risk of developing CRC, there is a discussion about whether this additional form of screening is advisable and desirable [1-4]. CRC is one of the most commonly diagnosed cancers and the leading cause of death among all cancer types worldwide [5]. Prognosis, treatment intensity and the 5-year survival rate significantly improve if CRC is diagnosed at an early stage [6, 7]. Moreover, CRC can actually be prevented because it is usually preceded by a slow progressive premalignant lesion (an adenomatous polyp) which may become cancer but can be detected and removed during colonoscopy [7]. Therefore, population-based screening programs for CRC are recommended and widely implemented in Western countries. Within these programs, there is little attention for genetically predisposed individuals who run a higher risk of developing CRC. About 5% of all diagnosed CRC's is of genetic origin [8-10]. This relatively small percentage actually reflects a substantial number of CRC patients given the high incidence of CRC in the general population. Genetically predisposed individuals and their families can be identified by offering genetic screening to participants in a population-based CRC screening program with a positive colonoscopy and/or a familial cancer history [11, 12]. By including genetic screening in current population-based CRC screening programs, CRC-related morbidity and mortality may further decrease due to increased surveillance of cases and their relatives [11-13].

However, genetic testing raises several ethical and counseling challenges [14, 15]. Positive test outcomes may have serious psychological and social consequences [14, 16]. For instance, knowing that one is at risk to develop cancer might induce fear of actually developing cancer, possibly with a negative impact on a person's quality of life [17, 18]. Positive test results may also have a severe impact on the family of the tested individual [16-18], as they themselves might run a higher risk of developing cancer as well. Moreover, the general population often holds unrealistic expectations about the accuracy with which genetic screening tests can predict future disease status [18, 19].

Despite these potential negative consequences, the general population shows great interest in genetic screening and has a positive attitude towards such screening initiatives [17, 20-22]. Previous research shows that individuals are willing to take part in genetic screening when the test aims to identify an increased risk for a monogenic form of a common disease, when adequate treatment and/or prevention options are available and when clinicians recommend screening [21, 23-25].

To date, no research has been conducted into studying the preferences of the general population for genetic screening for CRC specifically. Therefore, this study aims to explore individual preferences concerning genetic screening for CRC within a population-based CRC screening program. A further aim is to estimate whether individuals are willing to participate in genetic screening for (1) Lynch syndrome, (2) familial adenomatous polyposis (FAP), and (3) familial colorectal cancer (FCC).

MATERIALS AND METHODS

Discrete Choice Experiment (DCE)

DCEs are increasingly being used to determine an individual's preferences regarding different characteristics of interventions or medical treatments [26]. This method is based on the Random Utility Theory. This theory assumes that any intervention or treatment can be described by its characteristics or 'attributes', such as the probability of a positive test outcome. The preferences of an individual for an intervention or treatment is determined on the basis of the 'levels' of the attributes, such as 1%, 3% or 15% probability that the test outcome is positive [26]. Hypothetical scenarios are constructed by varying the levels of the attributes. Respondents are provided with a series of 'choice tasks' that consist of at least two scenarios. They are asked to choose the scenario they prefer most within every choice task.

DCE development

To construct the DCE used for this study, possible attributes were identified from previously published studies [21-24, 27], six expert interviews (i.e. a scientist with a specific interest in public health genomics, a scientist with a specific interest in ethics of genetics/genomics, a specialist in cancer genetics, and three medical specialists in gastroenterology), and five group interviews (n= 38) with the target population of men and women aged 55-65 years. These group interviews were conducted using the Nominal Group Technique (NGT) [28]. During these interviews, participants were asked to rank a number of potential attributes from most to least important, the mean group ranking of the attributes was then discussed in the group, after which participants could change their original individual ranking [28]. Finally, four attributes were selected for this DCE (Table 1). The levels that were used to describe the identified attributes were based on realistic numbers representing the three most common types of genetic and familial CRC; Lynch syndrome, FAP, and FCC. About 3% of all CRC patients are diagnosed with Lynch syndrome [4, 29-31]. Without surveillance, these patients have a 70% probability of developing CRC during their lifetime [4, 29-31]. Patients who are diagnosed with Lynch syndrome are offered a preventive colonoscopy every two years. On average, their 5-year survival rate is 92% if they are aware of their genetic predisposition and participate in biannual colonoscopies [4, 29-31]. FAP is present in 1% of all CRC patients [29-31]. The probability of developing CRC among these patients is 99% without surveillance and therefore they are advised to undergo an annual colonoscopy [29-31]. This results in a 5-year survival rate of 80% if CRC is discovered [29-31]. Finally, FCC is considered to be present in 15% of all CRC patients [29-31]. These patients an at least 15% probability of developing CRC based on the number and age of relatives with CRC. They are offered screening by means of a 5-yearly colonoscopy which increases their 5-year survival rate to 98% if CRC is found [29-31].

NGene 1.0 (ChoiceMetrics, 2011) software was used to develop a D-efficient design [26]. The DCE consisted of nine unique choice tasks each containing two scenarios. Following each choice task, participants were asked whether they would actually participate in the chosen scenario or not (i.e. opt-out). Before participants were asked to complete the choice tasks, they received detailed information on the meaning of all attributes and levels as well as an explanation on how to

Table 1. Attributes and levels that were included in this DCE

Attributes	Level 1	Level 2	Level 3
Probability of being genetically predisposed (genetic predisposition) The likelihood that you are genetically predisposed to develop colorectal cancer.	1% 1 out of every 100	3% 3 out of every 100	15% 15 out of every 100
Probability of developing CRC* (CRC risk) 5 out of every 100 (5%) Dutch individuals develop colorectal cancer. If you have a genetic predisposition to develop colorectal cancer and you do not participate in preventive colonoscopies, the likelihood that you will develop colorectal cancer is higher and varies between:	15% 15 out of every 100	70% 70 out of every 100	99% 99 out of every 100
Frequency of preventive colonoscopies (colonoscopy frequency) If the genetic test shows that you are genetically predisposed to develop colorectal cancer, you will be invited to participate in preventive colonoscopies. These colonoscopies are performed to prevent cancer from developing or to diagnose cancer in an early stage. These colonoscopies will be scheduled on a regular basis varying between:	Every year	Every 2 years	Every 5 years
Probability of surviving CRC (survival) 60 out of every 100 (60%) Dutch individuals with colorectal cancer survive over the next 5 years. If you know you are genetically predisposed to develop colorectal cancer and if you participate in the preventive colonoscopies the likelihood that you will survive colorectal cancer over the next 5 years will increase and varies between:	80% 80 out of every 100	92% 92 out of every 100	98% 98 out of every 100

* CRC = colorectal cancer

complete a choice task, illustrated by an example (see additional file 1). The draft questionnaire was pilot tested among a subgroup (n=90) of our target population. Four of these pilot-tests were 'think aloud' tests, during which a researcher was present when the participant completed the questionnaire, reading out loud. It was tested by means of this pilot whether correct wording was used and whether the target population understood the attributes, levels and choice tasks. Additionally, the attribute level estimates that were retrieved from the pilot study served as input for the design of the final DCE questionnaire.

Questionnaire

The final questionnaire consisted of three parts. The first section of the questionnaire comprised 25 questions on demographics such as gender, age, educational level, health literacy and ethnicity. Educational level was dichotomized into higher (i.e. tertiary education) or lower education (i.e. all other educational levels). Health literacy was measured by three validated Dutch questions of the Set of Brief Screening Questions (SBSQ-D) [32]. Participants scored these questions on a 5-point Likert scale, from zero to four. An average score of ≤2 indicates inadequate health literacy, while an average score >2 indicates adequate health literacy [32]. Furthermore, questions pertained to

information on experience with other national cancer screening programs, experience with genetic screening, and family cancer history. Respondents were asked to indicate to what extent they agreed or disagreed with several theorems about their attitude, social norm, self-efficacy and intention towards genetic screening for CRC. The second part of the questionnaire consisted of the actual DCE as explained above. The third part consisted of several theorems regarding the consequences of genetic testing, such as fear and worries, and on the possibility of incidental findings.

Study population

From 2014 onwards, all Dutch residents aged 55-75 years will receive a biannual invitation to participate in the national population-based screening program for CRC. Screening is carried out by means of the fecal immunochemical test (FIT). If the test result is positive, i.e. blood is detected in the stool, a colonoscopy will be planned and participants are asked to complete a family cancer history questionnaire. At present, it is expected that genetic screening for CRC might only become part of the Dutch CRC screening program for individuals with a positive colonoscopy and/or a familial cancer history.

Since it was expected that preferences for genetic screening for CRC are highly dependent on age and experience with CRC screening, individuals were eligible to participate in our study if they were 55-65 years of age and had not yet participated in the CRC screening program or one of the extensive pilot studies that preceded the decision to implement the Dutch population-based CRC screening. Respondents were recruited via an existing online panel of the general Dutch population. Respondents were selected to be representative for the entire target population with respect to age, gender and educational level. In total, 5500 individuals were invited to participate in this study and recruitment continued until at least 500 questionnaires were fully completed by a representative sample of the target population.

The Dutch Central Committee on Research involving Human Subjects concluded that formal testing by an Institutional Review Board was not necessary, as respondents were only required to complete an anonymous and non-invasive questionnaire once, which is in accordance with the Dutch legislation and guidelines laid down in the Declaration of Helsinki.

Statistical analysis

All results were considered statistically significant when $p<0.05$. All attributes were considered to be non-linear and were recoded using effect codes [26]. This coding procedure codes the reference category as -1 and the sum of the effect coded attribute levels is always zero.

Preferences for genetic screening for CRC

Nlogit 5.0 (econometric software) was used to conduct the panel latent class models for this study. Such models account for the multilevel structure of our data (i.e. every respondent answered nine choice tasks). Moreover, by means of such models, it can be determined whether preferences differ across unobserved subgroups of the population. This modelling procedure identifies whether there are 'classes' within the data based on respondents' answering patterns. Which respondents belong to what class is not assigned by researchers, but is latent. Each respondent has a certain probability to belong one of the identified classes. However, demographic characteristics can be incorporated into the modelling procedure, which provides insight in which respondents are more likely to belong to a certain class.

Based on model fit tests (AIC, Log likelihood) it was tested which model was most suitable for our data, and how many classes could be identified within the data. This resulted in a two-class model based on the utility equation displayed below. The utility component (V) describes the utility that respondent 'r' belonging to class 'c' reported for alternative 'a' in choice task 't'. β_0 represents the constant of the model. The attribute level estimates that indicate the relative importance of each attribute level are represented by $\beta_1 - \beta_8$. A significant attribute estimate within a certain class indicates that this attribute contributes to the decision-making process of respondents who belong to that class.

$$V_{rta|c} = \beta_{0|c} + \beta_{1|c}\ \text{genetic predisposition}_{3\%\ rta|c} + \beta_{2|c}\ \text{genetic predisposition}_{15\%\ rta|c} + \beta_{3|c}\ \text{CRC risk}_{70\%\ rta|c} + \beta_{4|c}\ \text{CRC risk}_{99\%\ rta|c} + \beta_{5|c}\ \text{colonoscopy frequency}_{2\text{years } rta|c} + \beta_{6|c}\ \text{colonoscopy frequency}_{5\text{years } rta|c} + \beta_{7|c}\ \text{survival}_{92\%\ rta|c} + \beta_{8|c}\ \text{survival}_{98\%\ rta|c}$$

After fitting the above-specified utility function, a class assignment model was fitted. All demographic variables and all theorems were tested for a significant contribution to the class assignment model, and the final class assignment utility function was:

$$V_{rc} = \beta_{0|c} + \beta_{1|c}\ \text{high educational level}_r + \beta_{2|c}\ \text{experience with genetic screening}_r + \beta_{3|c}\ \text{being anxious and worried about CRC predisposition}_r$$

A significant estimate in this function indicates that this variable contributes to the class assignment (e.g., if the higher education variable is positive and significant for class 1, this indicates that respondents with a higher educational level are more likely to belong to class 1).

Relative importance of the attributes

The relative importance of the attributes was estimated separately for both classes of the panel latent class models. The difference between the highest and lowest attribute level estimate was calculated for each attribute. The largest difference value received an importance score of 1,

representing the attribute that was deemed most important by respondents. The other difference values were divided by the largest difference value resulting in a relative distance between all other attributes and the most important attribute.

Utility scores for Lynch syndrome, FAP and FCC screening

For each of the three realistic screening scenarios, specific utility scores were calculated for both classes separately. The attribute levels that correspond with each of the three screening scenarios were entered into the utility function. The outcome (V) represents individuals' willingness to participate in one screening initiative compared to the other initiatives.

RESULTS

Respondents' characteristics

Of the individuals initially invited (n=5500), 798 (14.5%) respondents started the questionnaire within the first four weeks of data collection. Complete data was gathered for 532 eligible respondents (66.7% of those who started the questionnaire) and data-collection was closed.

Table 2. Demographic characteristics of the study population (n=532).

		Mean (SD)	Percentage
Age		59.5 (3.1)	
Gender	Female		50.9
Educational level	Low		26.3
	Average		37.1
	High		36.6
Health literacy	Inadequate		3.4
Ethnicity	Dutch		96.6
Previously participated in another cancer screening program			48.9
Previously diagnosed with cancer			14.0
Family member previously diagnosed with cancer			24.6
Previously participated in genetic screening			7.7

Table 2 describes the demographic characteristics of the study population. The majority of the respondents reported that genetic screening for CRC is important for themselves as well as for their family (Table 3). Although about half of the respondents expect to become seriously anxious and worried about developing CRC due to a suspected genetic predisposition, 89.0% reported that they would participate in genetic screening for CRC if such a program would become available (Table 3).

Table 3. Proportion of respondents who agree with the provided theorems concerning genetic screening for CRC (n=532)^.

	Percentage
I think genetic testing for CRC is useful	89.1
I think it is important to take part in genetic testing for CRC	86.9
I consider it self-evident to take part in CRC	77.4
It would not be difficult for me to take part in genetic testing for CRC	76.7
It is important that my family takes part in genetic screening for CRC	70.3
My family would take part in genetic screening for CRC	69.1
My family would consider it important that I take part in genetic screening for CRC	71.2
It is important to know whether I am genetically predisposed so my family can take precautions	87.4
I would inform my family if I was genetically predisposed to develop CRC	89.1
I would be seriously anxious if I was genetically predisposed to develop CRC	43.7
I would find it seriously worrying if I was genetically predisposed to develop CRC	65.2
I always want to know about incidental findings	75.9
I never want to know about incidental findings	3.2
I only want to know about incidental findings concerning diseases that can be prevented	8.3
I only want to know about incidental findings concerning diseases that can be treated	12.6
I would take part in genetic screening for CRC	89.1

^ CRC = colorectal cancer

Preferences for genetic screening for CRC

The average probability of respondents belonging to either of the two latent classes was 65% and 35%, respectively, but this depended on educational level, experience with genetic screening tests and being worried and anxious about being predisposed to develop CRC (Table 4). Respondents with a higher educational level, respondents who had no experience with genetic screening tests and respondents who were less worried and anxious about their predisposition to develop CRC were more likely to belong to class 1. The probability of belonging to class 2 increased when respondents had a lower educational level, when they had experience with genetic screening tests, and when they were worried and anxious about being predisposed to develop CRC.

In both classes, respondents preferred a genetic screening test when their probability of being genetically predisposed to develop CRC was high and their survival rate due to screening would increase the most. Respondents in class 1 preferred a genetic screening when the probability that they would develop CRC due to their genetic predisposition was highest, while respondents in class 2 preferred a genetic screening test if the probability that they would develop CRC due to their genetic predisposition was lowest. Respondents in class 1 preferred to have a biannual colonoscopy, while respondents in class 2 preferred to have an annual preventive colonoscopy.

Table 4. Preferences for genetic testing for colorectal cancer based on latent class analysis^

		Class1			Class 2		
		Estimate	**SE**	**RI**	**Estimate**	**SE**	**RI**
Constant		0.21***	0.04		0.14	0.12	
Genetic predisposition	1% (ref)	-0.20***	0.04		0.18*	0.11	
	3%	-0.07*	0.04	4	-0.31***	0.10	4
	15%	0.27***	0.04		0.13	0.10	
CRC risk	15% (ref)	-0.37***	0.04		0.63***	0.11	
	70%	0.20***	0.04	3	-0.22**	0.10	2
	99%	0.17***	0.04		-0.41***	0.11	
Colonoscopy frequency	Every year (ref)	-0.29***	0.04		1.13***	0.12	
	Every 2 years	0.29***	0.04	2	0.61***	0.11	1
	Every 5 years	-0.00	0.05		-1.74***	0.18	
Survival	80% (ref)	-0.57***	0.04		-0.51***	0.16	
	92%	-0.01	0.04	1	0.10	0.08	3
	98%	0.58***	0.05		0.41**	0.16	
Class probability model		0.78***	0.18				
Constant		1.02***	0.24				
Higher education		-0.87**	0.39				
Experience with genetic screening		-0.88***	0.22				
Being worried and anxious							
Average class probability		0.65			0.35		

^ SE = standard error, RI = relative importance, CRC = colorectal cancer, the attribute level estimate of the reference categories can be calculated as -1*(sum of the other attribute level estimates); *P<0.10; ** P <0.05; ***P <0.01

Relative importance of the attributes

Respondents in both classes reported different preferences with respect to genetic screening for CRC, which indicates preference heterogeneity. Respondents in class 1 found survival to be most important, followed by colonoscopy frequency, CRC risk and genetic predisposition (Table 4). For respondents in class 2, colonoscopy frequency was most important (relative importance score of 1) followed by CRC risk, survival and genetic predisposition (Table 4). Figure 1 shows these results in more detail, here the values of the attributes display the relative distance of all attributes to the most important attribute on a scale of 0-1. The range in those distances is large in class 2, while for respondents in class 1 most attributes were approximately equally important.

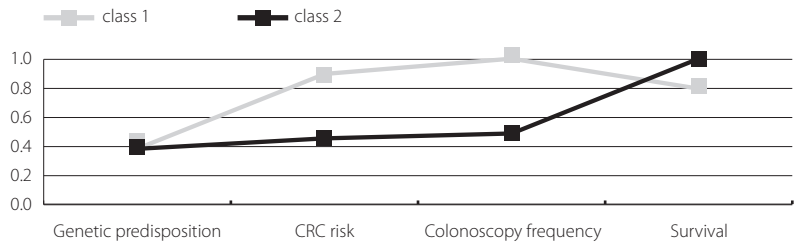

Figure 1. Relative importance of the attributes stratified by class*^.

* Values reflect the relative distance of all attributes to the most important attribute on a scale from 0-1 (1 indicating the most important attribute). ^CRC = colorectal cancer

Utility scores for Lynch syndrome, FAP and FCC screening

Respondents in class 1 (higher education, no experience with genetic testing and who are less anxious and worried about CRC predisposition) preferred Lynch syndrome screening (V = 0.62) over FAP screening (V = -0.68) and preferred FCC screening (V = 0.69) over both the other screening initiatives. Respondents in class 2 (respondents with a lower education, experience with genetic screening tests and respondents who are anxious and worried about CRC predisposition) preferred Lynch syndrome screening (V = 0.32) over FCC screening (V = -0.43) and preferred FAP screening (V = 0.53) over both the other screening initiatives.

DISCUSSION

This study shows that the probability of being genetically predisposed, the probability of developing CRC, the frequency of preventive follow-up colonoscopies and the probability of surviving CRC, all influence respondents' preferences for genetic screening for CRC. However, results also show heterogeneity in these preferences. Respondents with a lower education found colonoscopy frequency and the probability of developing CRC relatively more important and survival relatively less important than higher educated respondents. These differences in preferences were also found among respondents who had some versus no experience with genetic screening tests and among respondents with serious or little anxiety and worries about being genetically predisposed to develop CRC. Due to the differences in preferences among subgroups in the population, their willingness to participate in specific genetic screening initiatives also differed. Respondents in class 1 preferred FCC screening most and FAP screening least, while the respondents in class 2 showed complete opposite preferences for these screening initiatives.

This is the first DCE that studied the preferences of the general population for genetic screening for CRC. However, previous studies did measure preferences for population-based CRC screening program characteristics (without genetic screening) [33-36] or preferences for genetic screening test characteristics in general (not specifically applied to CRC) [24]. Although these studies focused on different topics and different target populations, their results do provide face validity for the results of the current study.

Insights into the preferences of the target population for genetic screening for CRC provide clear recommendations for effective communication about genetic screening between counselors and counselees [22, 37]. In turn, optimized communication may improve knowledge among the general population and may facilitate informed decision making among individuals who are offered genetic screening. First, respondents deemed survival as a highly important test characteristic of genetic screening. While increased survival rates as a result of participation in genetic screening should be communicated to those eligible for screening, counselors should also always ensure that individuals understand that their survival rates will only increase if they participate consistently in preventive colonoscopies. Second, the current study shows that some of the respondents preferred annual preventive colonoscopies, in particular those who had a lower educational level and who expressed serious anxiety and worries about a genetic predisposition for CRC. These respondents might have reasoned that frequent screening will increase the likelihood of early cancer detection and therefore will increase their probability of surviving CRC. However, annual colonoscopies are, at present, only recommended for surveillance of individuals diagnosed with FAP [29-31], individuals with Lynch syndrome or FCC are usually screened less often (biannually and every five or six years respectively) based on solid clinical evidence [4, 29-31]. For respondents with a lower educational level or those who are anxious and worried, effective communication and counseling is necessary to reduce their anxiety and to explain that screening frequency depends on the specific type of genetic or familial CRC. Third, respondents preferred a genetic screening test when their probability of being genetically predisposed increased. This result supports the fact that participation in screening will increase if all individuals suspected of genetic or familial CRC are actively informed about their personal risk of being predisposed to develop CRC. Since the relative occurrence of FAP within all genetic or familial CRC is relatively small (about 1%), some participants appeare less interested in FAP screening. However, this is the most aggressive and severe genetic variant of CRC for which active screening is of utmost importance [38, 39]. Fortunately, most FAP patients are aware of their predisposition from a young age due to a family history resulting in an acceptable surveillance compliance [40]. However, when clinicians suspect individuals may have FAP without a clear family history, they are advised to (continue to) stress the importance of active screening once genetic predisposition is confirmed.

This study is subject to some limitations. First, generalizability of our results to non-Dutch individuals may be limited because the number of non-Dutch respondents in our study population is relatively low compared to Dutch national population figures. Second, some respondents probably perceived the choice tasks to be difficult. This may be reflected by the valuation of the CRC risk attribute; respondents with a lower educational level preferred to participate in a genetic screening test for CRC if their risk of actually developing CRC as a result of their genetic predisposition would be lowest. They might have mistakenly interpreted this attribute as their probability of developing CRC in general. In addition, respondents tended to select the left scenario over the right scenario within the presented choice tasks, which might indicate that they were not fully committed to the task. Although such bias could potentially have a serious impact on the study outcomes, our sensitivity analysis showed that results did not significantly change by adjusting for this type of bias in the modelling procedure.

In conclusion, the current study suggests that the general population is willing to participate in genetic screening for CRC. Both screening situation characteristics and screening test characteristics influenced respondents' preferences for genetic screening for CRC. The increased survival rates as a result of genetic screening was the most important screening test characteristic for respondents with a higher educational level, respondents who have no experience with genetic testing and who are less anxious and worried about CRC predisposition. The frequency of colonoscopies was the most important screening test characteristic for respondents with a lower educational level, experience with genetic testing and who were anxious an worried about CRC predisposition. If individuals are suspected of genetic or familial CRC, they should at least be informed about their increased risk of being genetically predisposed, and about the importance of participating in all preventive follow-up colonoscopies in order to maximize their survival. Specifically, respondents with a lower educational level and those who express worries or anxiety should be informed about the frequency of preventive colonoscopies that is appropriate for the genetic or familial CRC they are diagnosed with.

References

1. Hickner, J., *Will screening open Pandora's box?* J Fam Pract., 2013. 62(9): p. 465.
2. Yazdi, S.M. and N.H. Robin, *We need to know our limitations: genetic testing for complex traits.* Curr Opin Pediatr., 2013. 25(6): p. 643-4.
3. Di Lena, M., E. Travaglio, and D.F. Altomare, *New strategies for colorectal cancer screening.* World J Gastroenterol., 2013. 19(12): p. 1855-60.
4. Giardiello, F.M., et al., *Guidelines on genetic evaluation and management of Lynch syndrome: a consensus statement by the US Multi-society Task Force on colorectal cancer.* Am J Gastroenterol, 2014. 109(8): p. 1159-79.
5. Ferlay, J., et al., *Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012.* Int J Cancer, 2014. 13(10): p. 29210.
6. Pignone, M., et al., *Screening for colorectal cancer in adults at average risk: a summary of the evidence for the U.S. Preventive Services Task Force.* Ann Intern Med., 2002. 137(2): p. 132-41.
7. Walsh, J.M. and J.P. Terdiman, *Colorectal cancer screening: scientific review.* JAMA., 2003. 289(10): p. 1288-96.
8. Bogaert, J. and H. Prenen, *Molecular genetics of colorectal cancer.* Ann Gastroenterol, 2014. 27(1): p. 9-14.
9. Haydon, A.M. and J.R. Jass, *Emerging pathways in colorectal-cancer development.* Lancet Oncol., 2002. 3(2): p. 83-8.
10. Fearon, E.R. and B. Vogelstein, *A genetic model for colorectal tumorigenesis.* Cell, 1990. 61(5): p. 759-67.
11. Jarvinen, H.J., et al., *Ten years after mutation testing for Lynch syndrome: cancer incidence and outcome in mutation-positive and mutation-negative family members.* J Clin Oncol., 2009. 27(28): p. 4793-7.
12. Vasen, H.F., et al., *One to 2-year surveillance intervals reduce risk of colorectal cancer in families with Lynch syndrome.* Gastroenterology., 2010. 138(7): p. 2300-6.
13. Bellcross, C.A., et al., *Implementing screening for Lynch syndrome among patients with newly diagnosed colorectal cancer: summary of a public health/clinical collaborative meeting.* Genet Med, 2012. 14(1): p. 152-62.
14. Bredenoord, A.L., N.C. Onland-Moret, and J.J. Van Delden, *Feedback of individual genetic results to research participants: in favor of a qualified disclosure policy.* Hum Mutat, 2011. 32(8): p. 861-7.
15. Lolkema, M.P., et al., *Ethical, legal, and counseling challenges surrounding the return of genetic results in oncology.* J Clin Oncol, 2013. 31(15): p. 1842-8.
16. Burton, A.M., S.R. Hovick, and S.K. Peterson, *Health behaviors in patients and families with hereditary colorectal cancer.* Clin Colon Rectal Surg, 2012. 25(2): p. 111-7.
17. Glanz, K., et al., *Correlates of intentions to obtain genetic counseling and colorectal cancer gene testing among at-risk relatives from three ethnic groups.* Cancer Epidemiol Biomarkers Prev., 1999. 8(4 Pt 2): p. 329-36.
18. Finkler, K., C. Skrzynia, and J.P. Evans, *The new genetics and its consequences for family, kinship, medicine and medical genetics.* Soc Sci Med., 2003. 57(3): p. 403-12.
19. Tambor, E.S., et al., *Mapping the human genome: an assessment of media coverage and public reaction.* Genet Med., 2002. 4(1): p. 31-6.
20. Leventhal, K.G., et al., *"Is it really worth it to get tested?": primary care patients' impressions of predictive SNP testing for colon cancer.* J Genet Couns., 2013. 22(1): p. 138-51.
21. Matro, J.M., et al., *Cost Sharing and Hereditary Cancer Risk: Predictors of Willingness-to-Pay for Genetic Testing.* J Genet Couns, 2014. 6: p. 6.
22. Walsh, J., et al., *Preferences for genetic testing to identify hereditary colorectal cancer: perspectives of high-risk patients, community members, and clinicians.* J Cancer Educ., 2012. 27(1): p. 112-9.
23. Kuppermann, M., et al., *Preferences for outcomes associated with decisions to undergo or forgo genetic testing for Lynch syndrome.* Cancer., 2013. 119(1): p. 215-25.
24. Severin, F., et al., *Eliciting preferences for priority setting in genetic testing: a pilot study comparing best-worst scaling and discrete-choice experiments.* Eur J Hum Genet., 2013. 21(11): p. 1202-8.
25. Huang, M.Y., S.A. Huston, and M. Perri, *Consumer preferences for the predictive genetic test for Alzheimer disease.* J Genet Couns., 2014. 23(2): p. 172-8.
26. Hensher, D.A., J.M. Rose, and W.H. Greene, *Applied Choice Analysis: A Primer.* 2005, New York: Cambridge University Press.

27. Hall, J., et al., *What influences participation in genetic carrier testing? Results from a discrete choice experiment.* J Health Econ., 2006. 25(3): p. 520-37.
28. Hiligsmann, M., et al., *Nominal group technique to select attributes for discrete choice experiments: an example for drug treatment choice in osteoporosis.* Patient Prefer Adherence, 2013. 7:133-9.
29. American Gastroenterological, A., *American Gastroenterological Association medical position statement: hereditary colorectal cancer and genetic testing.* Gastroenterology, 2001. 121(1): p. 195-7.
30. Dutch Society For Clinical Genetics, *CBO Guideline Hereditary Colorectal Cancer 2008.* 2008, Van den Boogaard: Oisterwijk.
31. Vasen, H.F., et al., *Familial colorectal cancer risk: ESMO clinical recommendations.* Ann Oncol, 2009. 20 Suppl 4: p. 51-3.
32. Fransen, M.P., et al., *Applicability of internationally available health literacy measures in the Netherlands.* J Health Commun, 2011. 16(Suppl 3): p. 134-49.
33. Benning, T.M., et al., ***T****he effect of presenting information about invasive follow-up testing on individuals' noninvasive colorectal cancer screening participation decision: results from a discrete choice experiment.* Value Health, 2014. 17(5): p. 578-87.
34. Groothuis-Oudshoorn, C.G., et al., *Public stated preferences and predicted uptake for genome-based colorectal cancer screening.* BMC Med Inform Decis Mak, 2014. 14: p. 18.
35. Hol, L., et al., *Preferences for colorectal cancer screening strategies: a discrete choice experiment.* Br J Cancer, 2010. 102(6): p. 972-80.
36. van Dam, L., et al., *What determines individuals' preferences for colorectal cancer screening programmes? A discrete choice experiment.* Eur J Cancer, 2010. 46(1): p. 150-9.
37. Pieterse, A.H., et al., *Initial cancer genetic counseling consultation: change in counselees' cognitions and anxiety, and association with addressing their needs and preferences.* Am J Med Genet A., 2005. 137(1): p. 27-35.
38. Nieuwenhuis, M.H. and H.F. Vasen, *Correlations between mutation site in APC and phenotype of familial adenomatous polyposis (FAP): a review of the literature.* Crit Rev Oncol Hematol, 2007. 61(2): p. 153-61.
39. Rozen, P. and F. Macrae, *Familial adenomatous polyposis: The practical applications of clinical and molecular screening.* Fam Cancer, 2006. 5(3): p. 227-35.
40. Douma, K.F., et al., *Long-term compliance with endoscopic surveillance for familial adenomatous polyposis.* Colorectal Dis, 2010. 12(12): p. 1198-207.

ADDITIONAL FILE 1:

Example choice task

Imagine that you have participated in the National Colorectal Cancer Screening Program. Blood was detected in your feces and you went to the hospital for a colonoscopy. Based on the colonoscopy and your family cancer history, your physician advises you to participate in a genetic test to establish whether you are genetically predisposed to develop colorectal cancer. In which scenario would you prefer to participate in a genetic test, scenario 1 or scenario 2?

	Scenario 1	**Scenario 2**
Probability of being genetically predisposed	1% 1 out of every 100	3% 3 out of every 100
Probability of developing CRC^	70% 70 out of every 100	99% 99 out of every 100
Frequency of preventive colonoscopies	Every 5 years	Every year
Probability of surviving CRC^	80% 80 out of every 100	98% 98 out of every 100
Tick the box of the scenario that you prefer:	☐	☐

^ CRC = colorectal cancer

PART 2

THE VALIDITY OF THE PREFERENCES THAT ARE MEASURED WITH DISCRETE CHOICE EXPERIMENTS IN PUBLIC HEALTH

CHAPTER 4

EXPLORING HOW INDIVIDUALS COMPLETE THE CHOICE TASKS IN A DISCRETE CHOICE EXPERIMENT: AN INTERVIEW STUDY

J. Veldwijk*, D. Determann*, M.S. Lambooij,
E.W. de Bekker-Brob, J.A. van Til, I. Korfage, G.A. de Wit

* Equal contribution

Submitted.

ABSTRACT

Objective: To what extent do participants evaluate and complete the choice tasks in Discrete Choice Experiments (DCE) according to the underlying methodological assumptions? What is the impact of education and health literacy of the participants?

Methods: Two existing DCE questionnaires on rotavirus vaccination and prostate cancer screening served as case studies. During structured interviews, 70 Dutch participants completed five choice tasks aloud.

Results: About 59% of the participants read the choice task attribute-wise from top-to-bottom. Nearly all (89%) participants chose the scenario with the optimal attribute levels (confirming the monotonicity axiom). In accordance with the continuity axiom, most participants (66%) mentioned three or more attributes when motivating their decisions. In total, sixteen participants continuously traded off two attributes or less when completing the choice tasks, of which seven found it hard to trade off multiple attributes or did not understand the meaning of certain attributes. Lower educated and less literate participants tended to cite fewer than three or more attributes when motivating their decision and used trading off between attributes less often as a decision-making strategy.

Conclusions: The majority of the participants complete a DCE as presumed by its underlying methodology. However, the assumptions did not hold for lower educated and less literate participants. Based on the participants' age, educational level and health literacy, additional measures such as conducting DCEs in mini-labs and thorough pilot testing, should be undertaken to enhance the participants' understanding of the attributes, the attribute levels and the choice tasks in a DCE.

INTRODUCTION

Do participants read and evaluate the choice tasks in a Discrete Choice Experiment (DCE) as researchers presume they do? DCEs are increasingly being used to make inferences on individuals' preferences for a wide range of products or services within a health care context [1, 2]. It is essential that researchers understand how participants read and interpret choice tasks, attributes, attribute levels, and that researchers have insight in how participants ultimately make their decision for one of the scenarios, to be able to make valid inferences on preference data.

The existing guidelines for designing and conducting DCEs are based on several assumptions regarding the participants' decision-making process [3-5]. Researchers quantify individuals' preferences while trusting that these assumptions hold. First, it is assumed that participants are actively involved in completing the choice tasks and that they use complex and rational decision-making processes to complete the choice task aiming at the most accurate response possible [6]. Second, participants are expected to understand and interpret the information presented as intended by the researcher [7, 8]. The responses are then analyzed by researchers according to economic theory (Lancaster's theory of demand [9] and random utility theory [10, 11] or random regret minimization [12, 13]). Third, participants are expected to prefer more attractive over less attractive attribute levels (monotonicity axiom) [13-15]. Fourth, it is expected that participants include all attributes in their considerations and make their decisions based on trade-offs between these attributes (continuity axiom) [13-15]. Fifth, participants are expected to have stable preferences, if scenario A is preferred over B at one point in the questionnaire, scenario A should also be preferred over B at other points (completeness axiom) [13-15]. Sixth, if participants prefer scenario A over scenario B, and scenario B over C, then respondents are expected to prefer scenario A over scenario C (transitivity axiom) [13-15].

Both within and outside the health care setting, qualitative and quantitative research that mainly focuses on the first four assumptions shows that these assumptions do not always hold since participants behave differently when making discrete decisions. First, completing a choice task can be a cognitive challenge [16], because choice tasks often contain a large amount of information and respondents are assumed to take into account and trade off multiple attributes at the same time [17]. Previously conducted research has shown that cognitively demanding decisions may induce the use of simplified heuristics [18-21], which may not be in accordance with the first assumption that participants apply complex and rational decision strategies to ensure they make the most accurate response. The second assumption (respondents use information as intended by the researcher), may be violated especially in health related DCEs. Those DCEs often contain risk attributes or other relatively abstract concepts. Research shows that respondents often misinterpret risk information [22, 23]. Difficulties with understanding risk attributes can trigger participants to interpret numerical values as categorical information (for instance, low-medium-high or good-intermediate-bad) [6, 24, 25], or to apply other simplifying decision strategies [26]. If such strategies are applied, marginal rates of substitution estimates might not reflect the participants' true preferences, as the numerical values of the attribute level were not the basis for the chosen option and thus cannot be interpreted as such. With regard to the third assumption (i.e. monotonicity), research shows that respondents

sometimes prefer less optimal attribute levels over more optimal attribute levels. This may be caused by a lack of understanding regarding the attribute (levels) or because respondents interpret the attribute levels using a specific context [25]. For instance, respondents might prefer higher cost levels over lower cost levels because they associate higher costs with better quality of the product or service. Research on the continuity axiom (the fourth assumption) shows that participants with dominant preferences base their decisions on one high priority attribute, and only trade this dominant attribute if the levels for this specific attribute are equal for both scenarios in the choice task [27]. Such non-compensatory decision-making could either reflect a true strong preference for one specific attribute or it may be a way to avoid complex decision-making by just focusing on that one attribute [28, 29]. Moreover, different quantitative studies, with different topics and among different target populations, show that up to 45% of the participants have dominant preferences [28, 30, 31]. On the other hand, it might also be the case that participants disregard certain attributes and base their decision on some, but not on all attributes (attribute-non-attendance) [25, 29, 32, 33]. All this is in conflict with the continuity axiom.

As described above, assumptions are not met mainly due to the cognitive burden of the choice task. Respondent characteristics, such as age and educational level are expected to influence cognitive ability. Highly correlated with educational level and age is health literacy, a concept that measures individual ability to access, understand, appraise and apply health-related information. Thus far, explorative research has shown that individuals with lower levels of health literacy are underrepresented in research, have difficulties with completing questionnaires and find it difficult to make informed decisions [23, 34-37]. To date, there is no insight into whether and how health literacy influences the ability of participants to complete a DCE. However, older participants as well as respondents with a lower educational level showed less understanding of a stated preference survey [38], were less consistent in their preference [39], made irrational choices more often [32] and had dominant preferences more often [28] compared to higher educated participants. Given this evidence, we expect that health literacy could possibly influence participants' ability to complete a DCE questionnaire according to the above-described assumptions.

This study explored whether the above-specified assumptions hold. Specifically, we tested how participants read the choice tasks, how many and which attributes participants take into account when making a decision per choice task, how participants interpreted the risk attributes included, what decision strategy they used to make their decision and whether the monotonicity and continuity axioms hold. The impact of educational level and health literacy was tested.

MATERIALS AND METHODS

Discrete Choice Experiments

Two previously administered DCE questionnaires were used as case studies for the current study [40, 41]. A sample of the respondents of those studies was re-contacted after previously indicating that they were willing to participate in further research. See additional file 1 for a description of both studies and Table 1a and 1b for a description of the included attributes and levels.

Table 1a. Attributes and levels for rotavirus DCE.

Attributes	Level 1	Level 2	Level 3
Vaccine effectiveness			
The percentage of children that will be protected against a rotavirus infection when vaccinated	55%	75%	95%
Frequency of severe side effects			
The number of vaccinated children that will suffer from intussusception due to vaccination. Intussusception is an acute condition in which part of the bowel telescopes into another adjacent part of the bowel, resulting in obstruction [45]	1 in 10,000	1 in 100,000	1 in 1,000,000
Protection duration			
The number of years that the vaccine protects against a rotavirus infection	1 year	3 years	6 years
Healthcare facility of vaccine administration			
Within the Netherlands all vaccines in the NIP are administrated at a child welfare center, The GP office was included because the rotavirus vaccine may not become part of the NIP; in that case it is likely that this vaccine is administrated here.	Child welfare center	General practitioner	
Out-of-pocket costs			
Parents may have to pay (part) of the vaccine costs out-of-pocket	€0	€30	€140

Participants

In total, we included 70 participants for the current study; 35 from the rotavirus DCE and 35 from the prostate cancer screening DCE. This number is a trade-off between practicality (a manageable amount of interviews) and usefulness and generalizability of results (i.e. enough opinions and interpretations, and being able to distinguish between educational level). An even distribution of both lower and higher educated participants in each cohort was guaranteed, because participants were randomly selected from the original studies based on their educational level. Educational level was dichotomized into a higher and a lower educational level, whereby a Bachelor's and/or Master's degree were defined as a higher educational level and all other educational levels were defined as a lower educational level.

If subjects agreed to participate in the current study, they received a package with materials (i.e. information letter, the original DCE questionnaire and two documents needed to measure the participant's health literacy) by mail.

Interviews

Both face-to-face (N=5 per cohort) and telephone interviews (N=30 per cohort) were scheduled. Interview guides were developed for both DCEs. Although the wording of the two interview guides was different with respect to the inherently different choice tasks, both guides described a similar

Table 1b. Attributes and levels for prostate cancer-screening DCE.

Attributes	Level 1	Level 2	Level 3	Level 4
Number of deaths from prostate cancer				
It was given that 35 out of 1000 men die because of prostate cancer when no screening program is provided.				
	32 deaths (3 deaths prevented)	28 deaths (7 deaths prevented)	25 deaths (10 deaths prevented)	18 deaths (17 deaths prevented)
Frequency of blood test				
	Every year	Every 2 year	Every 3 years	Every 4 years
Number of unnecessary biopsies				
Number of men, per 1000 men with an elevated PSA level, in which biopsies are unnecessary. Unnecessary biopsies were defined as biopsies in which no cancer was found, but in which PSA levels suggested that there was cancer.				
	200 unnecessary biopsies (800 justified biopsies)	400 unnecessary biopsies (600 justified biopsies)	600 unnecessary biopsies (400 justified biopsies)	800 unnecessary biopsies (200 justified biopsies)
Number of unnecessary treatments				
Number of men, per 1000 treated men, in whom treatment is unnecessary. Unnecessary treatment was defined as treatment that was not life prolonging, however it could lead to urine-loss and erection disorders due to treatment.				
	0 unnecessary treatments (1000 justified treatments)	200 unnecessary treatments (800 justified treatments)	500 unnecessary treatments (500 justified treatments)	800 unnecessary treatments (200 justified treatments)
Out-of-pocket costs				
	€0	€50	€100	€300

interview protocol to make the results of both groups comparable. The structured interviews were pilot tested (N=7) to optimize the interview guide, to test the duration of an interview and to ensure both interviewers conducted the interviews in the same manner. This resulted in minimal adaptations to the interview guides. All interviews started with a short introduction to the current study. Next, participants were given some time to read the introduction to the DCE questionnaire, because it had been a while since they completed the original DCE. The interview consisted of three parts, part one of the interview took place without any specific guidance by the interviewers in order to mimic non-lab questionnaire completion situations, while specific questions were asked during part two and three. During part one, we asked participants to complete four choice tasks (see additional file 2 for examples of choice tasks of both case studies) from the original DCE after having practiced with one additional choice task. We instructed the participants to read aloud and speak aloud what they thought when reading and completing the choice tasks. Furthermore, they were asked to clarify why they chose certain scenarios. If a participant was quiet for some time, the interviewer reminded him/her to keep reading aloud and to report his/her thoughts. During part two of the interview, specific questions were asked to test the interpretation of risk attributes, the monotonicity axiom, the decision strategy that participants used, and the continuity axiom. Finally, in part three of the interview, health literacy was measured both by means of a subjective

self-reported questionnaire (Set of Brief Screening Questions (SBSQ-D) of Chew [42]) and a validated objective measurement (Newest Vital Sign (NVS-D) [42]) (see additional file 3).

Two researchers (JV and DD) conducted the interviews in August-September 2013. Interviews were audio taped, the interviewers also made notes and wrote down specific observations during each interview. During a consensus meeting with all authors, both recordings and notes were used to discuss the categorization of answers.

Choice task reading

Interviewers used an answer form, in which they marked per participant which parts of the choice tasks he or she read aloud in part one of the interview. Interviewers also marked in which manner participants read the choice tasks, e.g. horizontally (attribute-wise) or vertically (scenario-wise).

Interpretation of the risk attributes and monotonicity axiom

In part one, the interpretation of the risk attributes was measured based on how participants read the risk attributes aloud, e.g. reading the actual values of attributes as presented in the choice task or translating them directly into an ordinal scale. During part two, it was tested if participants were able to give the exact definition of the risk attributes as described in the attribute explanation section of the DCE. To test the monotonicity axiom, interviewers asked participants twice to choose between two scenarios based on only one attribute. If participants chose the scenario with the 'best' attribute level they were marked as acting in accordance with the monotonicity axiom. Part two also included a control question in which participants were asked to make a simple calculation with respect to the risk attributes (for the rotavirus DCE: effectiveness of the vaccine and the frequency of severe side effects and for the prostate cancer screening DCE: number of unnecessary treatments) to test their understanding of these risk attributes. E.g. for the prostate cancer screening DCE: 'Imagine a screening program in which out of 1000 treatments, 200 are unnecessary. Imagine that 2000 men participate in this screening program. How many men will be treated unnecessarily?'.

Continuity axiom and decision strategy

To test the continuity axiom, interviewers marked on the answer form which attributes participants mentioned when motivating their decision for a certain scenario for the five choice tasks during part one. Participants were marked as acting in accordance with the continuity axiom if they mentioned three or more attributes when motivating their decision. It was also monitored which of the predefined decision strategies participants applied to make their decision (e.g., trading off attributes or counting pros and cons), or whether they used a different strategy. During part two of the interview, participants were specifically asked what strategy they applied to make a decision. If participants based their decision on less than three attributes (i.e. less than the majority of the five included attributes), they were asked why this was the case.

RESULTS

Table 2 describes the demographic characteristics of the participants who were interviewed. The face-to-face interviews did not differ from the telephone interviews with respect to duration as well as the amount and quality of information retrieved. The average duration of the interviews in the rotavirus cohort was 27 minutes, while the average duration of the interviews in the prostate cancer-screening cohort was 41 minutes.

Table 2. Demographics of participants in both cohorts

		Rotavirus cohort (n=35)	**Prostate cancer-screening cohort (n=35)**
		Mean (SD)	**Mean (SD)**
Age in years		30.4 (4.5)	67.6 (5.5)
		Proportion (%)	**Proportion (%)**
Gender	Female	94.3	0
Education	Lower	45.7	48.6
	Higher	54.3	51.4
Health literacy*	High subjective score	100	100
	High objective score	100	55.9

* High subjective score includes participants with a score >2 on the SBSQ-D. High objective score includes participants with a score of 4-6 on the NVS-D.

Choice task reading

Within both cohorts, the majority of the participants read the choice tasks attribute-wise (reading the levels for both scenarios per attribute), starting from the top and moving to the bottom (Table 3). In the rotavirus cohort, two other frequently used strategies for reading the choice tasks were 1) reading scenario-wise (reading the levels of all attributes per scenario), starting from the top and moving to the bottom, and 2) directly motivating which of the two scenarios was preferred based on the attribute levels instead of just reading all the attributes first. This latter strategy was also often applied in the prostate cancer-screening cohort. Additionally, a considerable number of participants used different reading strategies; only reading attributes that were of personal importance, only reading attributes that differed between the two scenarios, and reading choice tasks (completely) in a random manner. The prostate cancer-screening choice tasks included an opt-out option (i.e. no screening), that was specifically read aloud by 42.9% of the participants in choice task one, by 25.7% in choice task two, 20.0% in choice task three and 8.6% of the participants in choice task four.

Table 3. How respondents read the choice tasks

		Choice task 1 (%)	Choice task 2 (%)	Choice task 3 (%)	Choice task 4 (%)
Rotavirus cohort (n=35)	Attribute-wise from top to bottom	60.0	62.9	57.1	62.9
	Attribute-wise other	8.6	2.9	8.6	2.9
	Scenario-wise from top to bottom	14.3	14.3	17.1	14.3
	Scenario-wise other	0.0	0.0	0.0	0.0
	Starts motivating directly	17.1	11.4	14.3	14.3
	Otherwise	0.0	8.6	2.9	5.7
Prostate cancer-screening cohort (n=35)	Attribute-wise from top to bottom	54.3	54.3	60.0	57.1
	Attribute-wise other	5.7	2.9	11.4	14.3
	Scenario-wise from top to bottom	2.9	2.9	0.0	2.9
	Scenario-wise other	2.9	2.9	2.9	0.0
	Starts motivating directly	17.1	25.7	14.3	17.1
	Otherwise	17.1	11.4	11.4	8.6

Interpretation of the risk attributes and monotonicity axiom

In the rotavirus cohort, the participants' interpretation of two risk attributes (vaccine effectiveness and frequency of severe side effects) was tested. In all four choice tasks, the majority of the participants mentioned the actual values of the attribute levels for vaccine effectiveness while completing the choice task, i.e. 56.6%, and 45.9% mentioned this for frequency of severe side effects. For the attribute 'vaccine effectiveness', on average over the four choice tasks, 17.5% of the participants described the levels on an ordinal scale (low-intermediate-high) and 20.6% combined reading with interpretation, like:

'*In total 75 out of every 100 children are protected against a rotavirus infection, or three-quarters of the children do not become ill*'

With respect to the frequency of severe side effects, these percentages were 23.7% and 20.6% respectively.

After completing the four choice tasks, 20.0% of the participants was able to repeat the exact definition of vaccine effectiveness as described in the introduction section of the questionnaire. Another 57.1% described vaccine effectiveness as 'how well a vaccine works' and 22.9% provided a completely different definition. When asked about the meaning of the attribute side effects, the definition of side effects as provided in the questionnaire was mentioned by 37.1% of the participants, 54.1% interpreted side effects correctly but mentioned additional side effects that were not mentioned in the explanation of the attribute such as a high temperature, feeling sick or

dying, while 11.4% provided a completely different definition.

All participants chose the scenario with the highest effectiveness within both choice tasks when they were asked to choose based on this one attribute (monotonicity axiom). On average over two choice tasks, all but three (4.3%) participants chose the scenario with the lowest frequency of severe side effects (monotonicity axiom). Finally, 77.1% of the participants gave the correct answer to the control question for vaccine effectiveness, and 94.3% of the participants gave the right answer to the control question for frequency of severe side effects. These results indicate that most participants were able to interpret percentages and frequencies correctly.

The number of unnecessary biopsies and the number of unnecessary treatments were the risk attributes included in the prostate cancer-screening DCE. The majority (52.2%) of the participants mentioned the actual values of both of these attributes when reading the choice tasks. Additionally, 12.9% of the participants interpreted the number of unnecessary biopsies and 14.3% of the participants interpreted the number of unnecessary treatments when reading the choice tasks, for example:

'If I have to choose between 200 or 800 unnecessary biopsies/treatments, the likelihood of me having an unnecessary biopsies/treatment is four times as high in scenario two'

Others did not mention these attributes while reading the choice tasks (30.7% for the number of unnecessary biopsies and 29.3% for the number of unnecessary treatments). Many of the participants seemed to have experienced difficulties interpreting these two attributes. It seemed that participants who experienced such difficulties did not understand the difference between biopsies and treatment. Some participants even thought they were similar or at least had similar side effects. For instance, participants stated:

'An unnecessary biopsy is an unnecessary treatment.'

or

'Biopsy causes urine incontinence.'

Some participants stated that they ignored these attributes when reading the choice tasks for those reasons, while others misinterpreted the numbers. This misinterpretation was shown again, when participants were asked to give the definition of unnecessary treatments after completing all four choice tasks. Only 17.1% of the participants was able to give the exact definition of this attribute as described in the attribute explanation section of the questionnaire. Of all participants, 83% chose the scenario with the lowest level of unnecessary treatments (monotonicity axiom). Although the concepts might not have been completely clear to some participants, 88.6% answered the control question correctly, indicating that the participants were able to interpret the numbers of unnecessary treatment correctly.

Continuity axiom and decision strategy

In both the rotavirus and the prostate cancer-screening cohort, most participants mentioned three or more attributes (i.e. the majority of the included attributes) while motivating their choice for a scenario, which is in accordance with the continuity axiom (72.9% and 60.0% respectively on average over the four choice tasks). In both cohorts, the majority also traded off between the levels of those attributes when motivating their decision (Table 4), which is again in accordance with the continuity axiom. Within the rotavirus cohort, 20.0% mentioned two attributes and 7.2% only mentioned one attribute when motivating their decisions. In the prostate cancer-screening cohort 16.4% mentioned two attributes, 17.9% only mentioned one attribute and 5.7% did not mention any of the attributes when motivating which scenario they preferred (they chose to opt-out). A total number of 16 participants (one in the rotavirus cohort and 15 in the prostate cancer-screening cohort) continuously traded off two attributes or less when completing the choice tasks. Nine out of those 16 participants stated that they traded off so few attributes because only those attributes were important to them, the other seven mentioned that they did so because they found it hard to trade off more attributes at once or because they did not understand the meaning of certain attributes. This latter category also comprised participants of whom it is highly questionable whether they grasped the questions and understood the hypothetical nature of the choice tasks at all. This was reflected in the finding that the participants decided per attribute which scenario they preferred without making a final decision for one overall scenario, or they mentioned things such as:

'What is the difference between this question and the previous one?'

or

'Can I switch between scenarios within one question?'

Differences by educational level and health literacy

Overall, there is a trend showing that more educated and literate participants included three or more attributes when motivating their decision and that they traded off between attributes more often compared to participants with a lower educational level or lower health literacy score (Table 5). Additionally, higher educated and literate participants more often correctly explained the risk attributes and more often answered the risk attribute control question correctly (Table 5). Finally, lower educated and less literate participants who based their decision on two attributes or less, more often stated that they found it difficult to compare all attributes.

DISCUSSION

In general, participants were actively involved and willing to provide the response that was required. Most participants understood the attributes. More optimal attribute levels were preferred over less optimal attribute levels, which is in accordance with the monotonicity axiom. Furthermore, the majority of the participants based their decision on three or more attributes, which is in accordance with the continuity axiom. However, about a third of the participants used simplifying strategies such as basing their decision on less than three attributes (34%), or reported inverse preferences concerning attribute levels (4%).

Table 4. Continuity axiom and decision strategy

		Average over all four choice tasks (%)
Rotavirus cohort (n=35)	**Motivating decision (continuity axiom)***	
	Motivation based on one attribute	7.2
	Motivation based on two attributes	20.0
	Motivation based on three or more attribute	72.9
	Decision strategy	
	Traded off attribute levels between each other	85.6
	Based decision on one attribute	11.5
	Otherwise	2.9
Prostate cancer-screening cohort (n=35)	**Motivating decision (continuity axiom)*†**	
	Motivation based on one attribute	17.9
	Motivation based on two attributes	16.4
	Motivation based on three or more attribute	60.0
	Decision strategy	
	Traded off attribute levels between each other	60.0
	Based decision on one attribute	26.4
	Otherwise	13.6

***** Participants were marked as acting in accordance with the continuity axiom, only if they motivated their decision based on three or more attributes.

† These numbers do not add up to 100% because some men did not mention any of the attributes when motivating which scenario they preferred; they chose opt-out (5.7%).

Of the two axioms tested, the continuity axiom seems most difficult to uphold, which might not be a problem per se. In real life individuals might also not include all product characteristics when making their decision. However, within a DCE analysis, this may result in invalid conclusions on the attribute level estimates and estimated potential uptake rates, since a multi-attribute approach is undertaken to analyze DCE data [38, 43]. Therefore, estimated stated preferences might not reflect revealed behavior as accurately as possible. Previous research described that this non-compensatory decision-making behavior might have different causes; participants might actually have dominant preferences, it might be that the attribute levels are too similar, or that the participants lack understanding of certain attribute levels [14]. This latter was shown in the current study, since the analysis on risk attribute interpretation showed that, specifically within the prostate cancer-screening cohort, a majority of the participants indicated that they did not understand one or more attributes (mostly the risk attributes). Studies state that a lack of understanding of certain attribute (levels) might be due to a lower educational level, older age and a lower health literacy [6, 23, 37-39]. This study did indeed show that the number of attributes included in decision-making,

Table 5. Differences in educational level and health literacy*

	Rotavirus cohort		Prostate cancer-screening cohort	
	Educational level (n=35)		Educational level (n=35)	
	Lower (%)	Higher (%)	Lower (%)	Higher (%)
Including three or more attributes when motivating decisions	81.3	100.0	70.6	83.3
Trading off attribute levels as a strategy to make a decision	56.3	73.7	35.3	44.4
Right explanation of vaccine effectiveness	12.5	26.3	-	-
Right explanation of severe side effects	56.3	94.7	-	-
Right explanation of unnecessary treatments	-	-	11.8	22.2
Right answer to control question on vaccine effectiveness	18.8	52.6	-	-
Right answer to control question on severe side effects	87.5	100.0	-	-
Right answer to control question on unnecessary treatments	-	-	82.4	94.4
			Health literacy (n=34)	
			Low (%)	High (%)
Including three or more attributes when motivating decisions			80.0	73.7
Trading off attribute levels to make a decision			33.3	47.4
Right explanation of unnecessary treatments			6.7	21.1
Right answer to control question on unnecessary treatments			80.0	94.7
			Combined measure (n=20)†	
			Low (%)	High (%)
Including three or more attributes when motivating decisions			77.8	81.8
Trading off attribute levels to make a decision			33.3	54.5
Right explanation of unnecessary treatments			0.0	18.2
Right answer to control question on unnecessary treatments			77.8	100.0
Perceived it as difficult to trade off >2 attributes			60.0	33.3

* Differences in health literacy could only be calculated for the prostate cancer-screening cohort, because 100% of the participants in the rotavirus cohort had high objective health literacy scores. †Individuals that scored low on both educational level and objective health literacy (n=9) or scored high on both educational level and objective health literacy (n=11).

decision strategy, interpretation of the risk attributes and understanding of the risk attributes differed between participants with different educational levels and health literacy scores. This might also be reflected by the fact that the mean interview duration of the less literate and older prostate cancer-screening group was almost 15 minutes longer compared to the rotavirus cohort. Besides educational level and health literacy scores, the topic of the DCEs and the included attributes and attribute levels may have added to the differences that were found between the two cohorts.

The results of our study indicate that differences in educational level, age and health literacy might be the cause of the violation of the assumptions underlying a DCE, including the axioms observed in this study. We therefore recommend the conducting of DCE questionnaires among older and/or less health literate populations in, for instance, mini-labs, where participants complete DCEs in the presence of a researcher. Researchers have the opportunity to explain how to complete a DCE, including the hypothetical nature of the questionnaire and to answer questions that arise during the completion of the questionnaire, e.g. concerning the attributes and attribute levels. This is important especially among older target populations as participants in the prostate cancer-screening cohort sometimes indicated that they had difficulties with interpreting the questions (e.g., ***'In real life, I have a blood test to check my PSA levels every year, so I can only choose a scenario with that frequency of blood testing'***). This is in line with the findings of previous studies [32, 33]. Moreover, when conducting online research, the understanding of attribute levels among participants with a lower educational level and/or health literacy can be enlarged by providing the option to play a tape that reads out the attribute explanations to the participants or by other technical solutions, e.g. pop-ups when clicking on attributes or levels. In addition, the option to listen to the explanation again while completing the choice tasks could be offered. Another recommendation is that a thorough pilot testing phase is necessary while developing a DCE, which includes think aloud testing to identify possible problematic issues with the completion of the questionnaire a priori. Finally, age, educational level and health literacy should be standard measures to include in every DCE questionnaire as well as in the analysis of DCE data. Until options to correct DCE responses for possible differences in demographic characteristics become common practice, researchers should at least describe these measures in their population and explain the possible effects on the results retrieved.

This study was subject to some limitations. Firstly, the two DCEs that we used as case studies for this study were quite difficult, because each included two risk attributes. It is commonly known that the interpretation of such attributes is perceived as more difficult by participants than for instance qualitative attributes [22, 44]. Difficulties in interpreting attribute levels and making decisions might therefore be more pronounced in this study compared to DCEs that include no or less risk-related attributes. However, since most health-related decisions include risk information, the case studies used for this study may be representative for DCEs within a healthcare context. Secondly, although this study tested a wide range of assumptions, the completeness and transitivity axioms could not be tested due to the design of the study. Thirdly, although efforts were made to mimic non-lab choice situations, the fact that the interviewers were present during DCE completion might have influenced how participants completed the choice tasks. Participants therefore might have been more committed to completing the DCE. As a result, we might have overestimated the number of participants that acts in accordance with the tested assumptions. Fourthly, the sample size of 70 is relatively large for an interview study, at the same time, this sample size is too small to draw any conclusions based on statistical testing. However, the trends in the findings and the agreement of the current findings with the existing literature related to educational level and health literacy (non-DCE studies) provide face validity for the current study results. Confirmation of our findings is needed, e.g. from new DCEs including (preferably objective) health literacy measurements as well

as axiom testing questions in their study.

In conclusion, the majority of the participants seem to complete a DCE as presumed by its methodology. However, based on the participants' age, educational level and health literacy additional measures should be undertaken to ensure that every participant understands the choice tasks and completes the DCE as presumed.

References

1. Clark, M.D., et al., *Discrete Choice Experiments in Health Economics: A Review of the Literature.* Pharmacoeconomics, 2014.
2. de Bekker-Grob, E.W., M. Ryan, and K. Gerard, *Discrete choice experiments in health economics: a review of the literature.* Health Econ, 2012. 21(2): p. 145-72.
3. Bridges, J.F., et al., *Conjoint analysis applications in health--a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force.* Value Health, 2011. 14(4): p. 403-13.
4. Johnson, R.F., et al., *Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force.* Value Health, 2013. 16(1): p. 3-13.
5. Lancsar, E. and J. Louviere, *Conducting discrete choice experiments to inform healthcare decision making: a user's guide.* Pharmacoeconomics, 2008. 26(8): p. 661-77.
6. Lloyd, A.J., *Threats to the estimation of benefit: are preference elicitation methods accurate?* Health Econ, 2003. 12(5): p. 393-402.
7. Hensher, D.A., J.M. Rose, and W.H. Greene, *Applied Choice Analysis: A Primer.* 2005, New York: Cambridge University Press.
8. Ryan, M., K. Gerard, and M. Amaya-Amaya, *Using Discrete Choice Experiments to Value Health and Health Care.* The Economics of Non-Market Goods and Resources, ed. I.J. Bateman. 2008, Dordrecht: Springer.
9. Lancaster, K.J., *A new approach to consumer theory.* J Polit Econ, 1966: p. 132-157.
10. McFadden, D., *Conditional Logit Analysis of Qualitative Choice Behavior, in Frondtiers in Econometrics,* P. Zarembka, Editor. 1974, Academic Press: New York.
11. McFadden, D., *The Choice Theory Approach to Market Research.* Marketing Science, 1986. 5(4): p. 275–297.
12. de Bekker-Grob, E.W. and C.G. Chorus, *Random regret-based discrete-choice modelling: an application to healthcare.* Pharmacoeconomics, 2013. 31(7): p. 623-34.
13. Lancsar, E. and J. Louviere, *Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences?* Health Econ, 2006. 15(8): p. 797-811.
14. Cairns, J. and M. van der Pol, *Repeated follow-up as a method for reducing non-trading behaviour in discrete choice experiments.* Soc Sci Med, 2004. 58(11): p. 2211-8.
15. Thurstone, L.L., *The Method of Paired Comparisons for Social Values.* J Abnorm Soc Psychol, 1927. 21: p. 384-400.
16. de Bekker-Grob, E.W., J.M. Rose, and M.C. Bliemer, *A closer look at decision and analyst error by including nonlinearities in discrete choice models: implications on willingness-to-pay estimates derived from discrete choice data in healthcare.* Pharmacoeconomics, 2013. 31(12): p. 1169-83.
17. Bryan, S. and P. Dolan, *Discrete choice experiments in health economics. For better or for worse?* Eur J Health Econ, 2004. 5(3): p. 199-202.
18. Luce, M.F., *Choosing to avoid: coping with negative emotion-laden consumer desicions.* J Cons Res, 1998. 24(4): p. 409-433.
19. Luce, M.F., J.W. Payne, and J.R. Bettman, *Emotional trade-off difficulty and choice.* J Mark Res, 1999. 36(2): p. 143-159.
20. Iyengar, S.S. and E. Kamenica, *Choice overload and simplicity seeking.* Working paper, 2007.
21. Ritov, I. and J. Baron, *Status quo and ommission biases.* Journal Oof Risk and Uncertainty, 1992. 5: p. 49-61.
22. Galesic, M. and R. Garcia-Retamero, *Statistical numeracy for health: a cross-cultural comparison with probabilistic national samples.* Arch Inter Med, 2010. 170(5): p. 462-8.
23. Waters, E.A., et al., *Formats for improving risk communication in medical tradeoff decisions.* J Health Commun., 2006. 11(2): p. 167-82.
24. Reyna, V.F. and C.J. Brainerd, *Fuzzy-trace theory and framing effects in choice: Gist extraction, truncation, and conversion.* J Behav Decis Making, 1991. 4(4): p. 249-262.
25. Ryan, M., V. Watson, and V. Entwistle, *Rationalising the 'irrational': a think aloud study of discrete choice experiment responses.* Health Eco, 2009. 18(3): p. 321-36.
26. Tversky, A. and D. Kahneman, *Judgment under uncertainty: Heuristics and biases.* Science, 1974. 185(4157): p. 1124-1131.

27. Lagarde, M., *Investigating attribute non-attendance and its consequences in choice experiments with latent class models.* Health Econ, 2013. 22(5): p. 554-67.
28. Scott, A., *Identifying and analysing dominant preferences in discrete choice experiments: an application in health care.* J Econ Psych, 2002. 23(3): p. 383-398.
29. Alemu, M.H., et al., *Attending to the reasons for attribute non-attendance in choice experiments.* Environ Resource Econ, 2013. 54(3): p. 333-359.
30. Ryan, M. and A. Bate, *Testing the assumptions of rationality, continuity and symmetry when applying discrete choice experiments in health care.* Appl Econ Lett, 2001. 8(1): p. 59-63.
31. Bech, M., T. Kjaer, and J. Lauridsen, *Does the number of choice sets matter? Results from a web survey applying a discrete choice experiment.* Health Econ, 2011. 20(3): p. 273-86.
32. Miguel, F.S., M. Ryan, and M. Amaya-Amaya, *'Irrational' stated preferences: a quantitative and qualitative investigation.* Health Econ, 2005. 14(3): p. 307-22.
33. Cheraghi-Sohi, S., et al., *Making sense of patient priorities: applying discrete choice methods in primary care using 'think aloud' technique.* Fam Pract, 2007. 24(3): p. 276-82.
34. Berkman, N.D., et al., *Low health literacy and health outcomes: an updated systematic review.* Ann Intern Med, 2011. 155(2): p. 97-107.
35. Sorensen, K., et al., *Health literacy and public health: A systematic review and integration of definitions and models.* BMC Public Health, 2012. 12(1): p. 80.
36. Rudd, R.E., *Health literacy skills of U.S. adults.* Am J Health Behav., 2007. 31(Suppl 1): p. S8-18.
37. HLS-EU Consortium, *Comparative report of health literacy in eight EU member states. The European health literacy survey HLS-EU.* 2013.
38. Kenny, P., et al., *Do participants understand a stated preference health survey? A qualitative approach to assessing validity.* Int J Technol Assess Health Care, 2003. 19(4): p. 664-81.
39. Ozdemir, S., et al., *Who pays attention in stated-choice surveys?* Health Economics, 2010. 19(1): p. 111-8.
40. de Bekker-Grob, E.W., et al., *Men's preferences for prostate cancer screening: a discrete choice experiment.* Br J Cancer, 2013. 108(3): p. 533-41.
41. Veldwijk, J., et al., *Parental preferences for rotavirus vaccination in young children: a Discrete Choice Experiment.* Vaccine, 2014. In Press.
42. Fransen, M.P., et al., *Applicability of internationally available health literacy measures in the Netherlands.* J Health Commun, 2011. 16 Suppl 3: p. 134-49.
43. Hensher, D.A., *How do respondents process stated choice experiments? Attribute consideration under varying information load.* J Appl Economet, 2006. 21(6): p. 861-878.
44. Waters, E.A., et al., *Formats for improving risk communication in medical tradeoff decisions.* J Health Com, 2006. 11(2): p. 167-82.
45. Murphy, T.V., et al., *Intussusception among infants given an oral rotavirus vaccine.* N Engl J Med, 2001. 344(8): p. 564-72.

ADDITIONAL FILE 1:

Description of both case studies

Rotavirus DCE

Between January and March 2013 Veldwijk et al. conducted a DCE on parental preferences for rotavirus vaccination of newborns in the Netherlands [1]. A random sample of 2500 parents with newborns was selected from the Praeventis database (a national vaccination register in which the vaccination status of all Dutch newborns is registered) to participate in this study. The DCE questionnaire consisted of nine choice tasks. In the choice tasks, participants were asked to choose between two different rotavirus vaccine scenarios to protect their child from an infection. Each choice task was constructed based on five attributes with either two or three levels (Table 1a). In total, 959 participants completed the questionnaire, of which 202 gave permission to be re-contacted for further research.

Prostate cancer screening DCE

To investigate men's preferences and trade-offs for prostate cancer screening, de Bekker-Grob et al. conducted a DCE between January and May 2011 among a population-based random sample of 1000 men aged 55 to 75, living in the Rijnmond region of the Netherlands [2]. The DCE questionnaire comprised 16 choice tasks, in which participants were asked to choose between a no screening scenario (opt-out) and two prostate cancer-screening scenarios. Each scenario consisted of five attributes, with each four levels (Table 1b). In total, 459 men responded to the questionnaire and 373 gave permission to be contacted again for additional questions.

References

1. Veldwijk, J., et al., *Parental preferences for rotavirus vaccination in young children: a Discrete Choice Experiment.* Vaccine, 2014. 32(47): p. 6277-83.
2. de Bekker-Grob, E.W., et al., *Men's preferences for prostate cancer screening: a discrete choice experiment.* Br J Cancer., 2013. 108(3): p. 533-41.

ADDITIONAL FILE 2:

Example of a choice tasks of both case studies

Imagine that a vaccine against rotavirus infections would become available within the Netherlands. In what situation would you prefer to vaccinate your newborn, situation 1 or situation 2?

	Situation 1	Situation 2
Effectiveness	75%	95%
Frequency of severe side effects	1 in 100,000	1 in 10,000
Protection duration	3 years	6 years
Location	Your General Practitioner	Child Welfare Center
Costs	€ 30	€ 0
Tick the box of the scenario that you prefer:	☐	☐

Thirty-five out of every 1000 deaths among men are caused by prostate cancer. Which alternative do you prefer to reduce your risk of dying from prostate cancer: no screening, screening program 1, or screening program 2? ***(please, tick one box)***

	NO SCREENING	PROGRAM 1	PROGRAM 2
Amount of men per 1.000 men who will **die** from prostate cancer	**35 deaths**  (0 deaths prevented)	**25 deaths**  (10 deaths prevented)	**18 deaths**  (17 deaths prevented)
Frequency of a blood **test**	**No** blood test	Every **4 years** a blood test	Every **3 years** a blood test
Amount of men per 1,000 men with an increased PSA who receive an **unnecessary biopsy** (= no cancer detected, although the blood test suggested that a blood test was necessary)	**Not applicable**	**400 unecessary biopsies**  (600 correct biopsies)	**800 unnecessary biopsies**  (200 correct biopsies)
Amount of men per 1,000 treated men who receive an **unnecessary treatment** (= no increase in life expectancy, but there is a risk of urine incontinence and erection problems due to treatment)	Not applicable	**0 unnecessary treatments**  (1,000 correct treatments)	**500 unnecessary treatments**  (500 correct treatments)
Out of pocket cost per year during the period of the screening program	0 euro per year	**100 euro per year**	**50 euro per year**
Which alternative would you choose?	☐	☐	☐

ADDITIONAL FILE 3:

Description of the health literacy measures

The prostate cancer screening cohort was asked to complete the three validated Dutch questions of the Set of Brief Screening Questions (SBSQ-D) of Chew [1] to measure their subjective health literacy. This instrument was already included in the initial rotavirus DCE, these questions were therefore not repeated in the current study. The SBSQ-D contains questions on how often participants need help to read letters from their GP/specialist, how sure participants are that they complete medical forms correctly and how often participants find it difficult to find information about their health. Participants scored these questions on a 5-point Likert scale, from zero to four. An average score of ≤2 indicates inadequate health literacy, while an average score >2 indicates adequate health literacy [1]. The Dutch version of the Newest Vital Sign (NVS-D) was included as an objective measure of health literacy [1] for both groups. To measure the participants' health literacy status, they were asked six questions about an ice cream nutrition label. Participants scored one point for each correctly answered question, with a maximum of 6 points. A score of 4-6 indicates adequate health literacy [1].

Reference

1. Fransen, M.P., et al., *Applicability of internationally available health literacy measures in the Netherlands.* J Health Commun, 2011. 16 Suppl 3: p. 134-49.

CHAPTER 5

PREFERENCES FOR VACCINATION: DOES HEALTH LITERACY MAKE A DIFFERENCE?

J. Veldwijk*, I. van der Heide*, J. Rademakers,
A.J. Schuit, G.A. de Wit, E. Uiters, M.S. Lambooij

* Equal contribution

In revision: Medical Decision Making.

ABSTRACT

Purpose: To examine to what extent health literacy is associated with parental preferences concerning childhood vaccination.

Methods: A cross-sectional study was conducted among 467 Dutch parents of newborns aged 6 weeks (response rate of 37%). A self-reported questionnaire was used to measure health literacy by means of Chew's Set of Brief Screening Questions, as well as parental preferences for rotavirus vaccination by means of a Discrete Choice Experiment. Five rotavirus related characteristics were included (i.e. vaccine effectiveness, frequency of severe side effects, location of vaccination and protection duration and out-of-pocket costs). Panel latent class models were conducted, health literacy and educational level were added to the class probability model to determine the association between health literacy and study outcomes.

Results: Lower educated and lower health literate respondents considered protection duration to be more important and vaccine effectiveness and frequency of severe side-effects to be less important compared to higher educated and health literate respondents. While all respondents were willing to vaccinate against rotavirus when the vaccine was offered as part of the National Immunization Program (NIP), only lower educated and lower health literate parents were willing to vaccinate when the vaccine was offered on the free market.

Conclusion: Health literacy is associated with parents' preferences for rotavirus vaccination. Whether differences in vaccination decisions are actually due to varying preferences or might be better explained by varying levels of understanding should be further investigated. To contribute to more accurate interpretation of study results, it may be advisable that researchers measure and report health literacy when they study vaccination decision behavior.

INTRODUCTION

Two developments in the public domain appear to affect acceptance of vaccines in the Netherlands. First, there is a tendency to stimulate citizens to exert autonomy and to make well-considered decisions with respect to their health [1], such as vaccination decisions. At the same time, the necessity of vaccination has become a more prominent part of the public discourse in the Netherlands and other Western countries [2-6]. With respect to for instance Human Papilloma Virus vaccination, part of the target population was found to be hesitant [6, 7], which may in part be due to the distrust of parents in the communication of governmental institutions about the vaccine [7]. However, for the prevention of most infectious diseases including rotavirus, vaccination of the majority of the population is essential to reach herd immunity [8, 9]. The Netherlands so far has had a highly efficient National Immunization Program (NIP), with vaccination coverage rates of around 96% among young children aged 0 to 9 years [10]. To keep Dutch childhood vaccination rates high, insight into factors that determine parents' decisions about childhood vaccination is crucial in order to develop or revise vaccination education strategies.

Several studies already investigated the influence of psychosocial determinants (e.g., attitude towards vaccination and perceived severity of the disease at hand) on vaccination decisions [7, 11-14]. There is also a growing body of literature describing the effect of specific vaccine aspects such as vaccine safety and efficacy on parental preferences for vaccination and ultimately also their willingness to vaccinate [14-22]. Parental preferences are increasingly being elicited in order to guide policy measures such as the introduction of new vaccines, and may serve as a starting point for communication strategies when the vaccine is introduced [23].

Parental characteristics such as educational attainment is associated with the valuation of vaccine characteristics [13, 15-18]. A concept that is related to educational attainment and has received increasing attention in the field of prevention, is health literacy [24]. Health literacy skills, the ability to access, understand, appraise and apply health related information, are likely to influence parental preferences for vaccination as well, although such studies are currently lacking [25-28]. At the same time health literacy is considered an important set of cognitive, psychological and social skills in contemporary society [29], given the tendency that citizens are stimulated to exert autonomy and to make well-considered decisions with respect to their health [1]. Health literacy is highly relevant in the light of measuring vaccination preferences, since measuring preferences usually requires respondents to interpret and value risk information (e.g., risk of side effects). Previous research shows that specifically individuals with a lower educational level and a lower health literacy have difficulties processing such information [30-32]. Misinterpreting information or being unable to understand information probably affects parental preferences for vaccination and thereby their willingness to vaccinate their newborn [33].

Therefore, the current study aims to determine to what extent health literacy is associated with parental preferences for vaccination. Rotavirus vaccination served as a case for this study since this vaccine is currently considered for inclusion in the Dutch NIP. Rotavirus is the most common cause

of severe acute gastroenteritis in infants and young children worldwide [34] and can be prevented by vaccination of 6 to 10 weeks old infants, as is recommended by WHO [35].

MATERIALS AND METHODS

Sample Selection

This study is part of a larger study investigating parental preferences for rotavirus vaccination [14]. The target population was identified via Praeventis, which is a national register that registers the vaccination status of all Dutch newborns. A random selection of the parents of 1250 newborn babies aged six weeks received a questionnaire, to end up with a minimum of 250 completed questionnaires, based on a 20%-30% response rate. Due to confidentiality agreements with Praeventis, no reminder letters could be sent. Hence, we were unable to monitor whether response was selective (selection bias) and no non-response information could be gathered. The Institutional Review Board of the University Medical Centre Utrecht advised that formal testing by a medical ethical committee was not necessary, as parents were only required to complete an anonymous questionnaire once, which is in accordance with the guidelines laid down in the Declaration of Helsinki [36].

Questionnaire

The questionnaire was administered in Dutch and consisted of two parts. First, different demographic factors were included in the questionnaire among which, parents' age in years, gender, ethnicity (Dutch versus non Dutch) and highest attained level of education. Self-reported highest attained level of education was categorized into four categories: no or primary education, lower secondary education, upper secondary or vocational education, or tertiary education (bachelor's degree or higher). In this part, also health literacy was measures (see next subsection for details). Parental attitude towards vaccination and their intention to vaccinate their newborn were measured using statements. Parents could rate these statements on a five-point Likert scale ranging from totally agree to totally disagree. Attitude was measured by two items: 'I think vaccination is a good way to protect my child against rotavirus' and 'I think it is obvious to vaccinate my child against rotavirus'. Intention to vaccinate was used as a proxy for health action and was measured by one item 'I would vaccinate my child against the rotavirus when a vaccine would become available'. Second, parental preferences for rotavirus vaccination were measured.

Assessment of health literacy

Health literacy was assessed by Chew's Set of Brief Screening Questions (SBSQ), which is a validated subjective measure of health literacy containing three items [37-39]. The SBSQ provides a feasible and reliable indication of those who are likely to have lower health literacy skills [39]. Respondents were asked how often someone helps them to read letters from their GP/specialist, how sure they are that they complete medical forms correctly and how often they find it hard to learn more about their health because written information is not well understood. Responses were scored on a 5-point Likert scale ranging from zero to four. Respondents' sum scores and mean scores over the

three included items were calculated. An average score of ≤2 indicates inadequate health literacy, while an average score >2 indicates adequate health literacy.

Assessment of parental preferences

A discrete choice experiment was used to measure the preferences of parents for specific characteristics of rotavirus vaccination.

Discrete Choice Experiment

Discrete Choice Experiments (DCEs) are increasingly being used to determine the relative importance of different intervention characteristics [23, 40, 41]. The Random Utility Theory is the basis of this method, which assumes that any intervention can be described by its characteristics or 'attributes' (such as vaccine effectiveness). The individual's preference for an intervention is determined based on the levels (e.g., effectiveness of 50% versus 80% versus 95%) of those attributes [23, 40, 41]. Scenarios are constructed by varying the levels of the attributes. Respondents are provided with a series of 'choice tasks' that consist of at least two scenarios. They have to choose the scenario they prefer most within every choice task.

Attributes, levels and experimental design

To construct the current DCE, attributes and levels were identified in a stepwise manner. First, a list of barriers and facilitators for vaccinating children reported by both children and parents was compiled based on previously published literature [15, 17, 18, 20-22, 42, 43]. Second, the obtained list of barriers and facilitators was discussed with a pediatrician with specific interest in rotavirus infections and a scientist with specific interest in vaccination behavior. Based on these interviews the list of potential attributes was shortened and it was ensured that the potential attributes and levels were consistent with current practice. Third, four group interviews with 25 parents of newborns were conducted. Based on these interviews the most important attributes for the decision-making process of parents about whether to vaccinate their newborn against rotavirus were selected. Moreover, proper levels were appointed to each of the attributes.

For this DCE five attributes were selected (Table 1). The first attribute is vaccine effectiveness, which describes the percentage of children that will be protected against a rotavirus infection when vaccinated. Three levels were used for this attribute, namely 55%, 75% and 95% [44, 45]. The second attribute was the frequency of severe side-effects, which describes the number of vaccinated children that will suffer from intussusception due to vaccination. Intussusception is an acute condition in which part of the bowel telescopes into another adjacent part of the bowel, resulting in obstruction [46]. This number could be 1 in 10,000 children, 1 in 100,000 children and 1 in 1,000,000 children [47, 48]. The third attribute was the protection duration of the vaccine, which describes the number of years that the vaccine protects against a rotavirus infection. This duration was set at 1 year, 3 years or 6 years [44, 45]. The fourth attribute was health care facility. Within the Netherlands, all vaccines in the NIP are administered at a child welfare center, but the general practitioner (GP)

office was included because the rotavirus vaccine may not become part of the NIP; in that case it is likely that this vaccine is administered here. The final attribute was out-of-pocket costs, which was varied between €0, €30 and €140 for complete vaccination [49].

Table 1. Attributes and levels that were included in the DCE

Attributes	Level 1	Level 2	Level 3
Vaccine effectiveness			
The percentage of children that will be protected against a rotavirus infection when vaccinated			
	55%	75%	95%
Frequency of severe side effects			
The number of vaccinated children that will suffer from intussusception due to vaccination. Intussusception is an acute condition in which part of the bowel telescopes into another adjacent part of the bowel, resulting in obstruction.			
	1 in 10,000	1 in 100,000	1 in 1,000,000
Protection duration			
The number of years that the vaccine protects against a rotavirus infection			
	1 year	3 years	6 years
Healthcare facility of vaccine administration			
Within the Netherlands all vaccines in the NIP are administrated at a child welfare center, The GP office was included because the rotavirus vaccine may not become part of the NIP; in that case it is likely that this vaccine is administrated at the GP.			
	Child welfare center	General practitioner	
Out-of-pocket costs			
Parents may have to pay (part) of the vaccine costs out-of-pocket			
	€0	€30	€140

Testing all possible combinations of attributes and levels (full factorial design), would imply that 162 (34 * 21) different scenarios should be tested. Due to obvious methodological (bias) and cognitive (burden on participants) reasons, not all these scenarios were included. Using NGene 1.0 (ChoiceMetrics, 2011) software a D-efficient design was developed, which optimizes the variance-covariance matrix [50, 51]. The software was instructed to create a design using a panel mixed-multinomial-logit model, including beta priors from the pilot study, 250 Halton draws and 500 repetitions. It was assumed that there would be no interaction between attributes, while level balance and minimal overlap between attribute levels were optimized. The final design consisted of 18 unique choice tasks. These were divided into two sets of nine choice tasks, and each set of nine choice tasks was randomly distributed among half of the study population. Before participants were asked to complete these choice tasks, they received detailed information on the meaning of all attributes and levels as well as an explanation on how to complete a choice task, illustrated by an example. Every choice task started with the question: 'Imagine that there is a vaccine available against rotavirus infection, in which situation would you prefer to vaccinate your newborn, situation 1 or situation 2? After each choice task participants were asked whether they

would make the same choice in real life or whether they would rather not vaccinate their newborn (opt-out). This opt-out option was included because, in real life, parents can also choose not to have their child vaccinated [52]. For further details on this procedure see Veldwijk et al. [14].

Statistical analyses

Data were analyzed using panel latent class models [53-55]. Such models take into account the multilevel structure of our data (i.e. every respondent answered nine choice sets). By means of this latent class model it can be determined whether different preferences exist across unobserved subgroups of the population. Class membership is not assigned by researchers, but is latent, so each respondent has a certain probability to belong to a class. However, demographic measures can be incorporated into the modelling procedure, which provides some insights about which respondents belong to what class.

Respondents with >10% missing answers on their choice tasks were excluded from the analysis (n=12). All attributes were tested for linearity. All non-linear attributes were recoded using effect codes [56]. This coding procedure codes the reference category as -1 and the sum of the effect coded attribute levels is always zero [56, 57]. Based on model fit tests (AIC, Log likelihood) it was tested which model was most suitable for our data, and how many classes could be identified within the data. This resulted in a two-class model based on the utility equation displayed below.

$$V_{rta|c} = \beta_{0|c} + \beta_{1|c}\,\text{vaccine effectiveness}_{\,rta|c} + \beta_{2|c}\,\text{severe side effects}_{\,1\text{ in }100{,}000\ rta|c} + \beta_{3|c}\,\text{severe side effects}_{\,1\text{ in }10{,}000\ rta|c} + \beta_{4|c}\,\text{protection}_{\,3\text{ years }rta|c} + \beta_{5|c}\,\text{protection}_{\,6\text{ years }rta|c} + \beta_{6|c}\,\text{location}_{\,\text{Child Welfare Center }rta|c} + \beta_{7|c}\,\text{out-of-pocket costs}_{\,€30\ rta|c} + \beta_{8|c}\,\text{out-of-pocket costs}_{\,€140\ rta|c}$$

The systematic utility component (V) describes the measurable utility that respondent 'r' belonging to class 'c' reported for alternative 'a' in choice task 't'.

The β_0 represents the alternative specific constant and $\beta_1 - \beta_8$ are the attribute level estimates that indicate the relative importance of each attribute level. A significant attribute estimate within a certain class indicates that this attribute contributes to the decision-making procedure of respondents that belong to that class. The utility for the opt-out option was modelled as zero. In addition to the above-specified utility function, a class assignment model was fitted. All demographic measures were tested for a significant contribution to the class assignment mode, and the final class assignment utility function was:

$$V_{rc} = \beta_{0|c} + \beta_{1|c}\,\text{health literacy mean score}_{r} + \beta_{2|c}\,\text{high educational level}_{r}$$

A significant demographic variable indicates that this variable contributes to the class assignment (e.g., if the beta of the health literacy mean score variable is positive and significant for class 1, respondents with a higher health literacy mean score are more likely to belong to class 1).

Due to the unidentifiable scaling factor that is always present in DCE data [57, 58], the attribute level estimates cannot be compared directly between the two classes. Therefore, importance weights were calculated based on the results of the panel latent class models, separately for both classes. Per attribute, the difference between the highest and lowest attribute level estimate was calculated. The largest difference value received an importance score of one, representing the attribute that was deemed most important by respondents, the other difference values were divided by the largest difference value, resulting in a relative distance of all attributes to the most important attribute. Since these values can only be interpreted relatively to each other, the recoded values cannot be tested for statistical differences.

Situation specific utility scores were calculated for both classes separately. A realistic vaccine situation was used were the vaccine was expected to have a 95% effectiveness, a 1 in 100,000 frequency of severe side effects, and a protection duration of 3 years. Scores were calculated separately for a scenario in which the vaccine would be implemented within a NIP (i.e. free of charge and administered at a child welfare center) or on the free market (i.e. €140 out-of-pocket costs and administered at the general practitioner). A positive utility score that is larger than 0 implies that respondents prefer vaccination over no vaccination, while a utility below zero implies that respondents prefer not to vaccinate their newborn.

RESULTS

Sample characteristics

In total 467 out of 1,250 questionnaires were returned and included in the analyses (response rate of 37.4%). The mean age of the respondents was 31 years and most respondents were female (82%) (Table 2). The majority of the responders had a Dutch origin (92%), and attained a tertiary educational level (58%). On average respondents reported a positive attitude towards rotavirus vaccination, since 77% reported that they thought vaccination is a good way to protect their child against rotavirus and 90% thought it was obvious to vaccinate their newborn against rotavirus (Table 2). In addition, about 79% of the parents intended to vaccinate their child against rotavirus if this vaccine would become available (Table 2).

Health literacy

The majority of the respondents indicated that they never need any help reading letters and leaflets from their GP or the hospital (84%), they are very/fairly certain that they fill in medical forms correctly (93%) and that they never find it difficult to learn more about their health because they do not understand written information (64%) (Table 3). The internal consistency of the SBSQ (Cronbach's alpha = 0.61) was comparable to that found in the validating study of Franssen et al. [39]. The mean sum score of the respondent over these three items was 10.7 (min. 0, max. 12), resulting in a mean score of 3.6 (min. 0, max. 4). Lower educated respondents had a significantly lower mean health literacy score compared to high-educated respondents (3.4 (SD 0.55) and 3.7 (SD 0.37) respectively, $t = 8.16$, $p<0.05$).

Table 2. Sample characteristics (N= 466)

		Mean (SD)	Percentage
Age (in years)		31.2 (5.1)	
Gender	Female		81.5
Ethnicity	Dutch		91.8
Educational level	No or primary		0.4
	Lower secondary		7.8
	Upper secondary or vocational education		33.7
	High tertiary		58.0
Statements *			
Attitude	I think vaccination is a good way to protect my child against rotavirus		76.6
	I think it is obvious to vaccinate my child against rotavirus		89.5
Intention	I would vaccinate my child against the rotavirus when a vaccine would become available		79.4

* Proportion of respondents who (totally) agreed with these statements

Table 3. Health literacy scores, both combined and separate per item of the SBSQ*

Health literacy scores				
	Mean (SD)	Range		
Sum score over 3 items	10.7 (1.5)	3-12		
Mean score of 3 items	3.6 (0.5)	1-4		
Distributions of respondents answers to the three items separately				
	How often does someone help you to read letters and leaflets from your GP** or the hospital?	How often is it difficult for you to learn more about your health because you do not understand written information?		How certain are you that you fill in medical forms correctly?
Never	83.9	64.2	Very much	46.7
Now and then	9.4	26.6	Fairly	46.7
Sometimes	4.9	8.8	A bit	5.6
Often	1.7	0.4	A little bit	0.9
Always	0.0	0.0	Not at all	0.2

* SBSQ is the Set of Brief Screening Questions to measure health literacy, ** General Practitioner

Parental preferences

Two classes were identified in the panel latent class model (Table 4). All attributes were significant and influenced parental preferences for rotavirus vaccination as expected. In both classes, parents were more willing to vaccinate if vaccine effectiveness increased. Moreover, parents preferred the lowest frequency of severe side effects, a protection duration of three years, vaccine administration via a child welfare center and lowest out-of-pocket costs.

Table 4. The association between health literacy and parental preferences for rotavirus vaccination based on the panel latent class analysis†.

		Class 1		Class 2	
		Estimate	**SE**	**Estimate**	**SE**
Constant		-1.91***	0.26	-6.02***	0.36
Vaccine effectiveness		0.53***	0.03	0.54***	0.04
Frequency of severe side effects	1 in 1,000,000 (ref)	0.42***	0.08	0.80***	0.09
	1 in 100,000	-0.08	0.06	-0.16*	0.09
	1 in 10,000	-0.34***	0.09	-0.64***	0.11
Protection duration	1 year (ref)	-0.83***	0.06	-0.64***	0.09
	3 years	0.72***	0.09	0.40***	0.09
	6 years	0.11	0.07	0.24**	0.11
Healthcare facility	General Practitioner (ref)	-0.05	0.04	-0.12**	0.06
	Child Welfare Center	0.05	0.04	0.12**	0.06
Out-of-pocket costs	€0 (ref)	0.59***	0.06	1.12***	0.10
	€30	0.53***	0.07	0.18*	0.11
	€140	-1.12***	0.06	-1.30***	0.10
Class probability model					
Constant		2.82***	0.95	-	-
Health literacy score		-0.57**	0.28	-	-
Higher education		-0.76***	0.26	-	-
Average class probability		0.58		0.42	

† Due to the effects coding procedure, the estimate of the reference category can be calculated as -1 * (the sum of the other estimates within the same attribute). The standard error was calculated using the delta method.*** P<0.01, ** P<0.05, * P<0.10.

The average class probabilities were 0.58 and 0.42 for class 1 and 2 respectively. The probability of belonging to either class 1 or 2, was also dependent upon health literacy score and educational level (Table 5). The model that included health literacy in addition to educational level significantly improved the model fit (loglikelihood = -3062 and -3065 respectively, $P < 0.05$).

Respondents with a higher educational level or a higher health literacy score were more likely to belong to class 2, while the probability to belong to class 1 increased when the health literacy score or educational level decreased. Therefore, respondents with a lower health literacy score or educational level were more likely to belong to class 1. The probability of respondents with a lower health literacy to belong to class 1 increased if those respondents also attained a lower educational level and decreased if those respondents attained a higher educational level.

Table 5. Probabilities of respondents belonging to either class 1 or 2 of the latent class model based on their educational level and health literacy score*.

	Probability to belong to class 1	**Probability to belong to class 2**
High educational level		
Health literacy score = 4	0.45	0.55
Health literacy score = 3	0.58	0.41
Health literacy score = 2	0.71	0.28
Health literacy score = 1	0.81	0.18
Health literacy score = 0	0.88	0.11
Lower educational level		
Health literacy score = 4	0.63	0.37
Health literacy score = 3	0.75	0.24
Health literacy score = 2	0.84	0.16
Health literacy score = 1	0.90	0.10
Health literacy score = 0	0.94	0.06

* High education was defined as tertiary education, while all other education levels were conceptualized as lower educational level. Health literacy mean scores range between 0-4, a score between 0-2 indicates poor health literacy whereas a score ≥2 indicates adequate health literacy.

Respondents that belong to either class 1 or 2 reported different preferences with respect to rotavirus vaccination, indicating considerable preference heterogeneity (Figure 1). Respondents in both class 1 and 2 value out-of-pocket costs as most important (importance weight of 1) and health care facility as least important (lowest importance weight). Respondents in class 1 value protection duration as relatively more important compared to respondents belonging to class 2, while respondents in class 2 value vaccine effectiveness and frequency of severe side effects to be more important compared to respondents in class 1. Thus, respondents with a lower educational level and respondents with lower health literacy skills considered protection duration to be more important and vaccine effectiveness and frequency of severe side-effects to be less important compared to respondents with a high education and respondents with higher health literacy skills.

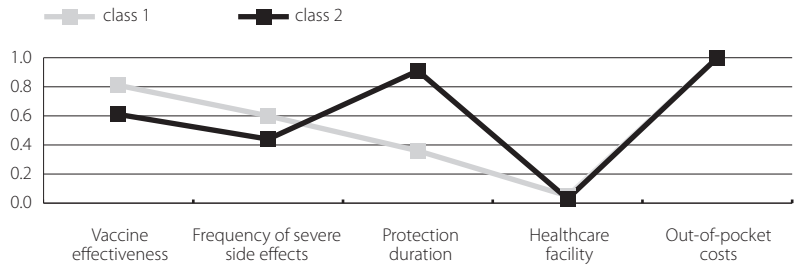

Figure 1. Importance weights of the attributes stratified by class*.

* Importance weights reflect the relative distance of all attributes to the most important attribute on a scale from 0-1 (1 indicating the most important attribute).

Vaccine specific utility scores

The utility scores for a realistic rotavirus vaccine (i.e. a 95% effectiveness, a frequency of severe side effects of 1 in 100,000 and a protection duration of 3 years) implemented within a NIP (i.e. free of charge and administered at the CWC) were 4.1 for class 1 (lower educated and lower health literacy skills) and 0.1 for class 2 (high educated and higher health literacy skills). These scores implicate that respondents in both classes would vaccinate their newborn against the rotavirus in this particular scenario (i.e. utility scores are larger than 0). However, if the same vaccine would be implemented on the free market (i.e. €140 out-of-pocket costs and administered at the GP), the utility scores of class 1 (lower educated and lower health literacy skills) and class 2 (high educated and higher health literacy skills) would be 2.3 and -2.6, respectively. In this situation, respondents belonging to class 1 (lower educated and lower health literacy skills) would still prefer to vaccinate their newborn against rotavirus, while respondents in class 2 (high educated and higher health literacy skills) would not.

DISCUSSION

Health literacy is associated with parental preferences concerning rotavirus vaccination. Current study results indicate that parents with lower health literacy skills and parents with a lower educational level value the protection duration of a vaccine as more important compared to parents with higher health literacy skills and parents with a higher education. Moreover, vaccine effectiveness and the frequency of severe side effects were perceived more important by parents with higher health literacy skills and educational level compared to parents with lower health literacy skills and educational level. Irrespective of health literacy and educational level, out-of-pocket costs were revealed to be the most important characteristic in the decision about rotavirus vaccination, while the health care facility where the vaccine would be administrated was least important in this decision making process.

Of particular interest is the finding that lower health literate parents valued vaccine effectiveness and the frequency of severe side effects as less important compared to higher health literate parents when deciding about vaccinating their newborn against rotavirus. Both these vaccine

characteristics were included in the DCE as a numerical value and depicted as a percentage and absolute frequency respectively. Such information is difficult to interpret, especially for parents with lower health literacy skills [31, 32, 59-61]. If parents had difficulties interpreting and understanding these risk attributes, they might deem those attributes as less important. This is in line with previous research indicating that if information is less well understood it is more likely to be neglected [33], or otherwise undervalued. Therefore, the difference in preference concerning vaccine effectiveness and the frequency of severe side effects between respondents with lower and higher health literacy skills might not be caused by actual differences in preference structures, but rather reflect a lack of understanding. This is of particular importance for all future vaccine-related DCE studies, since they are likely to include attributes such as vaccine effectiveness and frequency of side effects. By means of newly designed research methods, it should be explored if preferences of respondents with lower and higher health literacy skills still differ if respondents interpret the numerical value of risks as similar as possible. Such research will reveal to what extent a decision-making process is influenced by understanding certain vaccine characteristics. In addition, future research is required to explore if DCE's that are conducted online with the use of verbal support or in face-to-face settings might overcome issues with the understanding of vaccine characteristics. These considerations raise the question to what extent a DCE is a valid method to obtain insight into vaccination preferences among individuals with lower health literacy skills. However, in real life situations people with lower health literacy skills have to deal with complex information as well and may base their preferences on information that is not well understood.

Our findings showed that parents with higher health literacy were less likely to participate in any of the tested vaccination scenarios. This was mostly due to the large negative constant that was found for this class. The limited knowledge base with respect to health literacy and participation in vaccination programs suggests the opposite (i.e. lower health literacy is associated with lower participation rates) [25-27]. Besides differences in preferences for vaccine characteristics or differences in how well these characteristics were understood or how they were perceived, a possible interpretation of this finding may be derived from literature on the association between socio-economic position and vaccination participation. Although the concept of health literacy is not equal to socio-economic position, they are closely related. Literature suggests that individuals with a higher socio-economic position are more critical towards childhood vaccination, which could lead to higher vaccination hesitance and lower participation rates [1, 62]. It might therefore be that parents with higher health literacy are less likely to vaccinate their newborn against rotavirus because of more critical thoughts about rotavirus vaccination.

The present study showed that health literacy influenced study outcomes irrespective of educational level. Both factors influenced the results in a similar direction and they independently contributed to the class assignment models and all subsequent results. Moreover, health literacy significantly improved the model fit when added to the class assignment model. Health literacy can therefore not be adjusted for by only including educational level in future research on vaccination decisions. Health literacy should be measured and reported as a sample characteristic in all future research investigating vaccination decisions. This will lead to a more accurate interpretation of study results.

This study is subject to some limitations. Firstly, although a response rate of 37.4% is relatively high for a postal questionnaire without reminders [63-66], selective non-response seems plausible. Our sample shows relative many highly educated parents compared to Dutch national figures [67]. Furthermore, the number of non-Dutch parents in our study population is relatively low compared to the general population of the Netherlands [67]. Generalizability of our results to preferences of non-Dutch parents therefore may be limited. Secondly, although the applied measure of health literacy, the SBSQ, currently is the only short questionnaire validated in Dutch that can be assessed in writing, it provides a limited and subjective measure of functional health literacy. Compared to other (objective) measures of health literacy, the SBSQ may lead to an underestimation of the number of parents with lower health literacy [39]. Moreover, using a more comprehensive measure of health literacy, assessing the ability to critically judge information and apply it in various circumstances, may help explain differences in preferences based on difficulty with judging information. Therefore, our study may underestimate the true effect of health literacy on parental preferences and vaccination decisions. Thirdly, within a DCE respondents are asked to interpret and value attributes by comparing two or more options, while in real life, people are presented with only one option. Therefore, the interpretation and weighing of attributes in real life might be more complicated. Subsequently, a DCE can provide insight into conditions under which people are likely to choose for rotavirus vaccination, but in real life decisions are being affected by factors such as previous experiences and social influences as well [2, 3, 63, 68, 69]. Although DCE's provide an indication under what conditions people are likely to choose for a certain product, in this case vaccination, it remains unsure what decisions are made in real life.

This study shows that health literacy is associated with parents' preferences for rotavirus vaccination. When offered on the free market, parents with higher health literacy may be less likely to vaccinate their newborn against rotavirus than parents with lower health literacy. The results of this study call for health literacy as an important factor to take into account when studying vaccination behavior. Whether differences in vaccination decisions are actually due to varying preference structures or might be better explained by varying levels of understanding should be further investigated. Altogether, it may be advisable that researchers measure and report health literacy when they study vaccination decision behavior.

References

1. Peretti-Watel, P., et al., *Attitudes toward vaccination and the H1N1 vaccine: Poor people's unfounded fears or legitimate concerns of the elite?* Soc Sci Med, 2014. 109: p. 10-8.
2. Hak, E., et al., *Negative attitude of highly educated parents and health care workers towards future vaccinations in the Dutch childhood vaccination program.* Vaccine, 2005. 23(24): p. 3103-7.
3. Paulussen, T.G., et al., *Determinants of Dutch parents' decisions to vaccinate their child.* Vaccine, 2006. 24(5): p. 644-51.
4. Larson, H.J., et al., *Addressing the vaccine confidence gap.* Lancet, 2011. 378(9790): p. 526-35.
5. Shetty, P., *Experts concerned about vaccination backlash.* Lancet, 2010. 375(9719): p. 970-1.
6. Rondy, M., et al., *Determinants for HPV vaccine uptake in the Netherlands: A multilevel study.* Vaccine, 2010. 28(9): p. 2070-5.
7. Gefenaite, G., et al., *Comparatively low attendance during Human Papillomavirus catch-up vaccination among teenage girls in the Netherlands: Insights from a behavioral survey among parents.* BMC Public Health., 2012. 12:498.
8. Anderson, R.M. and R.M. May, *Immunisation and herd immunity.* Lancet, 1990. 335(8690): p. 641-5.
9. Patel, M.M., et al., *Real-world impact of rotavirus vaccination.* Pediatr Infect Dis J, 2011. 30(1 Suppl): p. S1-5.
10. van 't Schurink-van 't Klooster, T.M. and H.E. De Melker, *The National Immunisation Programme in the Netherlands: Surveillance and developments in 2013-2014.* 2014, National Institute for Public Health and the Environment: Bilthoven.
11. Harmsen, I.A., et al., *Psychosocial determinants of parents' intention to vaccinate their newborn child against hepatitis B.* Vaccine, 2012. 30: p. 4771-4777.
12. Paulussen, T.G.W., et al., *Determinants of Dutch parents' decisions to vaccinate their chils.* Vaccine, 2006. 24: p. 644-651.
13. Prislin, R., et al., *Immunization status and sociodemographic characteristics: the mediating role of beliefs, attitudes, and perceived control.* Am J Public Health., 1998. 88(12): p. 1821-6.
14. Veldwijk, J., et al., *Parental preferences for rotavirus vaccination in young children: a Discrete Choice Experiment.* Vaccine, 2014. 32(47): p. 6277-83.
15. Hall, J., et al., *Using stated preference discrete choice modelling to evaluate the introduction of varicella vaccination.* Health Econ, 2002. 11(5): p. 457-65.
16. Brown, D.S., et al., *Mothers' preferences and willingness to pay for vaccinating daughters against human papillomavirus.* Vaccine, 2010. 28(7): p. 1702-8.
17. Bishai, D., et al., *Conjoint analysis of French and German parents' willingness to pay for meningococcal vaccine.* PharmacoEconomics, 2007. 25(2): p. 143-54.
18. Poulos, C., et al., *Mothers' preferences and willingness to pay for HPV vaccines in Vinh Long Province, Vietnam.* Soc Sci Med., 2011. 73(2): p. 226-34.
19. Determann, D., et al., *Acceptance of vaccinations in pandemic outbreaks: a discrete choice experiment.* PLoS One., 2014. 9(7): p. e102505.
20. de Bekker-Grob, E.W., et al., *Girls' preferences for HPV vaccination: a discrete choice experiment.* Vaccine, 2010. 28(41): p. 6692-7.
21. Flood, E.M., et al., *A survey of children's preferences for influenza vaccine attributes.* Vaccine., 2011. 29(26): p. 4334-40.
22. Flood, E.M., et al., *Parent preferences for pediatric influenza vaccine attributes.* Clin Pediatr (Phila). 2011. 50(4): p. 338-47.
23. Clark, M.D., et al., *Discrete Choice Experiments in Health Economics: A Review of the Literature.* Pharmacoeconomics, 2014. 9: p. 9.
24. Berkman, N.D., et al., *Low health literacy and health outcomes: an updated systematic review.* Ann Intern Med, 2011. 155(2): p. 97-107.
25. Scott, T.L., et al., *Health literacy and preventive health care use among Medicare enrollees in a managed care organization.* Med Care, 2002. 40(5): p. 395-404.
26. Howard, D.H., T. Sentell, and J.A. Gazmararian, *Impact of health literacy on socioeconomic and racial differences in health in an elderly population.* J Gen Intern Med, 2006. 21(8): p. 857-61.

27. Bennett, I.M., et al., *The contribution of health literacy to disparities in self-rated health status and preventive health behaviors in older adults.* Ann Fam Med, 2009. 7(3): p. 204-11.

28. Sorensen, K., et al., *Health literacy and public health: A systematic review and integration of definitions and models.* BMC Public Health, 2012. 12(1): p. 80.

29. Kickbusch, I. and D. Nutbeam, *Advancing health literacy: a global challenge for the 21st century.* Health Promot Int, 2000. 15(3): p. 183-184.

30. Galesic, M. and R. Garcia-Retamero, *Statistical numeracy for health: a cross-cultural comparison with probabilistic national samples.* Arch Intern Med., 2010. 170(5): p. 462-8.

31. McCaffery, K.J., et al., *The influence of graphic display format on the interpretations of quantitative risk information among adults with lower education and literacy: a randomized experimental study.* Med Decis Making., 2012. 32(4): p. 532-44.

32. Waters, E.A., et al., *Formats for improving risk communication in medical tradeoff decisions.* J Health Commun., 2006. 11(2): p. 167-82.

33. Scheibehenne, B., R. Greifeneder, and P. Todd, *Can There Ever Be Too Many Options? A Meta-Analytic Review of Choice Overload.* J Cons Res 2010. 37(3): p. 409-25.

34. Parashar, U.D., et al., *Global illness and deaths caused by rotavirus disease in children.* Emerg Infect Dis, 2003. 9(5): p. 565-72.

35. Patel, M.M., et al., *Fulfilling the promise of rotavirus vaccines: how far have we come since licensure?* Lancet Infect Dis, 2012. 12(7): p. 561-70.

36. WMA. *Declaration of Helsinki - Ethical Principles for Medical Research Involving Human Subjects.* [cited 2015]. Available from: http://www.wma.net/en/30publications/10policies/b3/.

37. Chew, L.D., K.A. Bradley, and E.J. Boyko, *Brief questions to identify patients with inadequate health literacy.* Fam Med, 2004. 36(8): p. 588-94.

38. Chew, L.D., et al., *Validation of screening questions for limited health literacy in a large VA outpatient population.* J Gen Intern Med, 2008. 23(5): p. 561-6.

39. Fransen, M.P., et al., *Applicability of internationally available health literacy measures in the Netherlands.* J Health Commun, 2011. 16 Suppl 3: p. 134-49.

40. Ryan, M., K. Gerard, and M. Amaya-Amaya, *Using Discrete Choice Experiments to Value Health and Health Care.* The Economics of Non-Market Goods and Resources, ed. I.J. Bateman. 2008, Dordrecht: Springer.

41. Lancsar, E. and J. Louviere, *Conducting discrete choice experiments to inform healthcare decision making: a user's guide.* Pharmacoeconomics, 2008. 26(8): p. 661-77.

42. Brown, D.S., et al., *Mothers' preferences and willingness to pay for vaccinating daughters against human papillomavirus.* Vaccine., 2010. 28(7): p. 1702-8.

43. Stockwell, M.S., et al., *The effects of vaccine characteristics on adult women's attitudes about vaccination: a conjoint analysis study.* Vaccine., 2011. 29(27): p. 4507-11.

44. Rha, B., et al., *Effectiveness and impact of rotavirus vaccines in the United States - 2006-2012.* Expert Rev Vaccines, 2014.

45. Soares-Weiser, K., et al., *Vaccines for preventing rotavirus diarrhoea: vaccines in use.* Cochrane Database Syst Rev, 2012. 11: p. CD008521.

46. Murphy, T.V., et al., *Intussusception among infants given an oral rotavirus vaccine.* N Engl J Med., 2001. 344(8): p. 564-72.

47. Tate, J.E., et al., *Research priorities regarding rotavirus vaccine and intussusception: a meeting summary.* Vaccine, 2012. 30 Suppl 1: p. A179-84.

48. Weintraub, E.S., et al., *Risk of intussusception after monovalent rotavirus vaccination.* N Engl J Med., 2014. 370(6): p. 513-9.

49. Parez, N., et al., *Rotavirus vaccination in Europe: drivers and barriers.* Lancet Infect Dis., 2014. 14(5): p. 416-25.

50. Bliemer, M.C.J. and J.M. Rose, *Efficiency and sample size requirements for stated choice experiments, in Transportaion Research Broad Annual Meeting.* 2009: Washington DC.

51. Huber, J. and K. Zwerina, *The Importance of Utility Balance in Efficient Choice Designs.* J Market Res, 1996. 33(3): p. 307-317.

52. Veldwijk, J., et al., *The effect of including an opt-out option in discrete choice experiments.* PLoS One, 2014. 9(11): p. e111805.

53. Greene, W.H. and D.A. Hensher, *A latent class model for discrete choice analysis: contrasts with mixed logit.* Transport Res B Meth 2003. 37: p. 681-98.

54. Hess, S., et al., *Advantages of latent class over continuous mixture of Logit models.* 2011, Institute for Transport Studies, University of Leeds Leeds.

55. Swait, J., *A structural equation model of latent segmentation and product choice for cross-sectional revealed preference choice data.* J Retail Cons Service 1994. 1(2): p. 77-89.

56. Bech, M. and D. Gyrd-Hansen, *Effects coding in discrete choice experiments.* Health Econ, 2005. 14(10): p. 1079-83.

57. Hensher, D.A., J.M. Rose, and W.H. Greene, *Applied Choice Analysis: A Primer.* 2005, New York: Cambridge University Press.

58. Louviere, J.J., D.A. Hensher, and J.D. Swait, *Stated Choice Methods; analysis and application.* 2000, Cambridge: Cambridge University Press.

59. Galesic, M., R. Garcia-Retamero, and G. Gigerenzer, *Using icon arrays to communicate medical risks: overcoming low numeracy.* Health Psychol., 2009. 28(2): p. 210-6.

60. Dolan, J.G. and S. Iadarola, *Risk communication formats for low probability events: an exploratory study of patient preferences.* BMC Med Inform Decis Mak., 2008. 8(14).

61. Hildon, Z., D. Allwood, and N. Black, *Impact of format and content of visual display of data on comprehension, choice and preference: a systematic review.* Int J Qual Health Care., 2012. 24(1): p. 55-64.

62. Borjesson, M. and A. Enander, *Perceptions and sociodemographic factors influencing vaccination uptake and precautionary behaviours in response to the A/H1N1 influenza in Sweden.* Scand J Public Health, 2014. 42(2): p. 215-22.

63. Harmsen, I.A., et al., *Psychosocial determinants of parents' intention to vaccinate their newborn child against hepatitis B.* Vaccine, 2012. 30(32): p. 4771-7.

64. Hontelez, J.A., et al., *Parental attitude towards childhood HBV vaccination in The Netherlands.* Vaccine, 2010. 28(4): p. 1015-20.

65. Mollema, L., et al., *Participation in and attitude towards the national immunization program in the Netherlands: data from population-based questionnaires.* BMC Public Health, 2012. 12: p. 57.

66. van der Berg, J.D., J. Roorda, and M.J. Westerman, *[Reasons not to have your daughter vaccinated against the human papilloma virus in Twente: a questionnaire study].* Ned Tijdschr Geneeskd, 2010. 154: p. A1923.

67. Ladabaum, U. and J.M. Ford, *Lynch syndrome in patients with colorectal cancer: finding the needle in the haystack.* JAMA., 2012. 308(15): p. 1581-3.

68. Serpell, L. and J. Green, *Parental decision-making in childhood vaccination.* Vaccine, 2006. 24(19): p. 4041-6.

69. Brewer, N.T. and K.I. Fazekas, *Predictors of HPV vaccine acceptability: a theory-informed, systematic review.* Prev Med, 2007. 45(2-3): p. 107-14.

CHAPTER 6

THE PREDICTIVE VALUE OF DISCRETE CHOICE EXPERIMENTS IN PUBLIC HEALTH: AN EXPLORATORY APPLICATION

B.H. Salampessy, J. Veldwijk, A.J. Schuit,
K. van de Brekel-Dijkstra, R.E.J. Neslo,
G.A. de Wit, M.S. Lambooij

The Patient, in press.

ABSTRACT

Objective: To assess the predictive value of a Discrete Choice Experiment (DCE) in public health by comparing stated preferences and actual behavior.

Methods: 780 type 2 diabetes mellitus (T2DM) patients received a questionnaire containing a DCE with five attributes related to T2DM patients' willingness to participate in a combined lifestyle intervention. Panel mixed-multinomial-logit models were used to estimate the stated preferences based on 206 completed DCE questionnaires. Actual participation status was retrieved for 54 respondents based on patients' medical records and a second questionnaire. Predicted and actual behavior data were compared in at population level and at individual level.

Results: Based on the estimated utility function, 81.8% of all answers that individual respondents provided on the choice tasks were predicted correctly. Actual participation rate at the aggregated population level was minimally underestimated (70.1% versus 75.9%). Of all individual choices, 74.1% were predicted correctly with a positive predictive value of 0.80 and a negative predictive value of 0.44.

Conclusions: Stated preferences derived from a DCE can adequately predict actual behavior in a public health setting.

INTRODUCTION

Discrete Choice Experiments (DCEs) have been applied for many years in transportation economics, environmental economics and marketing, before being introduced in health economics [1]. Since then, outcomes from DCEs are increasingly being used to inform and support healthcare policy making [1]. However, critics question whether the outcomes of a DCE are a good proxy for actual behavior of patients and consumers. The hypothetical choices (stated preferences) made by respondents in DCEs only reflect what they presumably would have chosen given the specific set of program characteristics (i.e. attributes). These stated preferences may be different from choices made by the respondents in real life settings (revealed preferences) [2, 3].

Outside of health economics, stated preferences have been compared to actual behavior to some extent already [4-8]. These studies concluded that, in general, stated preferences could predict actual behavior in various study conditions, e.g., different elicitation methods, market segments, and time periods [4-8]. Within health economics, systematic reviews [1, 2, 9] identified only two studies that concerned the predictive value of DCEs [3, 10], and one study was recently published [11]. The first study conducted by Mark and Swait, who used both stated and actual behavior in their data analyses, concluded that SPs could be used to model market shares of newly introduced medication. However, no direct comparison was made between predicted and actual behavior at the individual level [2]. The one study that compared predicted behavior to actual behavior in a health care setting, was conducted by Ryan and Watson [3]. They asked women who visited a fertility clinic to indicate whether they were willing to participate in a Chlamydia screening test using multiple hypothetical scenarios. Subsequently, the women were asked to participate in an existing screening test offered at the clinic (this test was identical to one of the scenarios in the questionnaire). In the real life setting, 81% of the women behaved in accordance with what they had stated in the questionnaire. Most incorrect predictions concerned women who had stated to participate in the screening test but then declined this test when it was offered to them. The authors concluded their predictions overestimated actual behavior but emphasized that more research is needed to support their findings [3].

Several explanations have been suggested in literature for the discrepancy that was found in most studies that compared stated and actual behavior. Choices in DCEs may not have the same consequences (e.g., in terms of clinical effects, financial costs) for respondents as real life decisions, which is also referred to as hypothetical bias [12-14]. In addition, it is generally known that, in most cases, individual behavior is not solely based on whether the preferred program characteristics are present. Other circumstantial factors may also affect decisions of respondents in real life settings, e.g., suffering from illnesses, lack of time and lack of local facilities [15, 16]. Although the most important attributes and levels concerning a specific decision are preferably included in DCEs, these studies are limited in the number of attributes by nature, hence it may not always be possible to include all important attributes [5]. Another explanation is the intention-behavior gap. This is known to cause differences between planned and actual behavior, i.e. some respondents may change their initial intention prior to behavioral execution [17]. Since the outcomes of DCEs are based on stated behavior, the intention-behavior gap may cause incorrect predictions. All

these factors will influence individuals' behavior while not being accounted for in DCE studies. In the DCE literature, this is referred to as scale difference [5, 18]. Due to the fixed choice contexts and detailed information on a limited number of attributes, there is less 'noise' in stated preference data compared to revealed preference data (i.e. in SP data the error variance will be lower and scale will be higher). It can be questioned to what extent stated preferences can accurately predict actual behavior if such scale effects are not or cannot be accounted for.

Since outcomes of DCEs are increasingly used to support public health policy making [2], the extent to which stated preferences predict actual behavior is of societal interest. Therefore, this study aims to explore the predictive value of a DCE by assessing the consistency between stated preferences retrieved by a DCE and actual behavior in a specific health care context, i.e. type 2 diabetes mellitus (T2DM) patients and their participation in a combined lifestyle intervention (CLI).

MATERIALS AND METHODS

The study consisted of two stages: 1) stated preferences were derived from a DCE and 2) actual choices made by respondents in real life settings were determined. To investigate the predictive value of this DCE, we compared respondents' actual choices regarding participation to our predictions about their participation based on the stated preferences.

First stage: estimating stated preferences

All T2DM patients (except those who were terminally ill and those with a mental illness) in four general practices of Health Centers located in the area of Utrecht, The Netherlands were eligible to participate in this study. A questionnaire (questionnaire A) was sent to these 780 eligible patients by postal mail. Questionnaires were completed on a voluntary basis. A reminder was sent to patients who had not returned the questionnaire after three weeks. Questionnaire A contained questions concerning respondents' demographics and health status while ending with the DCE.

Attributes and levels of the DCE

The attributes and the levels that were used in the DCE were selected based on literature review, expert interviews and focus group interviews. These expert and focus group interviews were conducted to (1) determine the most important attributes, and (2) to ensure that the attribute levels were considered realistic and consistent with current practice. A detailed description of this process is described elsewhere [19]. The five attributes with the corresponding levels are shown in Table 1. The menu schedule and the physical activity schedule attributes, described the level of guidance provided by the lifestyle coach when establishing the goals of respondents concerning their diets and physical activity behavior. Respondents set these goals during consultations with their coach. Consultation structure, described whether these consultations took place individually, in small or large groups with other T2DM patients. The expected results in terms of weight loss and physical fitness that respondents had before starting the program, were reflected in the expected outcome attribute. Finally, the out-of-pocket (OOP) costs attribute reflected the amount that respondents had to pay when they participated in the CLI.

Table 1. The attributes and corresponding levels that were included in this DCE study *

Attributes	Level 1	Level 2	Level 3
Menu schedule			
A plan, which describes the aims of the participants with respect to improvements in their diet, developed by the participants of the program together with a coach			
	Flexible: primarily based on the participants' own initiatives and ideas	General: includes general information on a healthy diet and provides example recipes	Elaborate: a patient tailored schedule that is completely prepared by the lifestyle coach
Physical activity (PA) schedule			
A plan, which describes the aims of the participants with respect to improvements in their PA behavior, developed by the participants of the program together with a coach			
	Flexible: primarily based on the participants' own initiatives and ideas	General: includes general information on PA, and provides example exercises	Elaborate: a patient tailored schedule that is completely prepared by the lifestyle coach
Consultation structure			
The composition of the consults with the coach			
	Individually	Groups with 5 other T2DM patients	Groups with 10 other T2DM patients
Expected outcomes			
The results, in terms of weight loss and physical fitness expected by the respondent after completion of a lifestyle program			
	No weight loss but feeling more healthy	5 kilograms of weight lost and feeling more healthy	10 kilograms of weight lost and feeling more healthy
Out-of-pocket costs**			
Patients may have to pay (part) of the program costs out-of-pocket			
	75 euro per 3-6 months	150 euro per 3-6 months	225 euro per 3-6 months

* Attributes and corresponding levels are also described elsewhere [19]

** Levels of the linear attribute OOP-costs were coded as 0.75 (75 euro), 1.5 (150 euro) and 2.25 (225 euro).

Study design DCE

NGene 1.1 software (ChoiceMetrics, 2011) was used to create a D-efficient design for this study [18, 20]. The software was instructed to create a design using a panel mixed-multinomial-logit model (MIXL), with all beta-priors set at zero, 100 Halton draws and 500 repetitions. It was assumed that there would be no interaction between attributes, while level balance and minimal overlap between attribute levels were optimized. Utility balance between the alternatives within each choice task was optimized to be between 60%-40% and 80%-20%. The final design (D-error = 0.37) consisted of 18 unique choice tasks divided over two blocks. As shown in the example choice task of the DCE in figure 1, each choice task consisted of two unlabeled hypothetical CLI programs and an opt-out option. The latter was included to resemble real life settings more closely, since patients with T2DM could always decline the offer to participate.

Image your general practitioner or nurse practitioner would recommend you to participate in a combined lifestyle intervention program for 3-6months. In which situation would you prefer to participate, Situation 1 or situation 2? If you rather prefer not to participate in either of these situations, you can tick the opt-out option.

	Situation 1	**Situation 2**	**Opt-out option**
Menu schedule	General	Flexible	None
Physical activity schedule	General	Flexible	None
Consultation structure	Group with 5 others	Individual	None
Expected Outcomes	10 kilograms of weight loss and feeling better	10 kilograms of weight loss and feeling better	None
Out-of-pocket costs	€ 75	€ 75	€ 0
Tick the box of the scenario that you prefer:	☐	☐	☐

Figure 1. Example of a choice task.

Analysis of DCE data

Analyses of equation below were performed using the panel MIXL technique in Nlogit 5.0 (Econometric Software, Inc). This technique adjusts for the correlation between the answers within respondents, i.e. adjusting for the multilevel data structure, as well as preference heterogeneity between respondents [20, 21]. This model is fitted in an iterative manner until the log-likelihood function is optimized.

$U_{nj} = V_{nj} + \varepsilon_{nj} = \beta_{0i} + \beta_1$*General Menu schedule + β_2*Elaborate Menu schedule + β_3*General PA schedule + β_4*Elaborate PA schedule + β_5*Consultation structure in groups with 5 others + β_6*Consultation structure in groups with 10 others + β_7*Expected Outcome weight loss of 5 kilograms$_i$ + β_8*Expected Outcome weight loss of 10 kilograms$_i$ + β_9*OOP-costs$_i$ + ε_{nj}

The latent utility 'U' of individual 'n' concerning scenario 'j', can be estimated by taking the sum of the systematic utility element 'V' (i.e. the utility of individual 'n' concerning scenario 'j' calculated based on all attribute levels and covariates) and the random error term 'ε' (i.e. all unobserved and unobservable factors that influence the utility of person 'n' concerning scenario 'j'). This error term follows an extreme value type 1 distribution. β_0 represented the constant of the model. The constant describes the utility of T2DM patients for a lifestyle program versus no lifestyle program (opt-out) when all attributes are set at zero. β_1-β_9 represented the attribute level estimates. Four attributes (menu and PA schedule, consultation structure and expected outcome) were coded using effects coding. The reference category in effects coding was coded as -1, which summed the attribute in each category to zero. The estimates for the reference categories were calculated

using (-1) * ($\beta_{effectcode1}$+$\beta_{effectcode2}$) [5, 22]. Based on model fit tests (AIC, BIC, Log likelihood) it was tested which model fitted best to the data. Based on the significance level of the standard deviation of the attributes it was tested what attributes should be included as random parameters due to significant preference heterogeneity. In addition, different distributions of the random parameters were tested and based on the model fit results, all random parameters were included with a normal distribution. The constant variable, expected outcome and OOP-costs were included as random parameters (indicated by $_i$ in the utility equation) with a normal distribution. Since the panel MIXL model does not account for variability in individual errors (scale heterogeneity) [23], the modeling procedures described above were repeated using a Heteroscedastic Extreme Value (HEV) model (accounts for scale heterogeneity), and a Generalized-Mixed-Logit (G-MIXL) model (accounts for scale and preference heterogeneity) [23].

Second stage: determining actual behavior

A CLI that was implemented at the participating health centers was used to determine actual behavior of respondents. This CLI was offered to patients who had cardiovascular disease, diabetes mellitus type 2, COPD, anxiety or depressive disorders, and to patients who smoked or had Body Mass Index (BMI) scores over 30 kg/m^2. While this program was available for patients with a wider range of health conditions, in this study only patients with T2DM were selected. Within the program, goals were set to assist patients to improve their lifestyle with the help of lifestyle coaches, physiotherapists, dieticians and specialized nurses. Respondents who had completed questionnaire A (n=206) were eligible for the second stage of the study. After administration of questionnaire A, General Practitioner (GP) or Nurse Practitioners (NPs) discussed participation in the CLI with these respondents.

Actual behavior

In February 2013 actual behavior was determined for all eligible respondents of the second stage using patients' medical records and an additional questionnaire (questionnaire B). CLI participation of respondents was defined as having an intake appointment with the lifestyle coach. Respondents were marked as non-participant when they had been offered to participate in the program but declined this offer. The one respondent that did make an intake appointment but did not show up was marked as a non-participant. Fourteen respondents (n=14) were excluded from this stage of the study due to deregistration from the participating health centers, death or terminal illness.

Statistical analyses

To test the difference in demographic characteristics between the respondents that only completed the first stage of this study and the respondents that completed both the first and second stage of this study, independent sample t-tests were used. Results were considered statistically significant if $p<0.05$.

Within respondent consistency

Two within respondent consistency tests were performed to assess to what extent the stated preferences could reproduce the actual choices made by respondents in the DCE. In both tests, predicted choices were determined using the stated preferences and were then compared to the actual choices of the DCE. In the first test, these stated preferences were based on data from all respondents (n=206). Since only data derived from the questionnaire was used, the test assessed the predictive value of the fitted model itself. To determine which scenario in each choice task respondents would prefer, the individual utility scores that resulted from the MIXL were used. Subsequently, in agreement with the Random Utility Theory [12], utility maximization was assumed in respondents' decision making process. Therefore, it was expected that the respondent would choose the scenario with the highest utility score within the choice task. The procedure described above was then repeated in a slightly different manner. In the second test, it was tested whether the stated preferences that were measured among a random subgroup (50%, n=103) of respondents could be used to correctly predict the actual choices of the remainder of the sample for each of the choice tasks.

Predictive value at population level

The predictive value of DCEs at aggregated level was determined by comparing the estimated participation rate based on the stated preferences to the actual participation rate found in the second stage of the study. The participation rate was estimated based on the CLI as actually implemented at the health centers to allow comparison between stated and actual behavior. This CLI consisted of an elaborate menu schedule, a general PA schedule, an individual consultation structure and was offered free of charge. According to the guidelines of Dutch General Practitioners for managing patients with obesity or T2DM [24], a five percent weight loss of obese patients provides considerable health gains and is assumed to be realistically achievable without surgical treatment. Due to the current weight status and BMI of the respondents in our sample, a 5% weight loss equals a weight loss of five kilograms or slightly more. Therefore the expected outcomes attribute was set at five kilograms. To estimate the participation rate based on the stated preferences, utility scores were calculated. Since random parameters were included, the probability to participate could not be calculated directly. Therefore, the mean probability of 10.000 simulations was estimated by taking the average of all simulated probabilities given every tested CLI scenario, which was calculated as $1/(1+\exp^{-v})$ [5, 12]. Similar to the within respondent consistency test, the scenario with the highest utility score was expected to be the choice of the respondent. Finally, outcomes of these analyses were compared to the actual participation rate concerning the CLI that was observed in this study by means of a Chi-square test.

Predictive value at individual level

Finally, the predicted choices of respondents were compared to their actual choices in real life settings. Individual utility scores were calculated for the CLI that was implemented at the health centers and subsequently compared to the opt-out option. Fixed coefficients of the MIXL were used for three attributes (menu and PA schedule and consultation structure) while individual

coefficients were used for one attribute (expected outcome). The fifth attribute (OOP-costs) remained at zero. Again, the scenario with the highest utility score was expected to be the choice of respondents. Finally, these predicted choices were compared to the choice of respondents concerning participating in the CLI in the real life setting. Results were presented in percentages correctly predicted choices (correspondence level), the proportion of correctly identified participants (positive predictive value, PPV) and the proportion of correctly identified non-participants (negative predictive value, NPV). In addition, results were also described as in terms of sensitivity and specificity, and the Cohen's Kappa coefficient was calculated.

RESULTS

First stage: estimating stated preferences

Study population

The response rate of questionnaire A was 26.4% (n=206). As shown in Table 2, respondents had a mean age of 63 years (SD11.4). The majority had attained a medium educational level (43.3%) and was of West-European origin (64.4%). On average, respondents had been diagnosed with T2DM almost eight years prior to completion of questionnaire A (7.8 years, SD6.2). In addition, respondents had a mean HbA_{1c} of 52.6 mmol/mol (6.96%) (SD10.1) and a BMI-score of 29.7 kg/m^2 (SD5.4).

Table 2. Total study population (n=206)

		Mean (Standard deviation)	Percentage
Age		63.4 (11.47)	
Gender	Men		53.7
Education level	Low		37.0
	Medium		43.3
	High		19.7
Ethnicity	West-European		64.4
	Moroccan, Turkish		8.8
	Other or not specified		26.8
Status of disease			
Duration of diabetes (years)		7.8 (6.22)	
HbA_{1c} (mmol/mol)		52.6* (10.07)	
Body Mass Index (kg/m^2)**		29.7 (5.43)	
Dutch EQ-5D**		0.88 (0.18)	

* This equals an HbA1c of 6.96%; ** Significant differences ($p<0.05$) found between the respondents of only the first stage (n=152) and respondents of the first and second stage (n=54) of the study. On average, respondents that participate both in the first and second stage had higher BMI scores ((33.8 kg/m^2 (SD 6.4)) and lower EQ-5D scores (0.81 (SD 0.2).

Stated preferences

Of the 1818 (206*9) possible choice tasks, 1504 were completed by respondents. As shown in Table 3, three attributes showed significant attribute levels estimates, (i.e. PA schedule, expected outcome and OOP-costs) implying that the two other attributes did not significantly affect the choice for participation in a CLI. The negative coefficient of OOP-costs indicates that with a decrease in OOP-costs, the willingness to participate in a CLI increases. Respondents preferred an elaborate menu and a general PA schedule over all other menu and PA schedules. Individual consultations were preferred over consultations in groups, and expected weight loss of five kilograms was preferred over no weight loss or a 10 kilograms weight loss. Analyses with HEV and G-MIXL models showed no scale heterogeneity since the scale parameters were insignificant in both models (results not shown).

Table 3. Estimates of the attribute levels based on the panel mixed-logit model

Attribute			**Estimate**	**Standard error**
Constant		Mean	-1.09**	0.37
		SD	3.32	0.30
Menu schedule	Flexible (reference)		-0.21	0.12
	General		0.06	0.14
	Elaborate		0.15	0.16
PA schedule	Flexible (reference)		-0.16	0.10
	General		0.19*	0.08
	Elaborate		-0.03	0.10
Consultation structure	Individual (reference)		0.12	0.13
	Groups of 5		0.09	0.12
	Groups of 10		-0.21	0.18
Expected outcome	No weight loss (reference)	Mean	-0.80**	0.15
		SD	1.33	0.18
	Weight loss 5 kilograms	Mean	0.51**	0.12
		SD	0.20	0.19
	Weight loss 10 kilograms	Mean	0.29	0.19
		SD	1.31**	0.19
Out-of-pockets costs		Mean	-1.26**	0.23
		SD	-1.10**	0.15

SD= standard deviation, *significant at p<.05; ** significant at p<.01

The fitted model consisted of four effects coded attributes and a continuous coded attribute (i.e. OOP-costs). In addition, three parameters (i.e. the constant, expected outcome and the OOP-costs) were set random. The SD reflects the variance between the individual coefficients and the average coefficient.

Second stage: determining actual behavior

Actual choices were retrieved for 54 respondents based on patients' medical records (n=41) and questionnaire B (n=13). The latter showed a response rate of 43.1%. All 54 respondents were included in the second stage of study in which 41 respondents (75,9%) reported they chose to participate in the CLI while 13 respondents declined the offer to participate. All demographics of the 54 respondents of the second stage are equal to those of the respondents that only participated in the first stage, except they had higher BMI scores (33.8 kg/m^2, SD6.4) and lower self-reported health status (0.81 score SD0.2) compared to respondents of the first stage (n=152) (Table 2, footnote).

Within respondents consistency

Stated preferences in the first test reproduced 81.8% of the actual answers made by respondents in the DCE. Using the stated preferences of the randomly selected sample (50% of the total sample) resulted in accurate predictions in 45.0% of the choice tasks completed by the other half of the population.

Predictive value at population level

When participation rates of the offered CLI were compared at aggregated level, the estimated participation rate based on the stated preferences was somewhat but not significantly lower (70.1%) compared to the actual participation rate (75.9%) (Chi-square = 2.45, P >0.05).

Predictive value at individual level

As shown in Table 4, when stated preferences and actual behavior were compared at individual level, a correspondence level of 74.1% was found. In addition, the PPV of 0.80 implies that of those respondents who were predicted to participate in the offered CLI, four out of five actually participated in the CLI. Similarly, the NPV of 0.44 implies that 44% concerning non-participation were correct predictions when compared to actual behavior. Most of the incorrect predictions regarded respondents who were predicted to participate but declined the offer in a real life setting. Moreover, the sensitivity was 0.90 and the specificity was 0.35. Despite the number of correct predictions, the Cohen's Kappa resulted to be insignificant. Since the majority of T2DM patients decided to participate in the CLI (76%) and only a minority decided not to participate (24%), the distribution of patients' behavior was highly skewed, which probably caused the kappa coefficient to be insignificant [25, 26].

DISCUSSION

Comparisons between stated preferences and actual behavior at aggregated population level showed a slight but not statistically significant underestimation for the stated preferences (70.1% versus 75.9%). In 74.1% of the cases, the stated preferences corresponded with actual behavior at individual level, which resulted in a positive predicted value (PPV) of 0.80 and negative predictive

value (NPV) of 0.44, a sensitivity of 0.90 and a specificity of 0.35.

Table 4. Cross table of comparing stated with actual behavior at individual level

		Actual choices		Total
		Participation	No participation	
Predicted behavior	Participation	36	9	45
	No participation	5	4	9
	Total	41	13	54

Correspondence level, correctly predicted choices =(36+4)/54=74.1%

PPV, share of correctly predicted participations =36/(36+9)= 0.80

NPV, share of correctly predicted non-participations =4/(4+5)=0.44

Sensitivity =36/(36+4)= 0.90; Specificity = 5/(5+9)=0.35; Cohen's kappa =0.21 (approx. T= 1.57, ns)

Actual participation rate = 41/54=75,9%

Although actual behavior could partly be predicted based on the stated preferences elicited by the DCE, a discrepancy was found, namely 25.9% of the predicted choices differed from actual behavior. Results indicated an overestimation of the stated preference utilities. More respondents for whom participation was predicted actually opted-out compared to respondents who were predicted to opt-out to but actually participated.

Three distinct reasons may be underlying this finding. First, respondents might have incorporated other attributes in their decision concerning participation in a CLI in a real life setting. Since the actual choice leading to behavior was not limited to the attributes of the DCE but also included all unobserved attributes, predictions that are based solely on the DCE will inevitably lead to some prediction error. Therefore, it is stressed that the attribute (level) selection procedure is deliberate and concise [5, 12, 27]. Although this process was followed closely within this study [19], there is always a possibility that some important attribute was missed and therefore caused a discrepancy between the calculated participation rates based on the stated preferences and the actual behavior. Future research may focus on possible design or statistical solutions to reduce the error in calculated utilities due to missed attributes. It might, for instance be explored whether it would be feasible to have individuals decide to add certain attributes from a predefined list to the obligatory attributes within a DCE using an online setting. Second, respondents' decisions might be affected by different choice contexts. While in the DCE the respondents all evaluated the choice tasks within the same choice context, in real life respondents may differ with respect to for instance demographics, psychological determinants (e.g., attitude), and the priorities or skills of their GP and NP to motivate them to participate in the CLI. The presence of a context effect is underlined by the fact that participation rates of CLIs as reported in literature (23-79%) [28, 29], indicate the overrepresentation of certain groups. This form of selection bias, as was probably present in this study, is most likely always present in practice. For instance, GPs and NPs are often involved in the process of enrolling participants for CLIs, the degree in which they will motivate and persuade patients to participate might be subjective to their judgment about the extent to which that particular patient might benefit from participation. External factors as

described above may influence the relative importance of the attributes within the DCE (e.g., low income might be related to how the OOP-costs attribute is valued) and thus may influence the calculated utility for a certain scenario. Third, the intention-behavior gap probably always accounts for some error in the predictions based on DCEs. Perceived barriers and facilitators are likely to come into play when individuals actually decided whether to participate in a CLI. These barriers and facilitators might increase or decrease the utility of the opt-out option as compared to the utility of participating in the CLI without changing the relative importance of the attribute levels. For this reason, a complete correspondence level between stated preference and actual behavior may never be possible. In summary, external factors that are not included in the DCE but in real life, affect the utility of a particular scenario, cause an unknown discrepancy between the utilities of the stated and revealed preferences. Such differences are known as scale differences. Several initiatives might be undertaken to minimize the influence of scale on stated preferences. First, an online questionnaire that adapts the choice context of the decision to patient specific characteristics may be used as a tool to mimic the real life decision setting as closely as possible, and therefore may reduce the gap between the hypothetical and the real life choice situation of respondents. Second, analytical models that include context related covariates (e.g., respondent characteristics or context characteristics) might be used. Hybrid models or models that incorporate interaction terms between attribute levels and context factors may theoretically provide more accurate predictions, since these models incorporate the influence of relevant external factors. However, no study is likely to have sufficient power to incorporate all external factors. Sample size of the current study was sufficient for estimating main effects but not for incorporating several interaction terms. Future research that examines the external validity of a DCE should bear in mind to conduct hybrid models or to account for possible interactions when running sample size calculations beforehand. Results of comparisons at individual level of this study are in line with results of one other healthcare application [3] and the evidence base from other fields of research [4-8]. While the PPV of 0.80 seems promising, the NPV was not better than could have been expected by chance (0.44). However, values of PPV and NPV are affected by the number of respondents that participates or declines to participate [30], overrepresentation of participating respondents will obviously result in a higher PPV (i.e. the true positives will then always be higher than the false positives). In this study, a relatively large number of respondents decided to participate while only a small number of respondents declined participation. Subsequently, the reported PPV and NPV may be less accurate than anticipated beforehand due to under sampling of non-participating respondents.

Additionally, most DCEs are used to predict ***engaging*** behavior (i.e. the choice for participation) of respondents, e.g. uptake in new preventive programs, while it is less common to predict ***refraining*** behavior (i.e. the choice against participation). This implies that DCEs are likely to be more valuable in understanding why people engage in the behavior under study than understanding why people refrain from that behavior. Refraining behavior may be motivated by a combination of other (non-observed) attribute levels and external factors compared to engaging behavior. The PPV and sensitivity of the current study might be considered as good, since actual behavior was correctly predicted in more than three out of four respondents.

A key issue in the application of DCEs in health policy remains how policy makers should use the outcomes of DCEs. When predicting engaging behavior of respondents, stated preferences derived from DCEs can be used to predict actual behavior of respondents. However, when using the outcomes to predict refraining behavior, different research objectives should be formulated, probably different attribute levels should be identified, different external factors should be measured and different designs should be conducted.

In conclusion, stated preferences can adequately predict actual behavior in a public health setting. However, it remains unclear to what extent missed attributes, choice context and the intention-behavior gap play a part in the discrepancy between stated preferences and actual behavior and how these issues can be overcome. Moreover, it is unsure to what extent DCEs can predict refraining behavior, which is of particular importance when DCE results are translated into policy implications. Future research should assess the predictive value of DCEs in health economics using different approaches (both modeling engaging and refraining behavior) among different patient groups and different decision contexts. Because refraining behavior is not simply the opposite of engaging behavior, research on this specific topic is called for.

References

1. Clark, M.D., et al., *Discrete choice experiments in health economics: a review of the literature.* Pharmacoeconomics, 2014. 32(9): p. 883-902.
2. de Bekker-Grob, E.W., M. Ryan, and K. Gerard, *Discrete choice experiments in health economics: a review of the literature.* Health Econ, 2012. 21(2): p. 145-72.
3. Ryan, M. and V. Watson, *Comparing welfare estimates from payment card contingent valuation and discrete choice experiments.* Health Econ, 2009. 18(4): p. 389-401.
4. Adamowicz, W.L., *Combining revealed and stated preference methods for valuing environment amenities.* Staff paper. 1992, Edmonton, Canada: Dept. of Rural Economy, Faculty of Agriculture and Forestry, University of Alberta.
5. Louviere, J.J., D.A. Hensher, and J.D. Swait, *Stated Choice Methods: Analysis and Applications.* 2000: Cambridge University Press.
6. Hensher, D. and M. Bradley, *Using stated response choice data to enrich revealed preference discrete choice models.* Mark Lett, 1993. 4(2): p. 139-151.
7. Ben-Akiva, M. and T. Morikawa, *Estimation of switching models from revealed preferences and stated intentions.* Transport Res A Gen, 1990. 24(6): p. 485-495.
8. Adamowicz, W., et al., *Perceptions versus objective measures of environmental quality in combined revealed and stated preference models of environmental valuation.* J Environ Econ Manage, 1997. 32(1): p. 65-84.
9. Ryan, M. and K. Gerard, *Using discrete choice experiments to value health care programmes: current practice and future research reflections.* Appl Health Econ Health Policy, 2003. 2(1): p. 55-64.
10. Mark, T.L. and J. Swait, *Using stated preference modeling to forecast the effect of medication attributes on prescriptions of alcoholism medications.* Value Health, 2003. 6(4): p. 474-82.
11. Krucien, N., A. Gafni, and N. Pelletier-Fleury, *Empirical Testing of the External Validity of a Discrete Choice Experiment to Determine Preferred Treatment Option: The Case of Sleep Apnea.* Health Econ, 2014. 1(10).
12. Ryan, M., K. Gerard, and M. Amaya-Amaya, *Discrete Choice Experiments in a Nutshell, in Using Discrete Choice Experiments to Value Health and Health Care,* M. Ryan, K. Gerard, and M. Amaya-Amaya, Editors. 2008, Springer Netherlands. p. 13-46.
13. Johnson, F.R., et al., *How does cost matter in health-care discrete-choice experiments?* Health Econ, 2011. 20(3): p. 323-30.
14. Train, K. and W.W. Wilson, *Estimation on stated-preference experiments constructed from revealed-preference choices.* Transport Res B Meth, 2008. 42(3): p. 191-203.
15. Rhodes, R.E., R.C. Plotnikoff, and K.S. Courneya, *Predicting the physical activity intention-behavior profiles of adopters and maintainers using three social cognition models.* Ann Behav Med, 2008. 36(3): p. 244-52.
16. Thomas, N., E. Alder, and G.P. Leese, *Barriers to physical activity in patients with diabetes.* Postgrad Med J, 2004. 80(943): p. 287-91.
17. Sheeran, P., *Intention—Behavior Relations: A Conceptual and Empirical Review.* Eu Rev Social Psych, 2002. 12(1): p. 1-36.
18. Hensher, D.A., J.M. Rose, and W.H. Greene, *Applied Choice Analysis: A Primer.* 2005, New York: Cambridge University Press.
19. Veldwijk, J., et al., *Type 2 diabetes patients' preferences and willingness to pay for lifestyle programs: a discrete choice experiment.* BMC Public Health, 2013. 13: p. 1099.
20. Bliemer, M.C. and J.M. Rose, *Construction of experimental designs for mixed logit models allowing for correlation across choice observations.* Transport Res B Meth, 2010. 44(6): p. 720-734.
21. McFadden, D. and K. Train, *Mixed MNL models for discrete response.* Journal of applied Econometrics, 2000. 15(5): p. 447-470.
22. Bech, M. and D. Gyrd-Hansen, *Effects coding in discrete choice experiments.* Health Econ, 2005. 14(10): p. 1079-83.
23. Fiebig, D.G., et al., *The Generalized Multinomial Logit Model: Accounting for Scale and Coefficient Heterogeneity.* Mark Sci, 2010. 29(3): p. 393-421.
24. Van Binsbergen, J., et al., *NHG-Standaard Obesitas.* Huisarts en Wetenschap, 2010. 53(11): p. 609-625.

25. Cicchetti, D.V. and A.R. Feinstein, *High agreement but low kappa: II. Resolving the paradoxes.* J Clin Epidemiol, 1990. 43(6): p. 551-8.
26. Feinstein, A.R. and D.V. Cicchetti, *High agreement but low kappa: I. The problems of two paradoxes.* J Clin Epidemiol, 1990. 43(6): p. 543-9.
27. Lancsar, E. and J. Louviere, *Conducting discrete choice experiments to inform healthcare decision making: a user's guide.* Pharmacoeconomics, 2008. 26(8): p. 661-77.
28. James, D.V., et al., *Factors associated with physical activity referral uptake and participation.* J Sports Sci, 2008. 26(2): p. 217-24.
29. Diabetes Prevention Program Research, G., et al., *10-year follow-up of diabetes incidence and weight loss in the Diabetes Prevention Program Outcomes Study.* Lancet, 2009. 374(9702): p. 1677-86.
30. Altman, D.G. and J.M. Bland, *Diagnostic tests 2: predictive values.* BMJ, 1994. 309(6947): p. 102.

PART 3

THE INFLUENCE OF THE PRESENTATION OF CHOICE TASKS ON RESPONDENTS' CHOICE BEHAVIOR AND ON THE OUTCOMES OF A DISCRETE CHOICE EXPERIMENT

CHAPTER 7

THE EFFECT OF INCLUDING AN OPT-OUT OPTION IN DISCRETE CHOICE EXPERIMENTS

J. Veldwijk, M.S. Lambooij, E.W. de Bekker-Grob,
H.A. Smit, G.A. de Wit

PlosOne (2014), 9, e111805.

ABSTRACT

Objective: To determine to what extent the inclusion of an opt-out option in a DCE may have an effect on choice behaviour and therefore might influence the attribute level estimates, the relative importance of the attributes and calculated trade-offs.

Methods: 781 Dutch Type 2 Diabetes Mellitus patients completed a questionnaire containing nine choice tasks with an opt-out option and nine forced choice tasks. Mixed-logit models were used to estimate the relative importance of the five lifestyle program related attributes that were included. Willingness-to-pay (WTP) values were calculated and it was tested whether results differed between respondents who answered the choice tasks with an opt-out option in the first or second part of the questionnaire.

Results: 21.4% of the respondents always opted out. Respondents who were given the opt-out option in the first part of the questionnaire as well as lower educated respondents significantly more often opted out. For both the forced and unforced choice model, different attributes showed significant estimates, the relative importance of the attributes was equal. However, due to differences in relative importance weights, the WTP values for the PA schedule differed significantly between both datasets.

Conclusions: Results show differences in opting out based on the location of the opt-out option and respondents' educational level; this resulted in small differences between the forced and unforced choice model. Since respondents seem to learn from answering forced choice tasks, a dual response design might result in higher data quality compared to offering a direct opt-out option. Future research should empirically explore how choice tasks should be presented to make them as easy and less complex as possible in order to reduce the proportion of respondents that opts-out due to choice task complexity. Moreover, future research should debrief respondents to examine the reasons for choosing the opt-out alternative.

INTRODUCTION

There seems to be consensus regarding the inclusion of an opt-out option in Discrete Choice Experiments (DCEs) that aim to determine the potential participation in an elective program as such an option is in accordance with respondent's choice options in real life [1-4]. Moreover, when estimating the potential number of participants in any program, insight into the percentage of the target population that does not wish to participate in such a program is necessary. However, if individual preferences are measured to determine which components define the most preferred program or treatment, the inclusion of an opt-out option might not be a necessity but rather a threat to efficiency. Until now, the choice to include an opt-out option is determined by the objective of the DCE in the first place. Nevertheless, very little empirical evidence exists on the issue whether, and to what extent the inclusion of opt-out options in DCEs affect choice behavior of respondents. Which may therefore influence the precision of the estimates of attributes, the relative importance of the attributes, trade-offs (e.g., willingness to pay) calculated based on these estimates and thereby the conclusions that will be drawn from a DCE.

Most DCEs in health economics are rooted in the Random Utility Theory (RUT) [3, 5-7]. This theory assumes that respondents choose rationally and will select the scenario that generates the highest personal utility, that is, respondents will only select the opt-out option if none of the presented scenarios in that specific choice task is more attractive than the opt-out option [5, 8]. Additional research shows that from this perspective, forcing respondents to make a choice induces bias, as they would not always make that same choice in real life [3, 9, 10]. In such a forced-choice situation, people who would rather choose to opt-out, tend to randomly select either scenario from a choice task or select the most safe/least extreme scenario [9-12]. As a consequence, the standard error of the attribute estimates will increase while the external validity decreases [9, 10]. In summary, based on the RUT, an opt-out option can always be included, if this is accordance with the respondent's real-life decision context.

In practice, other motives than achieving the highest personal utility may be more important when people make their decisions [8-22]. This resulted in the hypothesis that only very few respondents act solely according to the assumptions of the RUT when choosing the opt-out option. Some individuals are more prone to choose the opt-out scenario even before they actually evaluate the different scenarios in a choice task. Baron and Ritov (1992) argued that individuals choose the opt-out scenario to protect themselves from poor choices, as negative outcomes based on taking action (choosing) are perceived as worse compared to negative outcomes due to inactivity (not choosing) [19]. This finding was confirmed by many others [13, 17, 18], among which a theory by Luce and colleagues who suggest that if people decide to make a choice, the tendency to choose to opt-out increases as the trade-off becomes more difficult and the decision at hand is emotion-laden [12, 16]. This indicates that people choose to opt-out to avoid making difficult trade-offs [12, 16]. Research by Dahr and colleagues (1997 and 2003) showed that choice task complexity (i.e. large number of choice scenarios per choice task or comparable choice scenarios with respect to their attractiveness) results in more opting out [9, 11]. In summary, it seems plausible

that respondents choose the opt-out option more often if they have to decide about a complex emotion-laden topic, if choice tasks are difficult, if scenarios are complex and if none of the scenarios is clearly superior. This way, respondents minimize their effort and reduce internal conflict induced by (negative) decision-making.

The above is of special interest within the public health setting. Decisions about personal health, public health and prevention are by definition complex and difficult and not part of an individual's everyday decisions [23-25]. Because DCE data are most often analysed according to the assumptions of the RUT, it can be discussed to what extent DCE results will be biased when respondents choose to opt-out as a result of reasons described above and not based on perceived personal utility. Until now, there is no empirical evidence on the effects of including an opt-out option on choice behaviour and the results of a health-related DCE. Therefore, the aim of this study is to investigate to what extent including an opt-out option in a DCE influences choice behaviour and thereby affects the attribute level estimates, relative importance of the attributes, calculated trade-offs, and the conclusions drawn for this DCE.

MATERIALS AND METHODS

Participants and recruitment

This study included participants diagnosed with type 2 diabetes mellitus (DM2), who were 35-65 years of age and who were not suffering from any serious complications due to their DM2. Participants were contacted via their diabetes care group. Within the Netherlands, diabetes care of all diagnosed DM2 patients is arranged in care groups, which are legal entities formed by multiple health care providers centered around general practitioners [26]. All Dutch care groups (n=94) were categorized by the province of the Netherlands in which they are located. Per province, one care group was randomly selected and contacted, until five care groups agreed to participate. These five care groups distributed 2,500 questionnaires in total, among all the eligible DM2 patients who were registered at these care groups. The Dutch National Ethics Board (Central Committee on Research involving Human Subjects) concluded that formal testing by a medical ethical committee was not necessary, as DM2 patients were only required to complete an anonymous questionnaire once, which is in accordance with the guidelines laid down in the Declaration of Helsinki.

Derived attributes and levels

Based on previously published literature on barriers and facilitators to participate in a lifestyle program among DM2 patients [27-37], interviews with experts (n=3) and four focus group interviews with DM2 patients (total n=24), five attributes with each three levels were selected for the current DCE. These included: menu schedule, physical activity (PA) schedule, consultation structure, expected outcome, and out-of-pocket costs.

A menu schedule and a PA schedule are plans that will be developed by the participants in the program together with a lifestyle coach. These plans describe the aims of the participants with

respect to improvements in their diet and PA behavior. A flexible schedule is a schedule that is based mostly on the participants own initiatives and ideas. A general schedule is a schedule that includes general information on either a healthy diet or PA and provides example recipes or exercises. An elaborate schedule comprises a tailored schedule that is prepared mostly by the lifestyle coach. Consultation structure describes the composition of the consultations with the coach (i.e. individual or in groups of 5 or 10 other patients). These are the consultations during which the participants develop their menu plan and PA schedule, and during which they discuss their progress. The expected outcome is meant to describe the results with respect to weight loss (0, 5 or 10 kilograms) and physical fitness, which the respondent can expect to achieve after completing the lifestyle program. Finally, as the costs of participating in a lifestyle program are not part of the participant's health insurance in general, the participant will have to pay for (part) of the program out-of-pocket. These costs can amount to either €75, €150 or €225 per year.

Experimental design

The scenarios in the DCE are constructed by combining different levels of each included attribute. The experiment comprises an unlabeled (generic) design with respect to the lifestyle program options. NGene 1.1 software (ChoiceMetrics, 2011) was used to create a D-efficient design for the current study. The software was instructed to create a design with two blocks using a panel-mixed-multinomial model with all beta-priors set at zero. It was assumed that there would be no interaction between attributes, while level balance, utility balance and minimal overlap between attribute levels were optimized [38, 39].

Finally, the design consisted of a sample of nine choice tasks per block (18 unique choice tasks in total). Within this design, each choice task contained two lifestyle program scenarios. To compare the possible differences in decision making when respondents are forced to make a choice or are offered an opt-out option (unforced choice option) and to obtain insight into the possible influence of the location of the opt-out option within the questionnaire [13], a within-sample design using four versions of the questionnaire was developed. This implies that version 1 and 2 of the questionnaire included the nine choice tasks of block 1 and version 3 and 4 included the nine choice tasks from block 2. Version 1 and 3 first offered the opt-out option and then forced respondents to choose, whereas version 2 and 4 started with forced choice tasks followed by choice tasks with an opt-out option (Table 1). To adjust for bias induced by the question order, the order of the choice tasks per version of the questionnaire was randomized. The opt-out option was included in the choice tasks as a third scenario, to prevent respondents from interpreting the opt-out option in different ways, all attributes were explicitly set to zero or 'none' in this scenario. Eventually, every respondent was asked to answer 18 choice tasks of which nine with the opt-out alternative and nine without.

Pilot test

The questionnaire was pilot tested among a subgroup (n=20) of the study population to ensure that the wording used in the questionnaire was correct and understood by the target population [40, 41]. During the pilot phase there was specific attention for the issue of interpretation of

the opt-out option. The pilot tests were distributed by means of postal questionnaires, respondents were asked to mark every question or answering category that they did not understand or found hard to grasp and they were asked to provide suggestions for improvement. Moreover, three think-aloud pilot tests were conducted to obtain more insight into the respondent's approach when answering the choice tasks. No changes in the attributes and/or levels were deemed necessary based on the results of this pilot study. Power/sample size calculations were performed based (partly) on the retrieved pilot data, to check a posteriori how large the sample size should be to find significant differences for each attribute at a 5% level in the final DCE [38, 42].

Table 1. Overview of the content of every version of the questionnaire

Questionnaire	Block	First nine choice tasks	Second nine choice tasks
Version 1	1	Including opt-out	Excluding opt-out
Version 2	1	Excluding opt-out	Including opt-out
Version 3	2	Including opt-out	Excluding opt-out
Version 4	2	Excluding opt-out	Including opt-out

Questionnaire

The questionnaire contained two parts. The first section of the questionnaire consisted of 28 questions about the respondent's demographic/background characteristics and on the patient's opinion with respect to lifestyle programs in general, accompanied by the EuroQol-5D health status questionnaire [43]. The second part was the actual DCE, which started with a detailed description of the attributes and levels and gave comprehensive guidance on how to answer the choice tasks provided. Every choice task started with the question: 'Imagine that your general practitioner advises you to participate in a lifestyle program for one year, which program would you prefer: the program in scenario 1 or scenario 2?' The following sentence was added to the above question in the choice tasks that included an opt-out option: 'If you prefer not to participate in either of the scenarios, you can tick the box 'none''.

Statistical analyses

According to the RUT, perceived utility (U) of a person 'n' in choice scenario 'j' is estimated by the sum of the systematic utility component (V) (i.e. the mean utility of the target population concerning a specific topic including the same attributes and levels) and the random error term (ε) (i.e. the deviation of the utility of every single person 'n' compared to the mean utility) (equation 1.1) [5, 6, 44].

[Eq. 1.1]

$$U_{nj} = V_{nj} + \varepsilon_{nj}$$

Based on the data retrieved by a DCE, the systematic utility component (V) of equation 1.1 can be estimated. This was estimated separately for the forced-choice data and the data that included an opt-out option. All analyses were conducted using mixed- logit (MIXL) models, to take preference heterogeneity into account. For the forced-choice data, the attribute estimates were estimated using equation 1.2 & 1.3.

[Eq. 1.2]

$V_{\text{alternative A}} = \beta_0 + \beta_1$ * flexible menu schedule + β_2 * elaborate menu schedule + β_3 * flexible PA schedule + β_4 * elaborate PA schedule + β_5 * consults in group of 5 + β_6 * consults in groups of 10 + β_7 * expected outcome + β_8 * out-of-pocket costs

[Eq. 1.3]

$V_{\text{alternative B}} = \beta_1$ * flexible menu schedule + β_2 * elaborate menu schedule + β_3 * flexible PA schedule + β_4 * elaborate PA schedule + β_5 * consults in group of 5 + β_6 * consults in groups of 10 + β_7 * expected outcome + β_8 * out-of-pocket costs

β_0 represents the alternative specific constant, while $\beta_1 - \beta_8$ are the attribute estimates that indicate their relative importance. All included attributes were tested for linearity, the attributes that appeared not to be linear were effects coded (i.e. menu schedule, PA schedule and consult structure). In contrast to dummy coding, effects coding enables one to compare the estimates of all attributes despite their categorization into non-linear levels, because the effects are uncorrelated with the intercept [6, 45]. This coding procedure codes the reference category -1, therefore the sum of the effect coded attributes is always 0. The coefficient for the reference category is therefore $-1*(_{\beta\text{effect code 1}} + _{\beta\text{effect code 2}})$.

After comparing the model fits (based on AIC, BIC and Chi-square) of different models including different (sets of) random parameters, two parameters were set at random for the final analysis: the constant and the attribute expressing the expected outcome of the lifestyle program. In addition, different distributions of the random parameters were tested and based on the model fit results, both random parameters were included with a normal distribution. It was expected that especially the constant and the outcome attribute would show high preference heterogeneity among the respondents, due to large differences in general opinions concerning lifestyle programs and the variation in Body Mass Index (BMI) among the respondents.

The attribute estimates for the data that included the opt-out option was retrieved via equations 1.4 - 1.6. The β-values in this equation are to be interpreted as explained above for the forced-choice model, except for the constant term β_0. Within this equation, β_0 represents an alternative specific constant for both A and B, as opposed to the opt-out. The systematic utility of both A and B are modelled using the same constant term because the separate alternative specific constants for scenario A and B did not significantly differ from each other (based on the Wald test statistic).

[Eq. 1.4]

$V_{alternative\ A} = \beta_1$ * flexible menu schedule + β_2 * elaborate menu schedule + β_3 * flexible PA schedule + β_4 * elaborate PA schedule + β_5 * consults in group of 5 + β_6 * consults in groups of 10 + β_7 * expected outcome + β_8 * out-of-pocket costs

[Eq. 1.5]

$V_{alternative\ B} = \beta_1$ * flexible menu schedule + β_2 * elaborate menu schedule + β_3 * flexible PA schedule + β_4 * elaborate PA schedule + β_5 * consults in group of 5 + β_6 * consults in groups of 10 + β_7 * expected outcome + β_8 * out-of-pocket costs

[Eq. 1.6]

$V_{opt\text{-}out} = 0$

Choice consistency of the study population was checked [46-49], as every respondent answered every choice task twice (once they were forced to choose and once they were offered an opt-out option). Respondents satisfied the consistency measure if they chose the same option for every choice task in the first and second part of the questionnaire. Respondents who chose to opt-out were automatically marked consistent.

Differences in opting out between respondents with a lower (i.e. primary education or lower secondary education) and higher (i.e. all other levels) were determined using independent sample t-tests.

Marginal willingness-to-pay (WTP) values were determined for all statistically significant attribute estimates of the main analysis. These results can be compared directly between the forced-choice DCE and the DCE including the opt-out option. In order to calculate the patient's WTP, the negative of the out-of-pocket attribute was used as a measure of the marginal utility of money. The ratio of either attribute estimate with this negative of the out-of-pocket attribute was calculated to estimate the patient's WTP concerning that specific attribute [3, 50].

NLogit 5.0 (Econometric Software, New York) was used to construct the models that were estimated within this study. Results were considered statistically significant when $p<0.05$.

RESULTS

Study population

Of all 2,500 distributed questionnaires, a total number of 781 (31.2%) questionnaires were returned and included in the analysis. The demographic and disease-specific characteristics of the study population are summarized in Table 2. Respondents were aged 57.8 years on average and mainly Dutch (92.7%). Approximately half of the population was male (55.1%) and most respondents had

an intermediate educational level (48.6%). With regard to the disease status of the respondents, they reported to have been diagnosed with DM2 on average 6.5 years ago and 79.7% used some form of medication. The mean BMI was 29.5 kg/m^2 and the mean HbA_{1c} was 49.1 mmol/mol (target value for DM2 patients is <53 mmol/mol). The majority reported no complications due to their DM2 (75.6%) and the mean EQ-5D score was relatively high with 0.91. There were no clinically relevant differences in these demographic and disease-related variables between the respondents who completed the different versions of the questionnaire.

Table 2. General description of the study population (N=781)

		Mean (SD)	**Percentage**
Age		57.8 (6.27)	
Gender	Male		55.1
Highest attained educational level	Low		31.2
	Medium		48.6
	High		20.2
Ethnicity	Dutch		92.7
Duration of diabetes (years)		6.5 (5.97)	
No complications			75.6
Medication	None		20.3
	Oral glucose lowering medication		66.7
	Insulin		4.0
	Both		9.0
BMI (kg/m^2)		29.5 (5.19)	
HbA_{1c} (mmol/mol)		49.1 (14.0)	
EQ5d score		0.91 (0.19)	

Choice behavior

With respect to the raw choice percentages, on average over all choice tasks, 54% of the respondents were willing to participate in a lifestyle program when the option to opt-out was offered.

Respondents did not have a strong tendency to choose either option A or B. This was expected as A and B were unlabeled and therefore did not have specific characteristics which would make one of them more attractive within every choice tasks for the same reason (e.g., A was not always cheaper than B or the other way around). The percentages of respondents choosing option A and B both dropped when the opt-out option was included. However, in 13 of the 18 choice tasks the difference in percentage of individuals moving from option A to the opt-out option and the

individuals moving from option B to the opt-out option was more than 5%.

In total 21.6% of the respondents always chose to opt-out, while 22.8% never chose to opt-out within the nine unforced choice tasks. Respondents with a lower educational level significantly more often chose to opt out compared to respondents with a higher educational level (t=2.31; P<.05). Except for choice task one and two, the frequency of choosing the opt-out option was significantly higher among respondents who first had the option to choose to opt-out and then were forced to make a choice, compared to respondents who first were forced to make a choice and later were able to choose the opt-out option (t=-2.94; P<.05) (Figure 1).

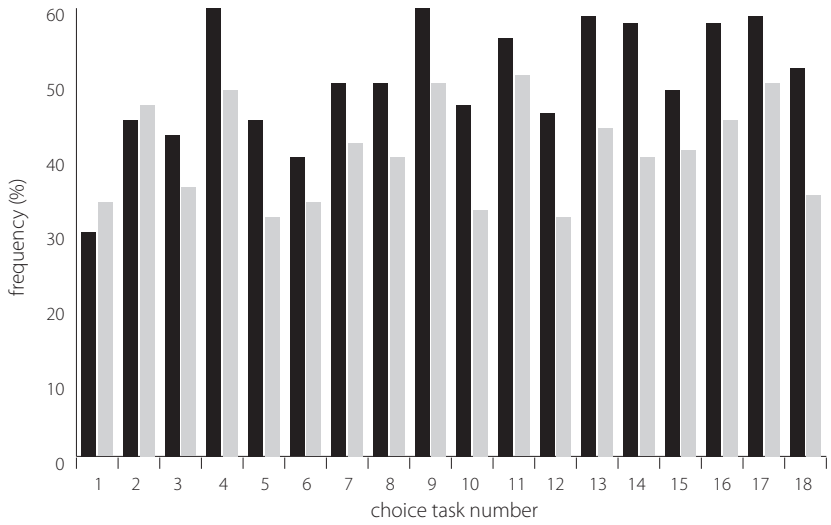

Figure 1. Proportion choosing to opt-out per choice task, stratified by respondents that started with the unforced choices (black) and respondent that were offered the opt-out in the second part of the questionnaire (grey).

Choice consistency

With respect to the consistency of the respondents when making their choices, overall 83.4% of the population answered consistently on all but one choice task. In total, 65.2% answered consistently on all choice tasks, while only 0.2% was consistent in six out of nine choice tasks (this was the lowest consistency score). The consistency measure was 93% on average per choice task.

Attribute estimates and relative importance

Within both the forced and unforced model, respondents did not have any specific preferences with respect to a menu schedule (Table 3). They reported a strong dislike for a flexible PA schedule, compared to a more general schedule. An elaborate PA schedule was preferred over a general schedule only in the opt-out model. A consultation structure in which the respondents would

work individually was preferred over consultations organized in groups with 10 other patients. Consultations in groups of five were preferred over groups of 10 only in the forced-choice model. In both models, respondents were more likely to choose a lifestyle program that was expected to result in higher weight loss and programs with the lowest out-of-pocket costs. Both models showed that out-of-pockets costs (per €100) was the most important attribute followed by expected outcome (per 10kg of weight loss) consultation structure and PA schedule. Both models show significant ($p<0.01$) preference heterogeneity among respondents concerning the attribute that reflects the expected outcome of the lifestyle program in terms of weight loss, as was shown by the statistically significant standard deviation of this attribute in both datasets.

Table 3. Attribute estimates (standard errors) of the MIXL model

		Forced choice model		**Opt-out model**	
Attribute		**Beta value**	**SE**	**Beta value**	**SE**
Constant	Mean	-0.11***	0.04	-0.30***	0.11
	Standard deviation	0.59	0.38	2.71***	0.52
Menu schedule	Flexible	-0.06	0.06	0.02	0.05
	General (ref)	0.10	0.12	0.02	0.04
	Elaborate	-0.04	0.04	-0.04	0.04
Physical activity schedule	Flexible	-0.25***	0.05	-0.14***	0.04
	General (ref)	0.18	0.07	0.01	0.05
	Elaborate	0.08	0.05	0.13***	0.04
Consultation structure	Individual (ref)	0.50***	0.06	0.51***	0.06
	Groups of 5	0.10**	0.05	0.02	0.04
	Groups of 10	-0.60***	0.08	-0.53***	0.06
Expected outcome (10kg)	Mean	0.75***	0.13	0.76***	0.11
	Standard deviation	2.82***	0.59	1.63***	0.36
Out-of-pocket costs (€100)		-0.97***	0.10	-0.78***	0.06

** Significant at p<.05; *** Significant at p<.001

Location was included as a covariate in the analysis and turned out to be statistically significant ($p<0.001$), indicating differences in attribute estimates and their relative importance based on the location of the opt-out option (i.e. either in the first or the second part of the DCE questionnaire). Therefore, location was included as an interaction term with all of the attributes, which resulted in a significant estimate with; out-of-pocket costs, expected outcome, consultation structure and menu schedule. Respondents who answered the choice tasks with opt-out first, showed stronger preferences for out-of-pocket costs and menu schedule and less pronounced preferences for the expected outcome and consultation structure compared to the respondents who first completed

the forced-choice tasks. Finally, attribute estimates were analyzed separately for these two groups (Table 4). Except for the menu schedule attribute, results are highly overlapping. Preference heterogeneity was shown in both models for the expected outcome attribute.

Table 4. Attribute estimates (standard errors) of the MIXL stratified by location of the opt-out

		Opt-out in first part		**Opt-out in second part**	
Attribute		**Estimate**	**SE**	**Estimate**	**SE**
Constant	Mean	0.11	0.13	-0.91***	0.25
	Standard deviation	1.93***	0.75	3.31***	0.97
Menu schedule	Flexible	0.12*	0.07	-0.08	0.08
	General (ref)	-0.04	0.03	0.11***	0.00
	Elaborate	-0.07	0.06	-0.02	0.06
Physical activity schedule	Flexible	-0.13**	0.06	-0.14**	0.07
	General (ref)	0.04	0.01	0.01	0.01
	Elaborate	0.09	0.06	0.13**	0.06
Consultation structure	Individual (ref)	0.48***	0.01	0.56***	0.01
	Groups of 5	-0.06	0.06	0.09	0.07
	Groups of 10	-0.42***	0.09	-0.65***	0.10
Expected outcome (10kg)	Mean	0.61***	0.19	0.87***	0.15
	Standard deviation	1.65***	0.75	2.12***	0.50
Out-of-pocket costs (€100)		-0.71***	0.08	-0.86***	0.09

* Significant at p<.10; **significant at p<.05; *** Significant at p<.001

Sensitivity analysis

To obtain more insight into the robustness of the retrieved results, three separate additional analyses were performed. First, respondents who answered inconsistently in more than one choice task (16.6%) were excluded from the analysis. Secondly, respondents were excluded if they indicated (in the first part of the questionnaire) that they would probably not or certainly not participate in a future lifestyle program (53.6%), no matter what attributes this lifestyle program would have. Third, all respondents who chose the opt-out option in all choice tasks were excluded (21.6%). The results of these analyses did not show any notable differences from the analysis described above with respect to the results on the forced choice data and opt-out data (results not shown).

Willingness to pay

Table 5 shows the different WTP values for the attribute levels that were statistically significant in both the forced-choice and the opt-out model. Based on the forced-choice model, respondents

reported a significant higher WTP estimate for a switch from a flexible PA schedule to a more general schedule, compared to the opt-out model (i.e. respectively 44 and 19 euro per year). Within the forced choice model, the WTP for individual consultations instead of consultation in groups with 10 other patients was approximately €114 per year, while in the opt-out model the WTP was estimated to be approximately €134 per year for this same switch in consultation structure. With respect to an expected additional weight loss of 10 kilograms, the forced-choice model showed a WTP estimate of €77 compared to a WTP estimate of approximately €98 for the opt-out model. Though these WTP values differ on an absolute scale, both difference were not statistically significant as the accompanying confidence intervals overlapped.

WTP values differed between the data where the opt-out was offered first or seconds, but these differences were not statistically significant due to the large confidence intervals around these estimates (Table 5).

Table 5. Willingness to pay values for all tested models from table 3 and 5.

	Switch from a flexible to a general PA schedule		**Switch from consultation in groups of 10 to individual consultation**		**Every 10kg of extra anticipated weight loss**	
	WTP (€)	**95%CI**	**WTP (€)**	**95%CI**	**WTP (€)**	**95%CI**
Forced model	44.3*	29.6 ; 59.0	113.5	68.4 ; 158.5	77.2	56.7 ; 97.8
Opt-out model	19.3*	13.5 ; 25.0	134.1	100.2 ; 168.1	98.1	69.8 ; 126.3
Opt-out in first part of the DCE	23.7	17.1 ; 30.4	125.8	75.8 ; 175.9	85.4	32.9 ; 137.9
Opt-out in second part of the DCE	17.8	8.9 ; 26.6	140.7	88.5 ; 192.8	101.2	66.0 ; 136.4

* Significant at p<.05

DISCUSSION

Results show differences in opting out based on the location of the opt-out option and the educational level of the respondents. The attribute estimates of the forced-choice and opt-out dataset differed, but no notable differences in the relative order of the attributes (as compared to each other) were present. However, because the importance weights of the attributes did differ between the datasets, there is a statistically significant difference in the WTP of patients for a PA schedule. This difference could lead to different conclusions and recommendations with regard to developing a lifestyle program that is most attractive for the target group.

Current findings underline the statements of Dahr and colleagues (2003) that the independence of irrelevant alternatives (IIA) assumption for forced-choice data does not hold in unforced data [9]. If this assumption would hold, including an opt-out option would not change study outcomes as it would take equal proportions of all attribute estimates [3, 6, 9]. If this were to be the case, including an opt-out option would not be necessary for the accurate prediction of a patient's' preferences. It is, however,

more likely that adding an opt-out option to a forced choice model will disproportionally change study results because this option competes more with one scenario than the other in the same choice task [51]. Current results confirm that this IIA assumption does not hold because a disproportional shift was shown for choosing option A or B in the forced choice tasks and then moving to opt-out in the unforced choice tasks. Moreover, study results change slightly if an opt-out option is included.

While the direct results of the DCE might not have differed much between the forced and opt-out data, the analysis of the influence of the location of the opt-out option (either in the first nine choice tasks or the second nine choice tasks), showed clearly that the location of the opt-out option in the questionnaire influences the results of a DCE. This was also shown in previous research [13]. The fact that, in general, fewer people chose to opt-out when this was offered in the second part of the questionnaire might be interpreted as a learning effect that respondents go through when completing a DCE. These respondents were forced to make a choice at first, so they became familiar with completing a DCE choice task. Respondents might therefore have had a lower tendency to opt-out when this option was offered later on (unless they really did not want to participate in a lifestyle intervention). These results are in line with findings from previous research, that indicate a decrease in negative emotions and decisions to opt-out when individuals become more familiar with making trade-offs [9, 11-13, 16]. It was hypothesized that respondents with a lower education would more often find choice tasks to be complex; this study showed that respondents with a lower educational level significantly more often opted out. This result underlines our hypothesis that respondents more often opt-out if they find the choice tasks complex.

Additionally to the effect on the data by simply including such an opt-out option, our study results indicate that choice behavior changes which influences DCE results when respondents are given the opportunity to opt-out. Including an extra choice option automatically implies reduced effectiveness, as there are more answering categories included. Specifically an opt-out option does not provide any insight on attribute level trade-offs. This is not an issue, if the choice to opt-out is due to the low perceived personal utility for the other scenarios. However, our analysis showed that it is likely to assume that a considerable number of respondents chose to opt-out for other reasons than a dislike for lifestyle programs. It can therefore be suggested that including an opt-out option in a DCE, leads to an 'unnecessary' loss of effectiveness. This is of special interest in the light of designing DCEs in an efficient manner (e.g., by minimizing D-error). Such designs strive to create choice tasks with an optimal utility balance between the scenarios of each choice task, by optimizing the variance-covariance matrix. This designing procedure results in a DCE that requires a minimal number of choice tasks per respondent and a minimal number of respondents per experiment (aside from model specifications (e.g., level restrictions or interactions)). At the same time, this may induce complexity of the generated choice tasks. Since there are indications for higher levels of opting out when choice tasks become more complex, the efficiency of designing DCEs in such a way may be at risk. Future research is necessary to identify subgroups among study populations that are most likely to opt-out due to other reasons than solely personal utility. Moreover, it should be explored how DCEs can be designed in an efficient manner while keeping in mind this phenomenon.

The current study has some limitations. Although an efficient DCE design was developed, an even more efficient design could have been created if the beta priors retrieved with the pilot-study were more stable. If more informed beta priors would have been used (instead of using zero as a beta prior for all attributes), the expected preferences of the target population would already be included in the design of the DCE. This way, the design varies the attribute levels based on their relative importance as displayed in the beta priors, resulting in more complex choice tasks. Since choice task complexity is expected to drive the choice to opt-out, using a more efficient design would probably have led to even more individuals that chose to opt-out and thereby more pronounced differences between the results of the model with and without an opt-out option. The response rate was 31.2%. Due to confidentiality agreements with the care groups that distributed the questionnaires, no reminder letters could be distributed and a non-response analysis could not be conducted. Non-response is likely to be selective, in the sense that DM2 patients who are not interested in a lifestyle program were less likely to participate in this study. It is therefore expected that this selective response resulted in an underestimation of the differences between the datasets with and without an opt-out option. Although the conclusions of this paper would probably not change, they might have been more pronounced if respondents with a negative attribute towards lifestyle programs had participated. The current study included DM2 patients in the age category 35-65. There is very limited information on the representativeness of the current study population compared to the target population. Additional analysis of DM2 patients aged 35-65 in a large Dutch Cohort study (EPIC-NL [52]) showed the same mean BMI values, It was not possible to compare other characteristics due to specific inclusion criteria of the EPIC-NL study. However, it is expected in the current sample that especially the number of respondents with a non-Dutch origin (7.3%) is relatively low compared to the average population of DM2 patients aged 35-65 years. If the hypothesized learning effect (i.e. people choose to opt-out because they do not understand the questions in a DCE or if the choice that has to be made is too difficult) is indeed present, it could very well be that the differences in results due to the inclusion of an opt-out option are underestimated assuming that, due to language difficulties, non-Dutch DM2 patients were more likely to opt-out.

In general, the choice for including an opt-out option in DCEs, depends evidently on the research objective. When the research objective is to determine the potential participation in a health program, an opt-out option should always be included; if in real life 'not participating' is an option as well. By doing so, researchers stay as close as possible to the actual choices of their target population. Introducing an additional loss of power, because respondents do not make any trade-offs and chose to opt-out, should then be accepted. However, the number of respondents that opt-out for other reasons than aiming for the highest personal utility should be minimized. Based on the learning effect that was shown in this study, future DCEs that include an opt-out option may want to incorporate multiple forced choice warm-up exercises. However, since DCE questionnaire are already cognitively demanding and time consuming, a more efficient solution might be to use a dual response design. In such a design, respondents are forced to make a choice and immediately after choosing, respondents are asked if they would like to opt out if given the choice [53-55]. This might diminish the risk that a direct introduction of an opt-out results in large numbers

of respondents avoiding to seriously weigh the different levels of attributes. Additionally, in order to minimize the proportion of respondents that chooses to opt-out because they find the choice tasks too complex or difficult, future research should empirically explore how choice tasks should be presented to make them as easy and less complex as possible. Finally, additional research that uses debriefing of respondents should be conducted to explore the reasons for choosing the opt-out alternative in depth.

References

1. Lancsar, E. and J. Louviere, *Conducting discrete choice experiments to inform healthcare decision making: a user's guide.* Pharmacoeconomics, 2008. 26(8): p. 661-77.
2. Ryan, M. and S. Farrar, *Using conjoint analysis to elicit preferences for health care.* BMJ, 2000. 320(7248): p. 1530-3.
3. Ryan, M., K. Gerard, and M. Amaya-Amaya, *Using Discrete Choice Experiments to Value Health and Health Care.* The Economics of Non-Market Goods and Resources, ed. I.J. Bateman. 2008, Dordrecht: Springer.
4. de Bekker-Grob, E.W., M. Ryan, and K. Gerard, *Discrete choice experiments in health economics: a review of the literature.* Health Econ., 2012. 21(2): p. 145-72.
5. Cascetta, E., *Random Utility Theory, in Transportation Systems Analysis: models and applications,* E. Cascetta, Editor. 2009, Springer: New York. p. 89-167.
6. Louviere, J.J., D.A. Hensher, and J.D. Swait, *Stated Choice Methods; analysis and apllication.* 2000, Cambridge: Cambridge University Press.
7. de Bekker-Grob, E.W. and C.G. Chorus, *Random regret-based discrete choice modelling: An application to health care.* PharmacoEconomics, 2013. 31(7): p. 623-34.
8. Karni, E. and A. Schwartz, *Search theory: the case of search with uncertain recall.* J Econ Theory, 1977. 16: p. 38-52.
9. Dhar, R. and S. Itamar, *The effect of forced choice on choice.* J Market Res, 2003. 40 (2): p. 146-160.
10. Krosnick, J.A., et al., *The impact of "no opinion" response options on data quality: non-attitude reduction or an invitation to satisfice?* Public Opin Quart, 2002. 66: p. 371-403.
11. Dhar, R., *Consumer preference for a no-choice option.* J Cons Res, 1997. 24(2): p. 215-231.
12. Luce, M.F., J.W. Payne, and J.R. Bettman, *Emotional trade-off difficulty and choice.* J Marke Res, 1999. 36(2): p. 143-159.
13. Boxall, P., W.L. Adamowicz, and A. Moon, *Complexity in choice experiments: choice of status quo alternative and implications for welfare measurement.* Aust J Arg Resour Ec, 2009. 53: p. 503-519.
14. Iyengar, S.S. and E. Kamenica, *Choice overload and simplicity seeking.* Working paper, 2007.
15. Kahneman, D., J.L. Knetsch, and R.H. Thaler, *The endowment effect, loss aversion and status quo bias.* J Econ Perspec, 1991. 5(1): p. 193-206.
16. Luce, M.F., *Choosing to avoid: coping with negative emotion-laden consumer desicions.* J Cons Res, 1998. 24(4): p. 409-433.
17. Meyerhoff, J. and U. Liebe, *Status quo effect in choice experiments: emperical evidence on attitude and choice task complexity.* Land Econ, 2009. 85(3): p. 515-528.
18. Nowlis, S.M., B.E. Kahn, and R. Dhar, *Coping with ambivalence: the effect of removing a neutral option on consumer attitude and preference Judgements.* J Cons Res, 2002. 29(3): p. 319-334.
19. Ritov, I. and J. Baron, *Status quo and ommission biases.* J Risk Uncertainty, 1992. 5: p. 49-61.
20. Tversky, A. and E. Shafir, *Choice under conflict: the dinamics of differed discision.* Psych Sci, 1992. 3(6): p. 358-361.
21. van Heafen, R.H., D.M. Massey, and W.L. Adamowics, *Serial nonperticipation in repeated discrete choice models.* Am J Agr Econ, 2005. 87(4): p. 1061-1076.
22. Cook, J., et al., *Reliability of stated preferences for cholera and typhoid vaccines with time to think in Hue, Vietnam.* Econ Inq, 2007. 45(1): p. 100-114.
23. Ciliska, D., H. Thomas, and C. Buffet, *An Introduction to Evidence-Informed Public Health and A Compendium of Critical Appraisal Tools for Public Health Practice (Revised).* Hamilton, 2012. ON: p. National Collaborating Centre for Methods and Tools.
24. Elwyn, G., D. Frosch, and S. Rollnick, *Dual equipoise shared decision making: definitions for decision and behaviour support interventions.* Implement Sci., 2009. 4(75).
25. Yousefi-Nooraie, R., et al., *Information seeking for making evidence-informed decisions: a social network analysis on the staff of a public health department in Canada.* BMC Health Serv Res., 2012. 12(118).
26. Struijs, J.N. and C.A. Baan, *Integrating care through bundled payments--lessons from The Netherlands.* N Engl J Med, 2011. 364(11): p. 990-1.
27. Dube, M.C., et al., *Physical activity barriers in diabetes: development and validation of a new scale.* Diabetes Res Clin Pract, 2006. 72(1): p. 20-7.

28. Forbes, C.C., et al., *Physical activity preferences and type 2 diabetes: exploring demographic, cognitive, and behavioral differences.* Diabetes Educ, 2010. 36(5): p. 801-15.
29. Lakerveld, J., et al., *Motives for (not) participating in a lifestyle intervention trial.* BMC Med Res Methodol, 2008. 8: p. 17.
30. Nagelkerk, J., K. Reick, and L. Meengs, *Perceived barriers and effective strategies to diabetes self-management.* J Adv Nurs, 2006. 54(2): p. 151-8.
31. Owen, K., et al., *Individual preferences for diet and exercise programmes: changes over a lifestyle intervention and their link with outcomes.* Public Health Nutr, 2010. 13(2): p. 245-52.
32. Gils van, P.F., et al., *Financial incentives and other factors on willingness to participate in a lifestyle intervention program in diabetic patients: a conjoint analysis.* Patient Pref Adher, 2011. 5: p. 537-546.
33. Roux, L., et al., *Valuing the benefits of weight loss programs: an application of the discrete choice experiment.* Obes Res, 2004. 12(8): p. 1342-51.
34. Thomas, D.E., E.J. Elliott, and G.A. Naughton, *Exercise for type 2 diabetes mellitus.* Cochrane Database Syst Rev, 2006. 3: p. CD002968.
35. Thomas, N., E. Alder, and G.P. Leese, *Barriers to physical activity in patients with diabetes.* Postgrad Med J, 2004. 80(943): p. 287-91.
36. Vijan, S., et al., *Barriers to following dietary recommendations in Type 2 diabetes.* Diabet Med, 2005. 22(1): p. 32-8.
37. Wood, F.G., *Ethnic differences in exercise among adults with diabetes.* West J Nurs Res, 2002. 24(5): p. 502-15.
38. Bliemer, M.C.J. and J.M. Rose, *Efficiency and sample size requirements for stated choice experiments, in Transportaion Research Broad Annual Meeting.* 2009: Washington DC.
39. Huber, J. and K. Zwerina, *The Importance of Utility Balance in Efficient Choice Designs.* J Marke Res, 1996. 33(3): p. 307-317.
40. Cheraghi-Sohi, S., et al., *Making sense of patient priorities: applying discrete choice methods in primary care using 'think aloud' technique.* Fam Pract, 2007. 24: p. 276-82.
41. Coast, J. and S. Horrocks, *Developing attributes and levels for discrete choice experiments using qualitative methods.* J Health Serv Res Policy, 2007. 12(1): p. 25-30.
42. Marshall, D., et al., *Conjoint Analysis Applications in Health - How are Studies being Designed and Reported?: An Update on Current Practice in the Published Literature between 2005 and 2008.* Patient, 2010. 3(4): p. 249-56.
43. Lamers, L.M., et al., *[Measuring the quality of life in economic evaluations: the Dutch EQ-5D tariff].* Ned Tijdschr Geneeskd, 2005. 149(28): p. 1574-8.
44. Louviere, J.J., et al., *Designing discrete choice experiments: do optimal designs come at a price?* J Cons Res, 2008. 35(2): p. 360-375.
45. Bech, M. and D. Gyrd-Hansen, *Effects coding in discrete choice experiments.* Health Econ, 2005. 14(10): p. 1079-83.
46. Lancsar, E. and J. Louviere, *Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences?* Health Econ, 2006. 15(8): p. 797-811.
47. Miguel, F.S., M. Ryan, and M. Amaya-Amaya, *'Irrational' stated preferences: a quantitative and qualitative investigation.* Health Econ, 2005. 14(3): p. 307-22.
48. Ryan, M., V. Watson, and V. Entwistle, *Rationalising the 'irrational': a think aloud study of discrete choice experiment responses.* Health Econ, 2009. 18(3): p. 321-36.
49. Sen, A., *Internal Consistency of Choice.* Econometrica, 1993. 61(3): p. 495-521.
50. Revelt, D. and K.E. Train, *Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level.* Rev Econ Stat, 1998. 80(4): p. 647-57.
51. Shafir, E., I. Simonson, and A. Tversky, *Reason based choice.* Cognition, 1993. 49: p. 11-36.
52. Beulens, J.W., et al., *Cohort profile: the EPIC-NL study.* Int J Epidemiol., 2010. 39(5): p. 1170-8.
53. Brazell, J.D., et al., *The no-choice option and dual response choice designs.* Market Lett, 2006. 17: p. 255-268.
54. Marshall, D.A., et al., *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., 2009. 18(12): p. 1420-39.
55. Marshall, D.A., et al., *Measuring patient preferences for colorectal cancer screening using a choice-format survey.* Value Health., 2007. 10(5): p. 415-30.

CHAPTER 8

WORDS OR GRAPHICS TO PRESENT A DISCRETE CHOICE EXPERIMENT: DOES IT MATTER?

J. Veldwijk, M.S. Lambooij, J.A. van Til,
G.C.M. Groothuis-Oudshoorn, H.A. Smit, G.A. de Wit

Submitted.

ABSTRACT

Objective: To test whether presenting attribute levels in words or graphics generates different results with respect to attribute level interpretation, relative importance and participation probabilities.

Methods: Parents of 959 newborns completed a DCE questionnaire that contained two versions of the same nine choice tasks in which the attribute levels were presented in words or graphics. Five attributes related to the decision of parents to vaccinate their newborn against rotavirus were included. Panel mixed-logit models were conducted to estimate the relative importance of the attribute levels.

Results: Respondents who started with the choice tasks in words produced the most consistent answer patterns. All respondents significantly preferred words to graphics. Part-worth utilities and the relative importance of the attribute levels differed based on the words and graphics data, resulting in different probabilities to participate in vaccination.

Conclusions: Words were preferred over graphics, resulted in higher choice consistency, and showed more valid attribute level estimates. Graphics did not improve respondents' understanding of the attribute levels. Future research on the use of either words or graphics is recommended in order to establish guidelines on how to develop a valid presentation method for attribute levels in the choice tasks of a DCE.

INTRODUCTION

There is an ongoing discussion about the complexity of choice tasks in health-related Discrete Choice Experiments (DCEs) and the extent to which respondents are capable of completing those choice tasks as intended by researchers [1-4]. At the same time, there has been an increase in the use of DCEs within the public health and health care research setting [5, 6]. Those DCEs aim to elicit respondents' preferences in order to advise on the development of preventive programs, medical therapies and/or policy measures [7-9]. Since DCEs are used for policymaking, the accuracy and validity of the measured (i.e. stated) preferences is essential. It is therefore vital that respondents understand the medical and/or health related information that is included, in order to make accurate choices that reflect their true preferences. The validity of a DCE is at risk if respondents do not fully understand how to complete the choice tasks, because they lack understanding of the attribute levels within the DCE.

There is great diversity in the way attribute levels are explained to respondents and how choice tasks are presented in DCEs on prevention or health related topics [6]. Some researchers have pointed out that the use of graphics or icons might help to make choice tasks easier and therefore improve respondents' understanding of the concepts in question [10-12]. Although there is no empirical evidence within the literature to support this assumption, these suggestions probably stem from the large amount of research on improving risk communication to enhance shared and informed decision-making as well as self-management [13-23]. Those studies showed that individuals in general but specifically individuals with a lower educational level and/or health literacy have greater difficulties with risk and health information displayed in words or numbers compared to graphics [13, 17, 24]. This is of particular importance for DCEs within the public health and health care setting, because individuals with a low educational level and health literacy use public health and/or health care interventions relatively more often [25, 26]. Studies investigating the validity of and preferences for communicating health related information showed that the use of icons or graphics might be helpful, if designed properly [13, 15, 16, 18-20, 22]. Currently, evidence on the effectiveness of depicting attribute levels using graphics within a DCE context is lacking. Pending such evidence, graphics are used under the assumption that individuals will be able to correctly decipher the actual numerical information captured in the graphics, to interpret the information and to reveal their preferences accordingly.

This study empirically tested whether DCE results with respect to attribute level interpretation, relative importance and participation probabilities differ when either words or graphics are used to present attribute levels in the choice tasks. Specifically, it was tested whether those results differ among respondents with a different educational level and health literacy status.

MATERIALS AND METHODS

Subject of the DCE and participant recruitment

A DCE on parental preferences for rotavirus vaccination among newborns was used as a case for this study. A random sample of 2500 parents of newborn babies aged six weeks was selected from a national vaccination register (Praeventis) to receive the DCE questionnaire. The Institutional Review Board of the University Medical Centre Utrecht concluded that formal testing by a medical ethical committee was not necessary, as parents were only required to complete an anonymous questionnaire once, which is in accordance with the guidelines laid down in the Declaration of Helsinki.

Attribute	Level 1	Level 2	Level 3
Vaccine effectiveness: The percentage of children that will be protected against a rotavirus infection when vaccinated	55%	75%	95%
			
Frequency of severe side effects: The number of vaccinated children that will suffer from intussusception due to vaccination. Intussusception is an acute condition in which part of the bowel telescopes into another adjacent part of the bowel, resulting in obstruction*.	1 in 10,000	1 in 100,000	1 in 10,000,000
			
Protection duration: The number of years that the vaccine protects against a rotavirus infection	1 year	3 years	6 years
			
Location: Within the Netherlands all vaccines in the NIP are administrated at a child welfare center, The GP office was included because the rotavirus vaccine may not become part of the NIP; in that case it is likely that this vaccine is administrated here.	General practitioner	Child welfare center	
			
Costs: Parents may have to pay (part) of the vaccine costs out-of-pocket	€0	€30	€140
	-	20 10	50 50 20 20

Figure 1. All attributes in levels as described in the questionnaire in words and graphics

* See reference list [45]

Attributes, levels and choice task presentation

Attributes and levels were identified based on previously published literature [27-34], interviews with experts (i.e. a pediatrician with specific interest in rotavirus infections and a scientist with specific interest in vaccination behavior), and four group interviews with in total 28 parents of newborns. Five attributes were selected for this DCE (Figure 1). A professional designer designed the graphics. Common symbols were used for the qualitative levels, and quantitative levels were depicted as close as possible to their actual number (Figure 1). The development of the graphics was further supported by extensive literature research, and consultations with an expert on scientific communication strategies and icon development.

Before participants were asked to complete the choice tasks, they received detailed information to enhance understanding of all attributes and levels as well as an explanation on how to complete a choice task, illustrated by an example. Also extensive guidance was provided on the meaning and interpretation of the graphics that were included. Every choice task started with the question: 'Imagine that a vaccine is available against rotavirus infections, in which scenario would you prefer to vaccinate your newborn, scenario 1 or scenario 2? After each choice task, participants were given the opportunity to state that they preferred not to vaccinate their child (opt-out).

Experimental design and pilot testing

NGene 1.0 (ChoiceMetrics, 2011) software was used to construct a D-efficient design [9, 35]. The DCE draft questionnaire consisted of 18 unique choice tasks. Ngene divided these 18 choice tasks into two blocks of nine choice tasks (i.e. block 1 and block 2), and each block was randomly distributed among half of the study population. A crossover design was used, so every respondent answered every choice task twice. All respondents answered the nine choice tasks of either block 1 or block 2 in words and the same nine choice tasks in graphics. Whether respondents started with the choice tasks in words or graphics was randomly assigned.

The draft questionnaire was pilot tested among a subgroup (N=48) of our study population to check whether correct wording was used and whether the target population understood the choice tasks, with special emphasis on the understanding of the graphics. The interpretation of the different graphics by respondents was tested, and multiple graphics were presented allowing the respondents to choose their preferred graphic per attribute (level). Four of these pilot tests were 'think aloud' tests, where a researcher was present while the participant completed the questionnaire by reading out loudly and sharing all thoughts, ideas and opinions. Finally, the attribute level estimates that were retrieved from the pilot study served as priors for the design of the final DCE questionnaire.

Questionnaire

The first section of the questionnaire comprised of 30 questions on demographics, among which educational level and three validated health literacy questions [36, 37]. The second part of the questionnaire consisted of the actual DCE. Finally, respondents' level of understanding of the

choice tasks in words and graphics, their preference for either presentation of the choice tasks and their interpretation of the words and graphics (i.e. either as a numeric value a or as a low-medium-high categorized variable) were measured.

Statistical analysis

Consistency of choices

Respondents were marked 'consistent' when they chose the same scenario within an identical choice task in words and graphics. Differences in the total number of inconsistent answers between the respondents that first answered the choice tasks in words and those who started with the graphic choice tasks was analyzed using independent sample t-tests.

Attribute level interpretation

Part-worth utilities for all attribute levels were estimated to analyze whether respondents interpreted them as an actual numerical value or on an ordinal scale. For that purpose a dummy variable representing the highest level of each of the attributes was added to the regression model. The difference in part-worth utilities between level one and two is significantly different from the part-worth utilities between level two and three (and thus nonlinear), if the dummy variable is significant.

Attribute estimates and relative importance

Both panel and generalized mixed logit models were constructed to adjust for the multilevel structure of the data, to be able to correct for preference heterogeneity as well as variability in individual errors (scale heterogeneity) [38].

All non-linear attributes (i.e. frequency of severe side effects, protection duration, location and out-of-pocket costs) were recoded using effects codes [9, 39]. Based on model fit tests (AIC, BIC, Log likelihood) it was tested which model fitted best to the data. Based on the significance level of the standard deviation of the attributes it was tested what attributes should be included as random parameters due to significant preference heterogeneity and what distribution should be assumed for those parameters. Based on this analysis the constant of the model and the out-of-pocket cost attribute levels were included as random parameters in the model with normal distributions. The systematic utility component (V), which describes the measurable utility of a specific vaccine based on the attributes that were included in the DCE, was testing using the following equation:

$$V = \beta_{0i} + \beta_1 * \text{vaccine effectiveness} + \beta_2 * \text{severe side effects}_{\text{1 in 100,000}} + \beta_3 * \text{severe side effects}_{\text{1 in 10,000}} + \beta_4 * \text{protection}_{\text{3 years}} + \beta_5 * \text{protection}_{\text{6 years}} + \beta_6 * \text{location}_{\text{GP}} + \beta_{7i} * \text{out-of-pocket costs}_{€30} + \beta_{8i} * \text{out-of-pocket costs}_{€140}$$

The β_0 represents the alternative specific constant and $\beta_1 - \beta_8$ are the attribute level estimates that indicate the relative importance of each attribute level. The utility for the opt-out option was set to be zero.

Using the above equation three main modelling steps were undertaken. First, based on the complete dataset, it was tested whether attribute levels significantly differed between the words and graphics choice tasks. Therefore, a variable that identified the use of either words or graphics was included in the model as a covariate and as an interaction term with the attribute levels. Second, to test whether the order in which respondents answered the choice tasks (i.e. words first or graphics first) influenced the outcomes, a variable that identified the order in which respondents completed the choice tasks was included in the model as a covariate and as an interaction term with the attribute levels. Third, separately for the data on the words choice tasks and the graphics choice tasks the attribute estimates as well as their relative importance was estimated.

Educational level and health literacy

The effects of educational level and health literacy score on respondents' opinion concerning the presentation of the attribute levels, and consistency of choices, was tested using independent sample t-tests.

Participation probabilities

Separately for the words and the graphics data, the probability to participate in a vaccination program was estimated for different vaccine scenarios based on the results of the generalized mixed logit regressions of the two datasets. Since random parameters were included, the probabilities could not be calculated directly but should take into account the standard deviation of those attribute level estimates. Therefore, the mean probability of 10,000 simulations was estimated by taking the average of all simulated probabilities given every tested vaccine scenario, which was calculated as $1/(1+\exp^{-v})$ [7-9]. Intraclass correlation was used to measure the level of agreement between the probabilities of every estimated vaccine scenario based on the two datasets.

RESULTS

Participant characteristics

In total 959 parents (38.4%) returned the questionnaire. Eighteen questionnaires contained more than 10% missing values. After excluding these, 941 respondents (37.6%) were included in the final analyses. Respondents were on average 31.2 years of age and 82.5% of the sample was female. The majority of the respondents had a high educational level (55.7%), reported high health literacy scores (83.8%) and was of Dutch ethnicity (91.1%).

Respondents' opinion on the use of words versus graphics

Although almost all respondents reported that both the choice tasks using words and using graphics were understood and clear to them (Table 1), the majority perceived the use of words in choice tasks as more clear (59.3%) and easier to interpret (54.0%) than the use of graphics. Irrespective of the order in which the choice tasks were presented, respondents reported that the use of words results in significantly more clear and easy choice tasks compared to the use of graphics. Nevertheless, respondents who started with the choice tasks in words significantly more often stated that choice tasks in words were most easy and clear (66.7% & 58.7%) compared

to respondents who started with the graphical choice tasks (51.9% & 49.4%). At the same time, respondents who started with the graphical choice tasks significantly more often stated that choice tasks in graphics were most easy and clear (13.1% & 10.0%) compared to respondents who started with the choice tasks in words (8.4% & 6.7%) (Table 1).

Consistency of choices

Overall, 13.1% of the respondents were consistent in all nine choice tasks (i.e. they chose the same scenario in every choice task in the words and graphics part of the questionnaire). In total, 13.4%, 18.9%, 16.8% and 19.3% of the respondents provided one, two, three or four inconsistent answers respectively. Additionally, 18.5% of the respondents gave five to eight inconsistent answers. Finally, one respondent answered inconsistently for all nine choice tasks. When these results were stratified by order of presentation (i.e. whether respondents started with the choice tasks in words or graphics) a significantly higher number of inconsistencies was found in the respondents who started with the choice tasks in graphics (t= -21.26; p<.001) (Figure 2).

Table 1. Respondents' opinion concerning the attribute level presentation

		Total (%)	**Words first (%)**	**Graphics first (%)**
Choice tasks in words were…	Understood	96.1	95.3	96.8
	Clear	93.4	93.3	93.4
Choice tasks in graphics were….	Understood	91.5	89.7	93.2
	Clear	83.5	80.9*	86.0*
The choice tasks that were most clear used…	Words	59.3	66.7***	51.9***
	Graphics	10.8	8.4***	13.1***
	No difference	30.0	24.9	35.0
The choice tasks that were most easy used…	Words	54.0	58.7*	49.4*
	Graphics	8.3	6.7*	10.0*
	No difference	37.7	34.6	40.7
With the word attributes I interpreted the different levels using…..	Actual values	77.0	78.5	75.4
	Low-medium-high categorization	20.8	19.4	22.2
With the graphic attributes I interpreted the different levels using…..	Actual values	64.5	66.7	62.2
	Low-medium-high categorization	31.1	29.0	33.1

* Significant difference between respondents who answered word choice tasks first compared to respondents who answered graphic choice tasks first, p <0.05.

*** Significant difference between respondents who answered word choice tasks first compared to respondents who answered graphic choice tasks first, p <0.001.

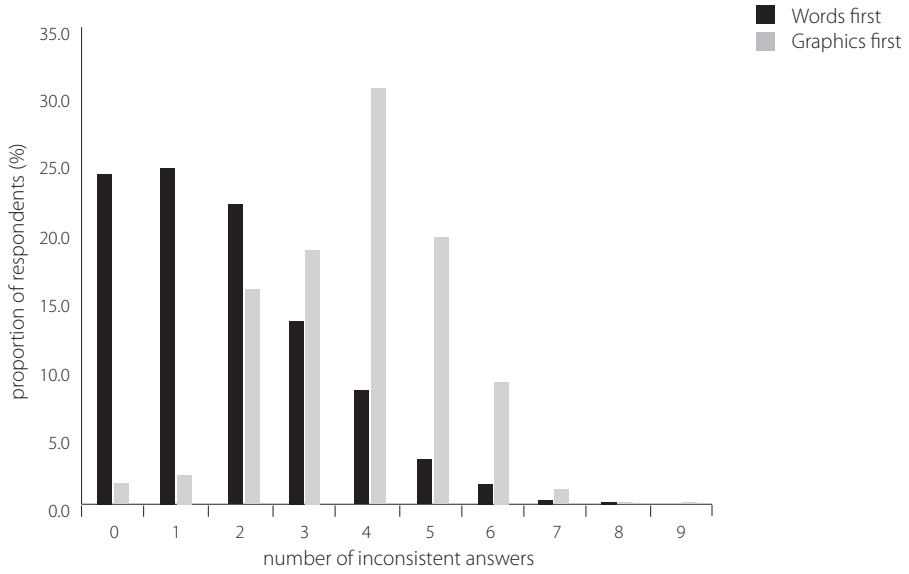

Figure 2. Proportion of inconsistent answers stratified by respondents who started with the choice tasks in words or graphics

Attribute level interpretation

In total, 77.0% of the respondents reported that they interpreted the attribute levels in words according to their actual value, while 20.8% reported they categorized the attribute levels before interpretation (e.g., low/intermediate/high or bad/intermediate/good) (Table 1). With respect to the graphical attribute levels, 64.5% of the respondents reported that they interpreted the attribute levels according to their actual value, and 31.1% reported they used a categorization method when interpreting the attribute levels.

Within the words dataset, the vaccine effectiveness attribute showed equal utility differences between all attribute levels (i.e. insignificant dummy variable), while for all other attributes this was not the case. This was expected given the ranges between the levels of all attributes included. However, within the graphics dataset, all attributes except for protection duration showed equal part-worth utilities between all attribute levels.

Attribute estimates and relative importance

Since there were no differences in the results between the panel mixed-logit and the generalized mixed-logit models, we will only present the results of the generalized mixed-logit models.

First, the variable that identified the use of either words or graphics showed significant interactions with all attributes except for location of vaccine administration, thus using either words or graphics to depict attribute levels significantly influences attribute level estimates. Second, including

interaction terms between order of presentation (i.e. words first or graphics first) and the attribute levels, resulted in a significant interaction with vaccine effectiveness, the frequency of severe side effects and out-of-pocket costs. Due to this ordering effect, all further models only included the data concerning respondents' first set of choice tasks. Third, when analyzing both datasets (either words first or graphics first) separately, results were as expected: parents were more willing to vaccinate their newborn when vaccine effectiveness increases, frequency of severe side effects is lowest, protection duration is longer than one year, and out-of-pocket costs are lowest (Table 2). Parents did not report significant preferences for location of vaccine administration in the words dataset, while in the graphics dataset the child welfare center was preferred to the GP. With respect to the relative importance, in the words dataset out-of-pocket costs were most important this was followed by vaccine effectiveness, protection duration, and frequency of severe side effects. In the graphics dataset the frequency of severe side effects was most important, this was followed by out-of-pocket costs, vaccine effectiveness, protection duration, and location.

Table 2. Outcomes of the generalized mixed logit model separate for the words and graphics dataset

			Words dataset		**Graphics dataset**	
Attribute			**Estimate**	**SE**	**Estimate**	**SE**
Constant		Mean	-4.91***	0.54	-3.24***	0.38
		SD^	3.66***	0.54	2.11***	0.40
Vaccine effectiveness			0.76***	0.07	0.44***	0.03
Frequency of severe side effects	1 in 1,000,000 (ref)		0.73***	0.11	0.99***	0.09
	1 in 100,000		-0.16***	0.06	0.20***	0.05
	1 in 10,000		-0.57***	0.11	-1.19***	0.10
Protection duration	1 year (ref)		-1.02***	0.09	-1.19***	0.09
	3 years		0.67***	0.09	0.71***	0.08
	6 years		0.35***	0.10	0.48***	0.08
Location	General Practitioner (ref)		0.06	0.05	-0.09**	0.04
	Child welfare center		-0.06	0.05	0.09**	0.04
Out-of-pocket costs	0 Euro per year (ref)	Mean	1.19***	0.14	1.16***	0.16
		SD^	1.25	1.21	1.09**	0.40
	30 Euro per year	Mean	0.66***	0.13	0.25**	0.12
		SD	0.60*	0.32	1.04***	0.23
	140 Euro per year	Mean	-1.85***	0.22	-1.41***	0.23
		SD^	1.10***	0.25	0.31	0.26
Tau^^			1.67**	0.78	1.56	1.36

^ Standard Deviation, only presented for the random parameters

^^ Tau (Ţ) is a measure of individual scale heterogeneity within our model.

* P<0.10; ** p <0.05; ***p <0.001

Educational level and health literacy

Compared to lower educated respondents, respondents with a high educational level reported significantly more often that they understood the choice tasks in words (t=7.95) and graphics (t=8.07). They also reported significantly more often that both the words (t=7.65) and graphics (t=5.92) choice tasks were clear to them. These same results were found when comparing respondents with a higher and lower health literacy score (respectively β=-0.56, β=-0.40 & β=-0.48, β=-0.30) (due to the relatively low number of respondents with a low health literacy score (1.3%) the sum score per respondent was used, and linear regressions were performed). Higher educated responders were significantly more consistent in answering the choice tasks compared to lower educated responders (t=-3.59). However, both high-educated and low-educated respondents were significantly more consistent when answering the tasks in words first compared to respondents who answered the choice tasks in graphics first (respectively t=12.56, t=7.40). Although respondents with a higher health literacy also showed a higher number of consistent answers compared to respondents with a lower health literacy, this difference was not statistically significant (β=0.03; p=0.30).

Participation probabilities

Table 3 shows estimated probabilities to participate in different vaccine scenarios, separately for the words and graphics dataset. Although the participation probabilities based on both models show a high correlation (ICC = 0.94; p<0.001), the estimated probabilities based on the graphics model are generally lower compared to the probabilities based on the words model. This is especially pronounced in the scenarios with a short protection duration and a high frequency of severe side effects. The change in participation probabilities when the frequency of severe side effect increases was larger based in the graphics dataset compared to the words dataset (Figure 3).

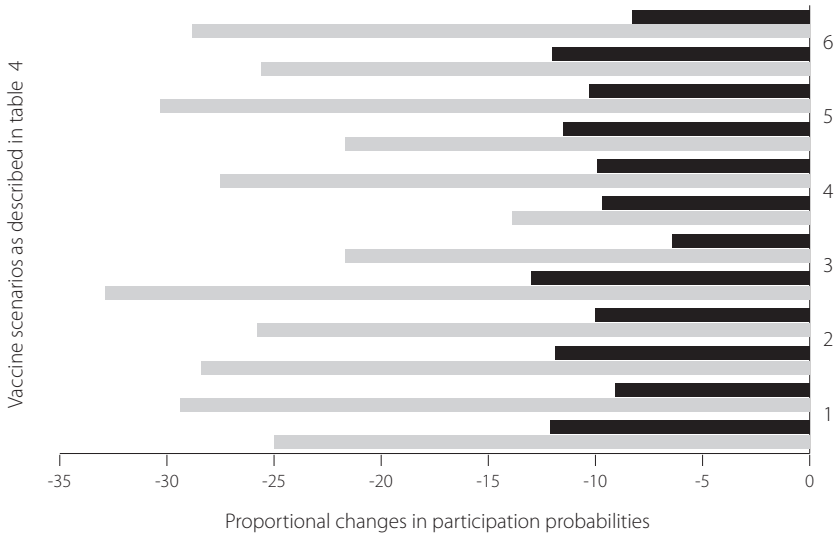

Figure 3. Changes in participation probabilities based on the words model (grey) and graphics model (black) when the frequency of severe side effects increase per vaccine scenario as calculated in Table 3.

Table 3. Participation probabilities for different vaccine scenarios separate for the words and graphics dataset.[§]

	Probabilities for words dataset	Probabilities for graphics dataset
Vaccine 1: Effectiveness 95%; 1 year protection; at CWC		
€0 + 1 in 10,000	69.0	43.9
€0 + 1 in 1,000,000	77.3	72.7
€140 + 1 in 10,000	38.9	16.0
€140 + 1 in 1,000,000	50.9	41.6
Vaccine 2: Effectiveness 75%; 1 year protection; at CWC		
€0 + 1 in 10,000	54.5	30.9
€0 + 1 in 1,000,000	64.8	61.2
€140 + 1 in 10,000	27.0	8.8
€140 + 1 in 1,000,000	38.5	30.5
Vaccine 3: Effectiveness 55%; 1 year protection; at CWC		
€0 + 1 in 10,000	40.8	22.4
€0 + 1 in 1,000,000	50.7	49.9
€140 + 1 in 10,000	16.0	6.0
€140 + 1 in 1,000,000	25.7	19.9
Vaccine 4: Effectiveness 95%; 6 years protection; at CWC		
€0 + 1 in 10,000	80.4	66.1
€0 + 1 in 1,000,000	86.8	87.8
€140 + 1 in 10,000	52.3	34.0
€140 + 1 in 1,000,000	65.3	66.9
Vaccine 5: Effectiveness 75%; 6 years protection; at CWC		
€0 + 1 in 10,000	66.8	54.0
€0 + 1 in 1,000,000	76.8	79.8
€140 + 1 in 10,000	38.7	24.3
€140 + 1 in 1,000,000	50.6	52.7
Vaccine 6: Effectiveness 55%; 6 years protection; at CWC		
€0 + 1 in 10,000	54.2	41.2
€0 + 1 in 1,000,000	63.3	70.6
€140 + 1 in 10,000	25.3	15.5
€140 + 1 in 1,000,000	37.4	40.5
Vaccine 7: Effectiveness 95%; 3 years protection; at CWC		
€0 + 1 in 1,000,000	89.3	89.3
Vaccine 8: Effectiveness 55%; 1 year protection; at GP		
€140 + 1 in 10,000	15.9	

[§] CWC is Child Welfare Center, GP is General practitioners' office, ICC 0.94, $p<0.001$ between probabilities calculated within the words and graphics dataset

DISCUSSION

This study shows that the way similar attribute levels are presented, either in graphics or in words, affects the results of the DCE. The most prominent and relevant result is that, due to the way of presenting attribute levels, the attribute level estimates and thereby the relative importance of the attribute levels is affected. Because of those differences, the estimated probabilities to participate in vaccination differed between the words and graphics datasets.

Now the important question is raised which results are more reliable, those elicited using words or using graphics. Respondents who were presented with choice tasks in words first showed significantly more consistent answer patterns compared to those who started with the graphics choice tasks. This could have been caused by the fact that the graphics were fairly unfamiliar to the respondents thereby requiring greater cognitive processing before they became meaningful and interpretable. However, in DCEs, especially on health-related topics, the graphics that will be used are always rather unfamiliar for respondents and respondents will be exposed to the graphics immediately at start of the questionnaire without any option to first practice the choice tasks in for instance words. Therefore, in this study consistency of choices was used as a proxy for direct attribute level understanding, and these results suggest that the use of words in choice tasks provides more accurate responses than the use of graphics.

Moreover, the differences in attribute estimates and relative importance between both datasets might be explained by how respondents interpreted the attribute levels when displayed in words or graphics. Although most respondents reported using the actual numerical values of the attribute levels when interpreting the graphical attribute levels, the results of this study seem to indicate otherwise. If the distance between the levels within an attribute differs, the part-worth utility difference between those levels should be higher for the two levels furthest apart [7-9]. This was only observed in the results of the questionnaire that used words to depict the attribute levels, and not in the graphics dataset. This suggests that respondents interpreted the attribute levels as ordinal utilities (e.g., low-medium-high or bad-average-good) when they were presented in graphics. This simplification in attribute level interpretation is unwanted, because a numerically small risk might be classified as the highest risk compared to the other levels within the same attribute. If broad classifications instead of actual values determine respondents' choice, these preferences might not reflect real life choices as accurately as intended. The possible negative effects of attribute level recoding may be further underlined by previous research stating that the use of icons may overestimate the importance of risk and willingness to pay values [12, 15, 18, 40]. The latter was also shown in the current study. Compared to the words dataset, within the graphics dataset, the participation probabilities decreased to a greater extent when the frequency of severe side effects increased. This has significant implications for the conclusions of a DCE and the estimated tradeoffs between decision criteria [7-9].

Independent of educational level and health literacy, respondents preferred choice tasks presented in words over choice tasks presented in graphics. Moreover, they were more consistent in their

answers when they started with the choice tasks in words compared to graphics. These findings are in contrast to expectations based on previous literature [12, 13, 15, 16, 18-22, 24, 41], which may have to do with the amount of graphics respondents had to process within the current study. Depicting one risk by means of a graphic might enhance respondents' understanding of that risk, however if respondents have to compare multiple graphics for risks and other characteristics, this might reduce the comparability between concepts and thereby make it more difficult to decide [42]. While in the current study either solely words or graphics were used for describing all attribute levels, most DCEs that use graphics only included them for certain attributes or as an add-on to a description in words. Although literature suggests that using both presentation option at once will not be beneficial for respondents [40], future research should be conducted to test whether respondents will benefit from or will be confused by using graphics as an add-on to (some) attributes described in words. Such research might then conclude whether the use of graphics should be further enhanced or avoided in DCEs.

This study is subject to some limitations. First, the response rate was 38.4%. Due to confidentiality agreements with Praeventis, no reminder letters could be sent and a non-response analysis could not be conducted. Non-response is likely to be selective. Compared to the general Dutch population and in agreement with other questionnaires and prior DCE's, the current sample was higher educated and included a lower number of non-Dutch individuals. Since education and ethnicity are expected to influence respondents' capacity to interpret the graphics in this DCE, and no gold standard is available, results might even underestimate the true effect of using either words or graphics. Analysing the demographics of the respondents that were excluded from the analyses due to missing data showed that, respondents with >10% missing answers on the choice tasks had a significant lower health literacy status and most often skipped the choice tasks that used graphics (results not shown). Finally, within this study only one graphic per attribute level could be included. Although the process of selecting and designing the graphics was very concise and thorough (i.e. based on literature, expert opinion, and pilot testing), we were not able to test whether the graphics used on this study are the best possible graphics for the attribute levels they depicted. Future research should be conducted to elaborate on sort and type of graphics that could be effective in certain DCEs with specific health related topics.

In conclusion, results differ when using words or graphics to present attribute levels in choice tasks. The use of words to depict attribute levels was preferred by the respondents of this study. Moreover, using words led to more consistent answer patterns, more accurate attribute level interpretation and more accurate attribute level estimates.

Currently, there are no clear guidelines on how to present a choice task, and this has not been discussed in the ISPOR tasks force on good research practices for conjoint analysis so far [43, 44]. This study shows a large impact of using two alternative forms of presentation on study outcomes. If the reason for using graphics is to improve respondents' understanding of the attribute levels, this study shows this is not the case by definition, also not among lower educated and less literate

respondents. Future research on this topic is recommended in order to establish guidelines on how to develop a valid presentation method for attribute levels in the choice tasks of a DCE. Until such guidelines exist, it is advised not to use only graphics to present the attribute levels in a choice task. Additionally, if researchers decide to use graphics, attribute level presentation should be incorporated in the focus group phase of the designing stage of a DCE and should be extensively pilot tested among the target population.

References

1. Johnson, F.R., S. Ozdemir, and K.A. Phillips, *Effects of simplifying choice tasks on estimates of taste heterogeneity in stated-choice surveys.* Soc Sci Med., 2010. 70(2): p. 183-90.
2. Bech, M., T. Kjaer, and J. Lauridsen, *Does the number of choice sets matter? Results from a web survey applying a discrete choice experiment.* Health Econ, 2011. 20(3): p. 273-86.
3. Maddala, T., K.A. Phillips, and F.R. Johnson, *An experiment on simplifying conjoint analysis designs for measuring preferences.* Health Econ., 2003. 12(12): p. 1035-47.
4. Kenny, P., et al., *Do participants understand a stated preference health survey? A qualitative approach to assessing validity.* Int J Technol Assess Health Care, 2003. 19(4): p. 664-81.
5. de Bekker-Grob, E.W., M. Ryan, and K. Gerard, *Discrete choice experiments in health economics: a review of the literature.* Health Econ., 2012. 21(2): p. 145-72.
6. Marshall, D., et al., *Conjoint Analysis Applications in Health - How are Studies being Designed and Reported?: An Update on Current Practice in the Published Literature between 2005 and 2008.* Patient, 2010. 3(4): p. 249-56.
7. Ryan, M., K. Gerard, and M. Amaya-Amaya, *Using Discrete Choice Experiments to Value Health and Health Care.* The Economics of Non-Market Goods and Resources, ed. I.J. Bateman. 2008, Dordrecht: Springer.
8. Louviere, J.J., D.A. Hensher, and J.D. Swait, *Stated Choice Methods; analysis and apllication.* 2000, Cambridge: Cambridge University Press.
9. Hensher, D.A., J.M. Rose, and W.H. Greene, *Applied Choice Analysis: A Primer.* 2005, New York: Cambridge University Press.
10. Ozdemir, S., et al., *Who pays attention in stated-choice surveys?* Health Econ., 2010. 19(1): p. 111-8.
11. Johnson, F.R., et al., *How does cost matter in health-care discrete-choice experiments?* Health Econ, 2011. 20(3): p. 323-30.
12. Hawley, S.T., et al., *The impact of the format of graphical presentation on health-related knowledge and treatment choices.* Patient Educ Couns., 2008. 73(3): p. 448-55.
13. McCaffery, K.J., et al., *The influence of graphic display format on the interpretations of quantitative risk information among adults with lower education and literacy: a randomized experimental study.* Med Decis Making., 2012. 32(4): p. 532-44.
14. Dolan, J.G. and S. Iadarola, *Risk communication formats for low probability events: an exploratory study of patient preferences.* BMC Med Inform Decis Mak., 2008. 8(14).
15. Hildon, Z., D. Allwood, and N. Black, *Impact of format and content of visual display of data on comprehension, choice and preference: a systematic review.* Int J Qual Health Care., 2012. 24(1): p. 55-64.
16. Hildon, Z., D. Allwood, and N. Black, *Making data more meaningful: patients' views of the format and content of quality indicators comparing health care providers.* Patient Educ Couns., 2012. 88(2): p. 298-304.
17. Galesic, M. and R. Garcia-Retamero, *Statistical numeracy for health: a cross-cultural comparison with probabilistic national samples.* Arch Intern Med., 2010. 170(5): p. 462-8.
18. Schapira, M.M., A.B. Nattinger, and T.L. McAuliffe, *The influence of graphic format on breast cancer risk communication.* J Health Commun., 2006. 11(6): p. 569-82.
19. Lipkus, I.M., *Numeric, verbal, and visual formats of conveying health risks: suggested best practices and future recommendations.* Med Decis Making., 2007. 27(5): p. 696-713.
20. Lipkus, I.M. and J.G. Hollands, *The visual communication of risk.* J Natl Cancer Inst Monogr, 1999(25): p. 149-63.
21. Chuang, M.H., et al., *Development of pictographs depicting medication use instructions for low-literacy medical clinic ambulatory patients.* J Manag Care Pharm., 2010. 16(5): p. 337-45.
22. Shrank, W., et al., *Effect of content and format of prescription drug labels on readability, understanding, and medication use: a systematic review.* Ann Pharmacother., 2007. 41(5): p. 783-801.
23. Galesic, M., R. Garcia-Retamero, and G. Gigerenzer, *Using icon arrays to communicate medical risks: overcoming low numeracy.* Health Psychol., 2009. 28(2): p. 210-6.
24. Waters, E.A., et al., *Formats for improving risk communication in medical tradeoff decisions.* J Health Commun., 2006. 11(2): p. 167-82.

25. HLS-EU Consortium, *Comparative report of health literacy in eight EU member states. The European health literacy survey HLS-EU*. Aailable from: HTTP://WWW.HEALTH-LITERACY.EU, 2013.
26. Rudd, R.E., *Health literacy skills of U.S. adults.* Am J Health Behav., 2007. 31(Suppl 1): p. S8-18.
27. de Bekker-Grob, E.W., et al., *Girls' preferences for HPV vaccination: a discrete choice experiment*. Vaccine, 2010. 28(41): p. 6692-7.
28. Hall, J., et al., *Using stated preference discrete choice modelling to evaluate the introduction of varicella vaccination.* Health Econ, 2002. 11(5): p. 457-65.
29. Brown, D.S., et al., *Mothers' preferences and willingness to pay for vaccinating daughters against human papillomavirus.* Vaccine., 2010. 28(7): p. 1702-8.
30. Poulos, C., et al., *Mothers' preferences and willingness to pay for HPV vaccines in Vinh Long Province, Vietnam.* Soc Sci Med., 2011. 73(2): p. 226-34.
31. Flood, E.M., et al., *Parent preferences for pediatric influenza vaccine attributes.* Clin Pediatr (Phila). 2011. 50(4): p. 338-47.
32. Flood, E.M., et al., *A survey of children's preferences for influenza vaccine attributes.* Vaccine., 2011. 29(26): p. 4334-40.
33. Stockwell, M.S., et al., *The effects of vaccine characteristics on adult women's attitudes about vaccination: a conjoint analysis study.* Vaccine., 2011. 29(27): p. 4507-11.
34. Bishai, D., et al., *Conjoint analysis of French and German parents' willingness to pay for meningococcal vaccine.* PharmacoEconomics, 2007. 25(2): p. 143-54.
35. Bliemer, M.C.J., J.M. Rose, and D.A. Hensher, *Efficient stated choice experiments for estimating nested logit models.* Transport Res, 2009. 43: p. 19-35.
36. Chew, L.D., K.A. Bradley, and E.J. Boyko, *Brief questions to identify patients with inadequate health literacy.* Fam Med., 2004. 36(8): p. 588-94.
37. Fransen, M.P., et al., *Applicability of internationally available health literacy measures in the Netherlands.* J Health Commun, 2011. 16(Suppl 3): p. 134-49.
38. Fiebig, D.G., et al., *The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity.* Market Sci, 2010. 29(3): p. 393-421.
39. Bech, M. and D. Gyrd-Hansen, *Effects coding in discrete choice experiments.* Health Econ, 2005. 14(10): p. 1079-83.
40. Feldman-Stewart, D., M.D. Brundage, and V. Zotov, *Further insight into the perception of quantitative information: judgements of gist in treatment decisions.* Med Decis Making, 2007. 27(1): p. 34-43.
41. Timmermans, D.R.M., C.F. Ockhuysen-Vermey, and L. Henneman, *Presenting health risk information in different formats: The effect on participants' cognitive and emotional evaluation and decisions.* Patient Educ Couns, 2008. 73: p. 443-447.
42. Schirillo, J.A. and E.R. Stone, *The greater ability of graphical versus numerical displays to increase risk avoidance involves a common mechanism.* Risk Analysis, 2005. 25(3): p. 555-566.
43. Bridges, J.F., et al., *Conjoint analysis applications in health--a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force.* Value Health., 2011. 14(4): p. 403-13.
44. Johnson, F.R., et al., *Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force.* Value Health., 2013. 16(1): p. 3-13.
45. Murphy, T.V., et al., *Intussusception among infants given an oral rotavirus vaccine.* N Engl J Med., 2001. 344(8): p. 564-72.

CHAPTER 9

SURVIVAL OR MORTALITY: DOES RISK ATTRIBUTE FRAMING INFLUENCE DECISION-MAKING BEHAVIOR IN A DISCRETE CHOICE EXPERIMENT?

J. Veldwijk, B.A.B. Essers, M.S. Lambooij,
C.D. Dirksen, H.A. Smit, G.A. de Wit

Submitted.

ABSTRACT

Objective: To test how attribute framing in a Discrete Choice Experiment (DCE) affects respondents' decision-making behavior and their preferences.

Methods: Two versions of a DCE questionnaire containing nine choice tasks were distributed among a representative sample of the Dutch population age 55-65. The DCE consisted of four attributes related to the decision about participation in genetic screening for colorectal cancer (CRC). The risk attribute included was framed positively as the probability of surviving CRC and negatively as the probability of dying from CRC. Panel mixed logit models were conducted to estimate the relative importance of the attributes.

Results: The majority (56%) of the respondents ranked survival as the most important attribute in the positively framed DCE, while only a minority (8%) of respondents ranked mortality as the most important attribute in the negatively framed DCE. Respondents made dominant choices based on survival significantly more often than based on mortality. The framing of the risk attribute significantly influenced all attribute level estimates and resulted in different preference structures among respondents in the positively and negatively framed dataset.

Conclusions: Risk framing affects how respondents value the presented risk. Positive risk framing led to increased dominant decision-making behavior while negative risk framing led to risk seeking behavior. Attribute framing should have a prominent part in the expert and focus-group interviews and different types of framing should be used in the pilot version of DCEs as well as in actual DCEs to estimate the magnitude of the effect of choosing different types of framing.

INTRODUCTION

When making health-related decisions, individuals have to weigh several benefits and risks. The trade-offs that people make between those benefits and risks can be measured by means of a Discrete Choice Experiment (DCE). The use of risk information in those DCEs is almost inevitable and assumed customary [1, 2]. However, accurately communicating risk information can be a challenge. Risks are often perceived as difficult to interpret, especially among certain subgroups of the population (i.e. individuals with a lower educational level or lower health literacy skills [3-5]). If not presented clearly, risk information might be misinterpreted, which limits the validity of the study outcomes. The framing that is used to present risk information is one of the many aspects of communicating risk information that influences how respondents interpret, perceive and value risks [1, 6, 7]. From the early experiments of Tversky & Kahneman [6, 7] and others [8-12], it is known that presenting otherwise identical risk information either positively or negatively will influence peoples' decision-making behavior. When risks are framed positively they are more often interpreted as a 'gain', resulting in risk averse behavior (i.e. choosing the safest option), while risk-seeking behavior is more common when risk information is framed negatively and interpreted as a 'loss'. Illustrative to this principle is Tversky & Kahneman's study where participants were faced with a scenario involving the outbreak of a disease that threatened to kill 600 people [7]. The majority (72%) of the participants preferred the 'safe' option that '200 people will be saved' over the 'risky' option that 'there is 1/3 probability that 600 people will be saved and 2/3 probability that no people will be saved'. When the risk information was framed negatively (but otherwise equivalent), the majority (78%) preferred the 'risky' option over the 'safe' option [7].

Therefore, it is expected that respondents value certain risk attributes in a DCE differently if they are framed positively or negatively. Specifically within a DCE, framing effects are of importance, since framing a risk does not only affect decision-making behavior with respect to the risk attribute at hand. Due to the multi-attribute approach of DCE studies [13, 14], framing a risk attribute might influence the valuation of all included attributes. Limited research outside health economics indicated that attribute framing within DCEs might affect decision behavior [15, 16]. Research on possible framing effects is even more scarce within health-related DCEs [17]. At the same time, the use of DCEs in the public health and health care research setting has increased [18, 19], and results are increasingly being used as input for policymaking [13, 14, 20]. Therefore, the accuracy and validity of the measured (i.e. stated) preferences are essential. Currently, both positive and negative framing is used to communicate risk attributes without exact knowledge of the effect of attribute framing on an individual's decision-making behavior or evidence of best practice. This study empirically tests whether framing a risk attribute either positively or negatively influences respondents' decision-making behavior and their preferences.

MATERIALS AND METHODS

DCE case study and participant recruitment

A DCE on preferences for genetic screening for colorectal cancer (CRC) was used as a case for this study. The decision to participate in cancer screening typically involves mastering concepts of risk. CRC is one of the most commonly diagnosed and the leading cause of death among all cancer types in developed countries [21, 22]. Several countries (among which the Netherlands) have implemented a population-based CRC screening program. However, within this population screening program there is no specific attention for genetically predisposed individuals, while about 5% of all diagnosed CRCs has a genetic origin [23, 24]. It has been argued that genetic screening will provide options to optimize surveillance of high-risk individuals, and consequently will reduce CRC related morbidity and mortality [25-28]. Therefore, individual preferences of the target population are important to consider within this context.

Respondents were recruited via an existing online panel of the general Dutch population. Respondents were selected based on their age (55-65 years) and were representative for the entire target population with respect to gender and educational level. In total, 11,000 individuals were invited to participate in this study and recruitment continued until at least 1000 questionnaires were fully completed. Of those initially invited, 1595 (14.5%) started the questionnaire within four weeks. Complete data was gathered for 1262 respondents (79.1%), after which data-collection was closed. Respondents were excluded from the dataset if they completed the questionnaire at an unlikely fast pace (within 10 minutes) (14%), or when they had already participated in the national CRC screening program (3.2%). The Dutch Central Committee on Research involving Human Subjects concluded that formal approval by an Institutional Review Board was not required, as respondents were only required to complete an anonymous questionnaire once, which is in accordance with the guidelines laid down in the Declaration of Helsinki.

Attributes, levels and framing

To construct the current DCE, potential attributes and levels were identified from previously published studies [29-33], six expert interviews, and five group interviews with members from the target population (n=38). These group interviews were conducted using the Nominal Group Technique (NGT) [34]. During these interviews, participants were asked to rank a number of potential attributes from most to least important, the mean group-ranking of the attributes was then discussed in the group, after which participants could change their primary individual ranking [34]. Following a consensus meeting with the authors, four attributes with each three levels were rated as most important and were selected for the DCE (Table 1). For the purpose of the current study, the survival attribute that describes the probability of surviving CRC (positive framing) was correspondingly framed in terms of mortality: the probability of dying from CRC (negative framing) (Table 1).

Table 1. Attributes and levels that were included in this DCE*

Attributes	Level 1	Level 2	Level 3
Probability of being genetically predisposed (genetic predisposition) The likelihood that you are genetically predisposed to develop colorectal cancer.	1% 1 out of every 100	3% 3 out of every 100	15% 15 out of every 100
Probability of developing CRC* (CRC risk) 5 out of every 100 (5%) Dutch individuals develop colorectal cancer. If you have a genetic predisposition to develop colorectal cancer and you do not participate in preventive colonoscopies, the likelihood that you will develop colorectal cancer is higher and varies between:	15% 15 out of every 100	70% 70 out of every 100	99% 99 out of every 100
Frequency of preventive colonoscopies (colonoscopy frequency) If the genetic test shows that you are genetically predisposed to develop colorectal cancer, you will be invited to participate in preventive colonoscopies. These colonoscopies are performed to prevent cancer from developing or to diagnose cancer in an early stage. These colonoscopies will be scheduled on a regular basis varying between:	Every year	Every 2 years	Every 5 years
Probability of surviving CRC (survival) 60 out of every 100 (60%) Dutch individuals with colorectal cancer survive over the next 5 years. If you know you are genetically predisposed to develop colorectal cancer and if you participate in the preventive colonoscopies the likelihood that you will survive colorectal cancer over the next 5 years will increase and varies between:	80% 80 out of every 100	92% 92 out of every 100	98% 98 out of every 100
Probability of dying from CRC (mortality) 40 out of every 100 (40%) Dutch individuals die from colorectal cancer within the next 5 years. If you know you are genetically predisposed to develop colorectal cancer and if you participate in the preventive colonoscopies the probability that you will die from colorectal cancer within the next 5 years will decrease and varies between:	20% 20 out of every 100	8% 8 out of every 100	2% 2 out of every 100

* All choice tasks included the first three attributes, for half of the population survival was added as a fourth attribute, while for the other half of the population mortality was added as a fourth attribute.

Experimental design, pilot testing and questionnaire

NGene 1.0 (ChoiceMetrics, 2011) software was used to develop a D-efficient design [35, 36]. The DCE consisted of nine unique choice tasks. Half of the population answered these choice tasks in the positive framing (i.e. survival was included as a risk attribute), while the other half of the population answered the choice tasks in a negative framing (i.e. mortality was included as a risk attribute). Respondents were randomly assigned to either the positively framed version or the negatively framed version of the questionnaire.

Before participants were asked to complete the choice tasks, they received detailed information

on the meaning of all attributes and levels as well as an explanation on how to complete a choice task, illustrated by an example. Every choice task started with the question: 'Imagine that you have participated in the national CRC Screening Program. Blood was found in your stool and you subsequently went to the hospital for a colonoscopy. Based on the colonoscopy and your family history regarding cancer, your physician advises you to participate in a genetic screening test to determine whether you have a familial and hereditary predisposition to develop CRC. In which scenario would you prefer to participate in a genetic screening, scenario 1 or scenario 2?' First, respondents were forced to choose between scenario 1 and scenario 2. Following each choice task, participants were asked whether they would actually participate in the chosen scenario or not (i.e. opt-out).

The draft questionnaire was pilot tested among members of the target population (n=90). Four of these pilot tests were 'think aloud' tests, during which a researcher was present when the participant completed the questionnaire, reading out loud. It was tested by means of this pilot whether correct wording was used and whether the target population understood the attributes, levels and choice tasks. Additionally, the attribute level estimates that were retrieved from the pilot study served as input for the design of the final DCE questionnaire.

The final questionnaire consisted of three parts. The first section of the questionnaire comprised 25 questions on demographics such as gender, age, educational level, health literacy [37] and ethnicity. Educational level was dichotomized into higher (i.e. tertiary education) or lower education (i.e. all other educational levels). Health literacy was measured by three validated Dutch questions of the Set of Brief Screening Questions (SBSQ-D) [37] and dichotomized into inadequate health literacy (an average score of ≤2) and adequate health literacy (an average score >2) [37]. The second part of the questionnaire consisted of the actual DCE as explained above (1 example choice task and 9 choice tasks to answer). The third part consisted of 14 questions regarding the consequences of genetic screening.

Statistical analysis

Direct attribute rating

Direct attribute ratings were obtained from respondents for both the positively and negatively framed DCE. Chi-square tests were used to test whether respondents in the positively framed dataset rated attributes significantly different from respondents in the negatively framed dataset.

The role of the risk attribute in decision-making

The proportion of respondents who always chose the scenario with the highest survival / the lowest mortality was calculated for each dataset. Chi-square tests were conducted to test whether this differed significantly between both datasets.

Differences by educational level and health literacy

Since respondents with a lower educational level or inadequate health literacy experience

difficulties understanding risk information [3-5], they might also be more prone to framing effects. Therefore, the differences in direct attribute ranking and the role of the risk attribute in decision-making between respondents who answered the positively and negatively framed DCE were tested among respondents with a higher educational level, a lower educational level, adequate health literacy and inadequate health literacy.

Attribute level estimates and relative importance

To adjust for the multilevel structure of the data and to correct for preference heterogeneity, panel mixed-logit models (MIXL) were constructed in Nlogit 5.0 (econometric software). All attributes were recoded using effects coding to account for potential non-linearity. This coding procedure codes the reference category as -1, so the sum of the effect coded attribute levels is always zero [13, 38]. Based on model fit tests (AIC, BIC, Log likelihood), it was tested which model was the most suitable for our data. Based on significance of the standard deviation estimates, it was determined which attributes should be included as random parameters due to significant preference heterogeneity. Similarly, it was tested which distribution (e.g., normal, lognormal, uniform) should be assumed for those parameters. Based on these analyses, CRC risk, colonoscopy frequency and survival/mortality were included as random parameters, all with normal distributions.

The systematic utility component (V) describes the measurable part of the utility of a specific genetic screening test based on the attributes that were included in the DCE, which was tested using the following equation:

$$V = \beta_0 + \beta_1 * \text{genetic predisposition}_{3\%} + \beta_2 * \text{genetic predisposition}_{15\%} + \beta_{3i} * \text{CRC risk}_{70\%} + \beta_{4i} * \text{CRC risk}_{99\%} + \beta_{5i} * \text{colonoscopy frequency}_{2\ \text{years}} + \beta_{6i} * \text{colonoscopy frequency}_{5\ \text{years}} + \beta_{7i} * \text{survival}_{92\%}/\text{mortality}_{8\%} + \beta_{8i} * \text{survival}_{98\%}/\text{mortality}_{2\%}$$

$\beta_1 - \beta_8$ are the attribute level estimates that indicate the relative importance of each attribute level. The sign of the estimate indicates whether the attribute level has a positive or a negative influence on the utility. β_0 was included as a constant term. Since the choice tasks were unlabeled and only the forced choices were analyzed, the constant term cannot be interpreted. It identifies respondents' preferences for scenario 1 over scenario 2 irrespective of the attribute levels, which in itself has no meaning. It merely reflects the optional presence of left-right bias (i.e. the tendency of respondents to opt for the left option).

Using the above equation, two main modelling steps were undertaken. First, based on the complete dataset, it was tested whether the frame of the risk attribute significantly influenced the attribute level estimates. For that purpose, a variable that identified the use of either positive or negative framing was included in the model as a covariate and as an interaction term with all attribute levels. Second, the attribute estimates were estimated separately for the positively framed data and the negatively framed data.

Importance weights were calculated based on the results of the separate MIXL models. The difference between the smallest and largest attribute level estimate was calculated for each attribute. The largest difference value received an importance score of one, the other difference values were divided by the largest difference value, resulting in a relative distance between all attributes to the most important attribute.

The panel-MIXL model does not account for variability in individual errors (scale heterogeneity) [39], while respondents might react differently (i.e. more or less consistent) to a positively or negatively framed risk. Therefore, the modelling procedures described above were repeated using a Heteroscedastic Extreme Value (HEV) model that accounts for scale heterogeneity, and a Generalized mixed-logit (G-MIXL) model that accounts both for scale and preference heterogeneity [39].

RESULTS

Participants' characteristics

In total, 1045 respondents were included in the analysis of which 532 (50.9%) respondents completed the positively framed DCE and 513 (49.1%) completed the negatively framed DCE. Demographic variables are shown in Table 2 and did not significantly differ between the respondents who answered either the positively or negatively framed DCE.

Table 2. Demographic characteristics of the study population

		Total dataset (n=1045)		Positive framing (n=532)		Negative framing (n=513)	
		Mean (SD)	Percentage	Mean (SD)	Percentage	Mean (SD)	Percentage
Age		59.7 (3.1)		59.5 (3.1)		59.9 (3.1)	
Gender	Female		49.7		50.9		48.3
Highest attained educational level	Low		26.1		26.3		25.9
	Medium		36.6		37.1		36.1
	High		37.3		36.6		38.0
Health literacy*	Adequate		96.7		96.6		96.7
Ethnicity	Dutch		97.6		96.6		98.6

***** Health literacy was measured using three items; mean scores range between 0-4, a mean score higher than 2 indicates adequate health literacy.

Direct attribute rating

Respondents who answered the positively framed DCE rated the attributes significantly different from the respondents who answered the negatively framed DCE (Chi-square=277.5, P< 0.01). Based on the positively framed data set, most of the respondents (55.6%) rated survival as most important (Figure 1), followed by genetic predisposition (19.4%). Based on the negatively framed data set, genetic predisposition was reported as most important by the majority of the respondents (43.6%). Only 8% of the respondents rated the probability of dying from CRC as most important (Figure 1).

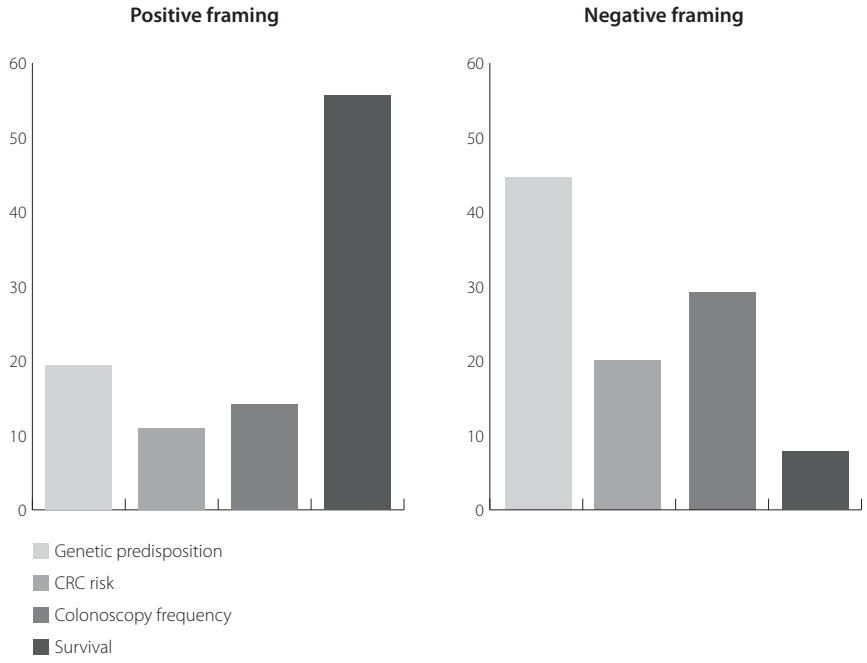

Figure 1. Direct attribute ranking based on the positive (top graph) and negative (bottom graph) framing.

The role of the risk attribute in decision-making

Based on the positive framing, a significantly higher proportion of respondents always chose the scenario with the highest level of the survival attribute (9.2%) compared to the proportion of respondents who always chose the scenario with the lowest level of the mortality attribute in the negatively framed DCE (4.9%) (Chi-square=7.47, P<0.01).

Differences by educational level and health literacy

Differences in direct attribute ranking based on the positively framed and negatively framed DCE were significant among respondents with a higher educational level (Chi-square=109.8, P<0.01), a lower educational level (Chi-square=168.1, P<0.01), adequate health literacy (Chi-square=266.7, P<0.01) and inadequate health literacy (Chi-square=13.1, P<0.01).

The proportion of respondents who always chose the scenario with the highest survival rate was significantly higher compared to the proportion of respondents who always chose the scenario with the lowest mortality rate. This dominant decision-making behavior was shown among respondents with a lower education level (Chi-square=2.98, P<0.10), respondents with a higher educational level (Chi-square=4.89, P<0.05), and respondents with adequate health literacy (Chi-square=6.96, P<0.01). This final analysis could not be conducted among respondents with inadequate health literacy due to the small size of this subgroup (n=35)

Attribute estimates, importance weights and relative importance

The framing of the risk attribute (survival versus mortality) had a significant impact on all attributes (i.e. significant interactions with all attributes) in the total dataset (Table 3).

Although respondents were initially forced to select their preferred screening option in all nine choice tasks, they were offered an opt-out option after answering each choice task. Of all respondents who completed the positively framed DCE, 3.6% chose to opt out in all nine choice tasks, while 88.9% never chose to opt out. Of all respondents who completed the negatively framed DCE, 5.8% always opted out and 86.6% never opted out. The mean number of times respondents chose to opt out differed between both datasets (t=-1.95, P=0.05).

Analyzing the data of the positive and negative framing of the risk attributes separately showed equal directions of the estimates for most attribute levels. Within both datasets, significant preference heterogeneity was shown for three out of the four attributes. Based on both the positive framing and the negative framing of the risk attributes, respondents preferred a survival of 92% and 98% over a survival of 80% and a mortality of 8% and 2% over a mortality of 20%. However, the preferences for the other attributes were different in both datasets. For instance, based on the positive framing, respondents preferred to undergo a colonoscopy every 2 years over having a colonoscopy every 5 years, while respondents in the negative framing preferred a yearly colonoscopy over all other levels. These differences in preferences were also reflected in the importance scores and thereby the relative importance of the attributes (Figure 2). Based on the positive framing, survival was most important, while colonoscopy frequency was most important based on the negative framing. Moreover, the relative distance of the attributes to the most important attribute differed between the datasets (Figure 2).

The results of the HEV model and the G-MIXL model showed no evidence for scale heterogeneity. The directions of the attribute level estimates as well as the relative importance of the attributes are similar to the results presented for the panel-MIXL model.

DISCUSSION

This study shows that framing a risk attribute either positively in terms of survival or negatively in terms of mortality affects how respondents perceive and value this attribute, ultimately affecting the results of the DCE. Based on the direct attribute ranking results, the majority of the respondents (56%) who answered the positively framed DCE rated survival as most important, while the minority of the respondents (8%) rated mortality as most important in the negatively framed DCE. This finding was further emphasized by evaluating the number of dominant choices; in the positively framed DCE a significantly higher proportion of respondents always chose the scenario with the highest survival compared to the proportion of respondents who always chose the lowest mortality in the negatively framed DCE. This indicates that respondents in a positively framed DCE might be less inclined to make a trade-off between survival and the other attributes compared to respondents in the negatively framed DCE. Framing risks negatively probably resulted in a shift of respondents'

Table 3. Preferences for genetic testing for colorectal cancer based on the MIXL model for the total data set as well as stratified by frame^

			Total data set		Positive framing^^		Negative framing^^	
			Estimate	**SE**	**Estimate**	**SE**	**Estimate**	**SE**
Constant			0.33***	0.04	0.20***	0.04	0.32***	0.04
Genetic predisposition	1% (ref)		-0.06	0.04	-0.07	0.03	-0.06*	0.03
	3%		-0.18***	0.04	-0.27**	0.04	-0.18***	0.04
	15%		0.24***	0.03	0.34***	0.04	0.24***	0.04
CRC risk	15% (ref)	Mean	0.07	0.04	-0.10***	0.04	0.08*	0.04
		SD	0.46***	0.03	0.42***	0.05	0.50***	0.05
	70%	Mean	0.05	0.03	0.15***	0.04	0.05	0.03
		SD	0.00	0.07	0.12	0.07	0.05	0.08
	99%	Mean	-0.12***	0.04	-0.05	0.04	-0.13***	0.04
		SD	0.46***	0.03	0.41***	0.05	0.50***	0.05
Colonoscopy frequency	Every year (ref)	Mean	0.39***	0.06	0.25***	0.07	0.37***	0.06
		SD	1.16***	0.05	1.24***	0.07	1.11***	0.06
	Every 2 years	Mean	0.22***	0.03	0.39***	0.03	0.23***	0.03
		SD	0.00	0.06	0.07	0.16	0.02	0.07
	Every 5 years	Mean	-0.61***	0.07	-0.64***	0.07	-0.60***	0.06
		SD	1.16***	0.05	1.24***	0.07	1.11***	0.06
Survival / mortality	80% / 20% (ref)	Mean	-0.41***	0.05	-0.73***	0.05	-0.41***	0.05
		SD	0.44***	0.04	0.46***	0.05	0.42***	0.05
	92% / 8%	Mean	0.14***	0.03	0.03	0.03	0.14***	0.03
		SD	0.02	0.04	0.01	0.05	0.02	0.06
	98% / 2%	Mean	0.27***	0.05	0.70***	0.05	0.27***	0.05
		SD	0.44***	0.04	0.46***	0.05	0.42***	0.05
Framing effect			-0.13**	0.05	-	-	-	-
Framing * genetic predisposition		3%	-0.58***	0.17	-	-	-	-
		15%	0.24***	0.08	-	-	-	-
Framing * CRC risk		70%	0.47***	0.13	-	-	-	-
		99%	-0.17	0.14	-	-	-	-
Framing * colonoscopy frequency		Every 2 years	0.37***	0.08	-	-	-	-
		Every 5 years	-0.09	0.12	-	-	-	-
Framing * survival / mortality		92% / 8%	0.29**	0.14	-	-	-	-
		98% / 2%	0.45***	0.11	-	-	-	-

^ Standard Deviation, only presented for the random parameters. ^^Model fit measures: positive frame LL -2785.6; Pseudo-R^2 0.16; AIC 5599.3, negative frame LL -2825.8; Pseudo-R^2 0.12; AIC 5679.6.

* P<0.10; ** p <0.05; ***p <0.001

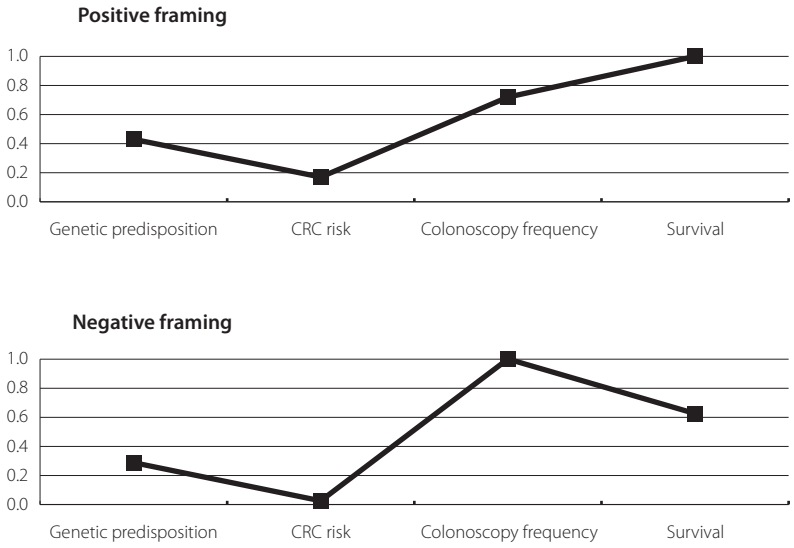

Figure 2. Importance weights and relative importance of the attributes based on the positive (top graph) and negative (bottom graph) data.

attention from mortality to other attributes. Indeed the regression results showed that, based on the positive framing, respondent valued survival relatively as most important, while respondents valued colonoscopy frequency relatively as most important based on the negative framing.

These results are as expected, respondents attach significantly higher value on survival (positive framing) than on mortality (negative framing) and are more inclined to focus on survival compared to mortality [7, 17]. In the negatively framed DCE, respondents often showed risk seeking behavior since they opted-out from genetic screening significantly more often compared to respondents in the positively framed DCE. Individuals valued equivalent risk information differently if framed positively or negatively, which influenced the valuation of all attributes that were included in the DCE. For instance, respondents who answered the negatively framed DCE were willing to accept a higher frequency of colonoscopies compared to respondents in the positively framed DCE. In summary, the current study shows that if this DCE was conducted using only a positive or a negative framing to present the included risk attribute, different conclusions would have been drawn with respect to respondents' decision-making behavior and preference structures.

All these results were found to be independent of educational level or health literacy score. This adds to the evidence that framing effects have little to do with understanding risk information, but that framing effects are 'likely to persist even among careful thinkers, and emerge from sincere attitudes towards either survival or mortality, and careless processing of probabilistic risk information' [40, 41].

All the aspects mentioned above call for greater attention and more research on the impact of risk attribute framing in DCEs that aim to elicit preferences within the health care and public health context. Research outside health economics suggests that framing risks both positively and negatively at once (you have a 80% probability to survive CRC so you have a 20% probability to die from CRC) in the same choice tasks might be a valid solution to overcome framing effects in a DCE [17, 42]. Whether this strategy helps to overcome framing effects in DCEs that study health-related decision-making behavior should be further investigated. Alongside such research, it should be debated whether we, as scientists, should strive to overcome framing effects when measuring stated preferences, or whether it is more informative to know the magnitude of the effect of attribute framing on decision-making behavior and the outcomes of DCEs. To date, large differences are reported in the communication of risks in practice [4, 5, 43]. Clinicians and health care and public health institutes use different methods and different frames to communicate risks. Since DCEs should mimic real life decision contexts as much as possible, it has to be discussed how to anticipate for framing differences in practice in order to develop a valid DCE. It might be worthwhile to conduct both a positively and a negatively framed pilot version or actual DCE more often to study three separate issues in further depth. First, the effect of framing on decision-making behavior and preferences might differ for different risk attributes [44, 45]. Individuals might react stronger to survival versus mortality compared to, for instance, effectiveness versus ineffectiveness or correctly diagnosed versus incorrectly diagnosed. Further research is needed to explore the differences in framing effects between different attributes. Second, individuals might react stronger to framing effects in individual health related decision-making compared to decisions within a public health setting [44, 45]. Decisions about treatment options are probably more emotional, personal and complex compared to decisions about participating in preventive programs such as vaccination. Additional research should be conducted to determine if framing effects differ based on the choice context. Third, DCEs might also be used as an effective tool to illustrate the effects of framing risks to clinicians in health care and public health institutes. Increased knowledge about framing effects might help them to develop effective communication strategies, which in turn might assist individuals in informed decision-making.

This study is subject to some limitations. First, given the context of this study (CRC screening), the current sample was relatively older (mean age of 59 years) compared to the general Dutch population. Since age might affect how respondents react to different frames, current study results cannot be generalized to the younger population. Especially with respect to survival and mortality, more pronounced framing effects are expected among younger individuals [46, 47]. Second, in agreement with other questionnaires and previous DCEs, the sample included a lower number of non-Dutch individuals than present in the general population. Language difficulties or cultural differences might influence how individuals with a different ethnicity interpret and value risks if they are framed either as survival or mortality [48]. Third, although relative measures were used to compare the outcomes of the positively and negatively framed DCE and the HEV and G-MIXL models provided no evidence for the presence of scale effects, attribute specific scale effects cannot be ruled out completely and might explain (part of) our findings.

In conclusion, risk framing affects how respondents value the presented risk thereby affecting the outcomes of the DCE study. Positively framed risks were significantly more important in the decision-making process than negatively framed risks. Moreover, positive risk framing led to increased dominant decision-making behavior while negative risk framing led to risk seeking behavior. Currently, there are no clear guidelines on how to present risks in a DCE and how decisions about risk framing impact study outcomes [49, 50]. Therefore, future research on this topic is recommended in order to establish guidelines on how to overcome or otherwise deal with framing effects in DCEs. Until such guidelines are developed, risks should be framed carefully. Attribute framing should have a prominent part in the expert interviews and focus group interviews that are conducted to select the attributes and attribute levels of a DCE in order to adapt them to current practice as much as possible. Preferably, different frames should be used in a pilot version of the DCE as well as in the actual DCE to enable researchers to estimate the magnitude of the effect of choosing different frames.

References

1. Harrison, M., et al., *Risk as an attribute in discrete choice experiments: a systematic review of the literature.* Patient, 2014. 7(2): p. 151-70.
2. Clark, M.D., et al., *Discrete Choice Experiments in Health Economics: A Review of the Literature.* Pharmacoeconomics, 2014. 9: p. 9.
3. Galesic, M. and R. Garcia-Retamero, *Statistical numeracy for health: a cross-cultural comparison with probabilistic national samples.* Arch Intern Med., 2010. 170(5): p. 462-8.
4. McCaffery, K.J., et al., *The influence of graphic display format on the interpretations of quantitative risk information among adults with lower education and literacy: a randomized experimental study.* Med Decis Making., 2012. 32(4): p. 532-44.
5. Waters, E.A., et al., *Formats for improving risk communication in medical tradeoff decisions.* J Health Commun., 2006. 11(2): p. 167-82.
6. Tversky, A. and D. Kahneman, *Judgment under uncertainty: Heuristics and biases.* Science, 1974. 185(4157): p. 1124-1131.
7. Tversky, A. and D. Kahneman, *The framing of decisions and the psychology of choice.* Science, 1981. 211(4481): p. 453-8.
8. Levin, I.P. and G.J. Gaerth, *How consumers are affected bythe frming of attribute information before an after consumer the product.* J Consum Res, 1988. 15(3): p. 374-378.
9. Peng, J., et al., *Framing effects in medical situations: distinctions of attribute, goal and risky choice frames.* J Int Med Res., 2013. 41(3): p. 771-6.
10. Just, D.R. and B. Wansink, *One man's tall is another man's small: how the framing of portion size influences food choice.* Health Econ., 2014. 23(7): p. 776-91.
11. Gong, J., et al., *The framing effect in medical decision-making: a review of the literature.* Psychol Health Med, 2013. 18(6): p. 645-53.
12. Edwards, A., et al., *Presenting risk information--a review of the effects of "framing" and other manipulations on patient outcomes.* J Health Commun., 2001. 6(1): p. 61-82.
13. Hensher, D.A., J.M. Rose, and W.H. Greene, *Applied Choice Analysis: A Primer.* 2005, New York: Cambridge University Press.
14. Louviere, J.J., D.A. Hensher, and J.D. Swait, *Stated Choice Methods; analysis and apllication.* 2000, Cambridge: Cambridge University Press.
15. Kragt, M.E. and J.W. Bennett, *Attribute framing in choice experiments: how do attribute level descriptions affect value estimates?* Environ Resour Econ, 2012. 51: p. 43-59.
16. Rolfe, J., J. Bennett, and J. Louviere, *Stated values and reminders of substitude goods: testing for framing effects with choice modelling.* Aust J Arg Resour Ec, 2002. 46(1-20).
17. Howard, K. and G. Salkeld, *Does attribute framing in discrete choice experiments influence willingness to pay? Results from a discrete choice experiment in screening for colorectal cancer.* Value Health, 2009. 12(2): p. 354-63.
18. de Bekker-Grob, E.W., M. Ryan, and K. Gerard, *Discrete choice experiments in health economics: a review of the literature.* Health Econ., 2012. 21(2): p. 145-72.
19. Marshall, D., et al., *Conjoint Analysis Applications in Health - How are Studies being Designed and Reported?: An Update on Current Practice in the Published Literature between 2005 and 2008.* Patient, 2010. 3(4): p. 249-56.
20. Ryan, M., K. Gerard, and M. Amaya-Amaya, *Using Discrete Choice Experiments to Value Health and Health Care.* The Economics of Non-Market Goods and Resources, ed. I.J. Bateman. 2008, Dordrecht: Springer.
21. Ferlay, J., et al., *Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012.* Int J Cancer, 2014. 13(10): p. 29210.
22. Ferlay, J., et al., *Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012.* Eur J Cancer, 2013. 49(6): p. 1374-403.
23. Bogaert, J. and H. Prenen, *Molecular genetics of colorectal cancer.* Ann Gastroenterol, 2014. 27(1): p. 9-14.
24. Haydon, A.M. and J.R. Jass, *Emerging pathways in colorectal-cancer development.* Lancet Oncol, 2002. 3(2): p. 83-8.
25. Burn, J., et al., *Long-term effect of aspirin on cancer risk in carriers of hereditary colorectal cancer: an analysis from the CAPP2 randomised controlled trial.* Lancet, 2011. 378(9809): p. 2081-7.

26. Burn, J., J. Mathers, and D.T. Bishop, *Genetics, inheritance and strategies for prevention in populations at high risk of colorectal cancer (CRC).* Recent Results Cancer Res, 2013. 191:157-83.
27. Jarvinen, H.J., et al., *Ten years after mutation testing for Lynch syndrome: cancer incidence and outcome in mutation-positive and mutation-negative family members.* J Clin Oncol., 2009. 27(28): p. 4793-7.
28. Vasen, H.F., et al., *One to 2-year surveillance intervals reduce risk of colorectal cancer in families with Lynch syndrome.* Gastroenterology., 2010. 138(7): p. 2300-6.
29. Hall, J., et al., *What influences participation in genetic carrier testing? Results from a discrete choice experiment.* J Health Econ., 2006. 25(3): p. 520-37.
30. Kuppermann, M., et al., *Preferences for outcomes associated with decisions to undergo or forgo genetic testing for Lynch syndrome.* Cancer, 2013. 119(1): p. 215-25.
31. Matro, J.M., et al., *Cost Sharing and Hereditary Cancer Risk: Predictors of Willingness-to-Pay for Genetic Testing.* J Genet Couns, 2014. 6: p. 6.
32. Severin, F., et al., *Eliciting preferences for priority setting in genetic testing: a pilot study comparing best-worst scaling and discrete-choice experiments.* Eur J Hum Genet., 2013. 21(11): p. 1202-8.
33. Walsh, J., et al., *Preferences for genetic testing to identify hereditary colorectal cancer: perspectives of high-risk patients, community members, and clinicians.* J Cancer Educ., 2012. 27(1): p. 112-9.
34. Hiligsmann, M., et al., *Nominal group technique to select attributes for discrete choice experiments: an example for drug treatment choice in osteoporosis.* Patient Prefer Adher, 2013. 7:133-9.
35. Bliemer, M.C.J. and J.M. Rose, *Efficiency and sample size requirements for stated choice experiments, in Transportaion Research Broad Annual Meeting.* 2009: Washington DC.
36. Huber, J. and K. Zwerina, *The Importance of Utility Balance in Efficient Choice Designs.* J Market Res, 1996. 33(3): p. 307-317.
37. Fransen, M.P., et al., *Applicability of internationally available health literacy measures in the Netherlands.* J Health Commun, 2011. 16(Suppl 3): p. 134-49.
38. Bech, M. and D. Gyrd-Hansen, *Effects coding in discrete choice experiments.* Health Econ, 2005. 14(10): p. 1079-83.
39. Fiebig, D.G., et al., *The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity.* Market Sci, 2010. 29(3): p. 393-421.
40. LeBoeuf, R.A. and E. Shafir, *Deep thoughts and shallow frames: on the susceptibility to framing effects.* J Behav Dec Making, 2003. 16: p. 77-92.
41. Whitney, P., C.A. Rinehart, and J.M. Hinson, *Framing effects under cognitive load: the role of working memory in risky decisions.* Psychon Bull Rev, 2008. 15(6): p. 1179-84.
42. Druckman, J.N., *Evaluating framing effects.* J Econ Psychol, 2001. 22: p. 91-101.
43. Timmermans, D.R.M., C.F. Ockhuysen-Vermey, and L. Henneman, *Presenting health risk information in different formats: The effect on participants' cognitive and emotional evaluation and decisions.* Patient Educ Couns, 2008. 73: p. 443-447.
44. Kuhberger, A., *The Influence of Framing on Risky Decisions: A Meta-analysis.* Organ Behav Hum Decis Process, 1998. 75(1): p. 23-55.
45. Kuhberger, A., M. Schulte-Mecklenbeck, and J. Perner, *The Effects of Framing, Reflection, Probability, and Payoff on Risk Preference in Choice Tasks.* Organ Behav Hum Decis Process, 1999. 78(3): p. 204-231.
46. Mata, R., et al., *Age differences in risky choice: a meta-analysis.* Ann N Y Acad Sci, 2011. 1235: p. 18-29.
47. Tymula, A., et al., *Like cognitive function, decision making across the life span shows profound age-related changes.* Proc Natl Acad Sci U S A, 2013. 110(42): p. 17143-8.
48. Fraser-Mackenzie, P., M.C. Sung, and J.E. Johnson, *Toward an understanding of the influence of cultural background and domain experience on the effects of risk-pricing formats on risk perception.* Risk Anal, 2014. 34(10): p. 1846-69.
49. Bridges, J.F., et al., *Conjoint analysis applications in health--a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force.* Value Health., 2011. 14(4): p. 403-13.
50. Johnson, F.R., et al., *Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force.* Value Health, 2013. 16(1): p. 3-13.

GENERAL DISCUSSION

The three main objectives of this thesis were (1) to measure individuals' preferences for specific public health interventions with a DCE, (2) to study the validity of the preferences that are measured with DCEs in public health, and (3) to assess how the presentation of choice tasks influences respondents' decision-making behavior and the outcomes of a DCE. The findings will be discussed in consecutive subsections.

Part 1: Measuring individuals' preferences for public health interventions with a DCE

In the first part of this thesis, we used DCEs to determine individuals' preferences for specific preventive interventions in three different public health domains, i.e. healthy lifestyle, vaccination, and screening. We describe the four main outcomes, as commonly reported in DCE studies [1, 2]. First, we measured individuals' preferences for the specific characteristics (attributes) of the three different preventive interventions (a lifestyle program, a rotavirus vaccination and genetic screening for colorectal cancer (CRC)). We also determined the relative importance of each of the attributes with regard to the decision to participate in a specific intervention and in comparison to the other attributes. For instance, the offered physical activity schedule, consultation structure, expected outcome, and out-of-pocket costs all influenced the decision of type 2 Diabetes Mellitus (T2DM) patients to participate in a lifestyle program, but the out-of-pocket costs were relatively the most important **(Chapter 1).** Second, based on these preferences, we estimated trade-offs among which willingness-to-pay. For instance, parents were willing to trade 20.2%-point vaccine effectiveness for a reduction in the frequency of severe side effects from 1 in 10,000 to 1 in 1,000,000 **(Chapter 2)**, and T2DM patients were willing to pay €128 a year for individual consultation instead of group consultation **(Chapter 1)**. Third, we estimated potential participation rates for different hypothetical preventive programs. For instance, 86% of the parents were willing to vaccinate their six-week-old baby against rotavirus if this vaccine would become part of the Dutch National Immunization Program (NIP), and 68% of the parents were willing to vaccinate their newborn baby if the vaccine would be offered on the private market only **(Chapter 2)**. Fourth, in all studies, we observed that specific subgroups of the population differed slightly in their preferences for (some of the) attributes (preference heterogeneity). For instance, when deciding about participation in genetic screening for CRC, respondents with a lower educational level found colonoscopy frequency relatively more important and survival relatively less important compared to higher educated respondents **(Chapter 3)**.

Each of these four outcomes provides valuable information for program developers and policy makers. First, knowledge about which program characteristics are preferred by the target population facilitates the development of preventive interventions that are perceived as attractive by the target population and hence are more likely to achieve higher participation rates. Second, specific insight into the trade-offs that people make between preferred and less desirable program characteristics helps to balance preferred program characteristics against the often scarce resources for implementation. For example, to realize a lifestyle program with an individual consultation structure as opposed to a group-based intervention might be too costly unless T2DM patients are willing to pay (part of) these costs out-of-pocket. Third, estimating potential participation rates

for different versions of the same intervention helps to determine the optimal implementation strategy of a new preventive program. For example, an 18%-point higher vaccination coverage can be expected if rotavirus vaccination is implemented as part of the NIP, compared to only being available on the private market. Fourth, knowledge about the heterogeneity in preferences among respondents assists in developing effective and/or personalized (tailored) preventive interventions and communication strategies, which might benefit participation rates and improve informed decision-making [3-6]. For instance, some T2DM patients can participate in a group-based lifestyle program, while others have individual consultation structures within the same lifestyle program. For individual level interventions, such as lifestyle programs or consultations regarding genetic screening for CRC, there are multiple tailoring opportunities. However, tailoring to this extent is not always feasible, since many public health interventions will be offered to the entire target population at once. For instance, parents might differ in their preferences for vaccine-specific characteristics, but only one vaccine will ultimately become available for the entire population. Creative solutions should be considered to respond to different preferences in the population when tailoring the intervention is not possible. For instance, web pages with detailed and structured information can be developed to inform the part of the population, for which the intervention itself is not the preferred option, about the background of the policy decisions that have been made and why certain implementation approaches are undertaken.

Although DCEs provide valuable information for policy makers and program developers, there are three closely related issues to consider given that, respondents are faced with hypothetical situations within a highly demarcated and controlled situation. First, all respondents are provided with a large amount of exactly the same information about the decision they are being asked to make. It is unlikely that, in real life, each individual also reads and carefully considers all the information that is provided when making public health related decisions. Therefore, it is unclear to what extent choice situations in DCEs are comparable to choice situations in real life, and to what extent differences in choice contexts influence the decisions being made [7-10]. Second, since respondents are valuing hypothetical situations, their observed preferences are likely to differ from their actual decisions because they have neglected or disvalued the importance of specific program characteristics (hypothetical bias) [7-9, 11, 12]. For instance, respondents often overestimate their willingness to pay for a preventive intervention [12-14], while they underestimate their willingness to accept risks [15]. Third, public health decisions are largely influenced by psychosocial factors such as emotions, perceived susceptibility and social norm (**Chapter 2 & 3**, [16-18]). Although such factors are increasingly accounted for in DCE studies, additional research is needed to determine the influence of such factors on respondents' preferences and their decision-making behavior. Different approaches exist that might help to account for the issues described above. Cheap-talk or decision-making under oath may reduce hypothetical bias [12, 19, 20], and analyzing the data using hybrid modelling techniques might capture the influence of psychosocial factors on decision-making behavior in DCEs [21, 22]. A time-to-think (TTT) approach may be used to tackle the influence of hypothetical bias and psychosocial factors at once [11, 23, 24]. In a TTT-DCE respondents are asked to read all the information in the questionnaire along with the choice tasks, without completing it. They can complete the questionnaire at least one day later. This gives them the

opportunity to discuss the topic with their partner, family or friends, or to look up additional information. Studies that used this method show promising results [11, 15, 25, 26]. These methods should be explored more extensively to determine whether they actually contribute to deriving (even) more accurate study outcomes for public health related DCEs.

In conclusion, the outcomes of a DCE can facilitate the development, implementation and marketing of effective preventive interventions, which may increase participation rates and improve informed decision-making among the target population. The influence of the choice context, hypothetical bias and psychosocial factors on decision-making behavior deserves further attention. Future DCE studies should make an effort to account for these issues by adapting their questioning (e.g., TTT-approach) or modelling (e.g., hybrid models) method. In doing so, DCEs seem a promising method to contribute to the improvement of the overall public health.

Part 2: The validity of the preferences that are measured with DCEs in public health

In part 2 of this thesis, we studied three aspects regarding the validity of the preferences that were measured by a public health related DCE. First, we showed that the majority of the respondents indeed completed the choice tasks in a DCE as explained by the normative theoretical assumptions **(Chapter 4)**. Some respondents expressed difficulties with understanding and interpreting the information provided, which might be the cause of their dominant decision-making behavior or their inability to understand the DCE questions. These 'deviations' from the theoretical assumptions about decision-making were more common among older respondents and among respondents with a lower educational level or lower health literacy. Second, we found differences in preferences between respondents with higher and lower health literacy **(Chapter 5)**. This heterogeneity in preferences is probably due to a lack of understanding or a misinterpretation, which caused respondents with a lower health literacy to neglect or otherwise undervalue specific information. Respondents with a lower health literacy deemed the two characteristics that included probabilistic information (vaccine effectiveness and frequency of severe side effects) as less important than respondents with a higher health literacy. Third, we showed that DCE results can be used to predict specifically engaging behavior (i.e. decisions to participate) since the positive predictive value and sensitivity were high (0.80 and 0.90) while the negative predictive value and specificity were low (0.44 and 0.35) **(Chapter 6)**. In 74% of the cases, we correctly predicted whether a specific T2DM patient would or would not participate in a lifestyle program, based on their stated preferences in the DCE.

Overall, it can be concluded that the vast majority of the population is capable of completing a public health related DCE as anticipated by the normative theoretical assumptions. The outcomes of DCEs can be used to predict engaging behavior. However, DCEs in their current form might not measure preferences validly among all members of the target population. First, older respondents and respondents with a lower education and/or lower health literacy more often do not (fully) understand the information provided, fail rationality or consistency tests or show dominant decision-making behavior (**Chapter 4**, [27-37]). This is of particular interest given that the population

at large is aging [38]. The increasing number of elderly people and their susceptibility to several (chronic) diseases make them a suitable target population for public health initiatives. However, it is uncertain whether DCEs, as currently practiced, can accurately measure the preferences of this population and thus assist in developing effective and attractive preventive interventions. Second, in the Netherlands, about 21% of the population is a first- or second-generation immigrant [39]. Due to cultural differences, respondents of non-Dutch origin may have different preferences for public health interventions [23]. At the same time, respondents from ethnic minorities are often underrepresented in DCE studies, probably because they are automatically but unintentionally excluded if they are unable to read Dutch. It therefore remains unclear to what extent their preferences are comparable to those of the general population. Significant effort is required to involve respondents from ethnic minorities in DCE studies and to ensure that the entire target population (including older respondents and respondents with a lower educational level and/or lower health literacy and respondents from ethnic minorities) is able to accurately complete a DCE. Cooperation with specific schools, church communities or general practitioners might help to increase participation rates among immigrants in DCE studies [23]. Surveying methods such as mini-labs or telephone interviews might also increase immigrant participation rates and additionally assist respondents with completing the DCE accurately [23, 40]. During such surveying methods, researchers explain the information provided and the process of completing a DCE in respondents' native language, which might contribute to their understanding. This will provide the unique opportunity to personalize (tailor) the assistance, since some respondents may require more explanation and help than others. A drawback is that this is time consuming for respondents and researchers, and potentially induces interviewer bias (respondents tend to be more positive to satisfy the interviewer [11, 23]). Online questionnaires might be a more efficient solution for communicating information and conducting a DCE. Several built-in tools, such as spoken texts, short video messages or demonstrations might assist respondents with correctly interpreting the information provided and completing the DCE [3, 4]. It is worthwhile to also investigate whether next of kin can be used as proxy respondents to complete a DCE when respondents themselves are unable to do so because of cognitive impairments or language difficulties [41, 42]. It should be tested in further detail to what extent the preferences of proxy respondents reflect the preferences of the individual at hand.

Part 3: The influence of the presentation of choice tasks on respondents' choice behavior and on the outcomes of a DCE

In the third part of this thesis, we studied the influence of three aspects of choice task presentation on respondents' decision-making behavior and on the outcomes of the DCE. First, we showed that presenting an opt-out option as one of the possible scenarios in a choice task, which is common in DCEs, makes respondents more prone to choosing this option **(Chapter 7)**. If respondents were forced to make a choice at first, they opted out less often if they were presented with that option later on. This can probably be attributed to a learning effect, as respondents familiarize themselves with making (difficult) decisions. Second, we showed that respondents preferred to have attribute levels described in words instead of graphics **(Chapter 8)**. Using graphics led to more inconsistent answering patterns, and the vaccine-related attributes that were included in this study were

more often interpreted incorrectly when presented in graphics. Third, we showed that framing a risk attribute either positively or negatively influenced respondents' decision-making behavior **(Chapter 9)**. Respondents attached a significantly higher value to survival (positive framing) than to mortality (negative framing). Framing risks positively induced dominant decision-making behavior, while framing a risk negatively induced risk-seeking behavior.

Currently, undoubtedly related to the absence of guidelines on this specific topic, DCE studies do not adequately report how choice tasks were presented to respondents [2, 7, 43]. Presumably, researchers decide on the presentation of choice tasks based on their best knowledge, but without empirical evidence of the consequences of presenting choice tasks one way or another. Research fields, such as social sciences, have already conducted extensive research on how the presentation of information affects human decision-making behavior [44-49]. Although such knowledge provides a promising opportunity to improve the DCE method, this seems to have gone largely unnoticed so far. Such evidence combined with the results described in **Chapter 7-9**, teaches us that people do not always behave rationally and that people are influenced greatly by how information is presented. However, health economists continue to model human decision-making behavior in DCEs as if respondents were behaving rationally; thereby misjudging the enormous effect the presentation of choice tasks has on the choices people make [7, 43].

The presentation of choice tasks needs careful consideration. For example, if researchers want to include an opt-out option in future DCEs, they can benefit from the learning effect we observed in **Chapter 7**. They may want to use a design that forces respondents to make a choice at first and then immediately after making their choice asks them if they would like to opt out [50]. Supplementary, preferably qualitative, research is required to determine why respondents opt out and whether there are subgroups of respondents who opt out more often than others do. Such empirical evidence can be used to improve the understanding of individuals' decision-making in complex situations. Theoretically it may be argued that respondents will only opt out if none of the alternatives is sufficiently attractive [51, 52], but our study adds to the empirical evidence that respondents opt out when faced with complex choice situations to protect themselves from poor decision-making [46, 53-60]. Although this finding still has to be supported by additional DCE research, it might be worthwhile to start identifying options to reduce the complexity of choice tasks. The risk information that is often included in public health related DCEs certainly contributes to the complexity of the choice tasks [61-63]. In any case, 'simply' replacing words with graphics to depict attribute levels will not reduce the complexity of a choice task, but might unintentionally increase the complexity of the decision-making task even further **(Chapter 8)**. It is worth investigating whether adding graphics to choice tasks that are presented in words might make choice tasks easier to understand. Moreover, some (types of) attributes might be more suitable to depict in graphics than others. Future research should try to validate certain graphics among specific target populations. Based on those results, a universal database with validated graphics can be developed, which assists researchers in the otherwise time-consuming task of developing graphics, while improving the validity of the measured preferences at the same time. Apart from the influence of depicting attributes in words or graphics, framing a risk attribute either as survival or as mortality

affects respondents' decision-making behavior **(Chapter 9).** It is advisable that future DCE studies discuss risk attribute framing during expert interviews and focus-group interviews. Such an approach ensures that the framing of the included risk attribute(s) resembles how those risks are communicated in practice as closely as possible. To determine the exact influence of attribute framing on respondents' decision-making behavior, extensive research is needed to map the effects of framing different attributes in different choice situations.

In conclusion, presenting choice tasks in different ways influences respondents' decision-making behavior and DCE study outcomes. Additional research should investigate these effects in detail. Learning from those experiences, adapting the DCE method according to the findings, and complementing existing guidelines [64, 65] for best practice DCEs will contribute to the quality of the data and the validity of the results of DCE studies. This is of vital importance, since poor data quality cannot be compensated for by using advanced designs and profound analytical modelling techniques. Improved transparency about choice task presentation in individual study papers and the adaptation of DCE best practice guidelines will ensure that the DCE method remains or even becomes a more valid and high quality method to measure preferences, also within the public health research setting.

Future scientific challenges

The future challenges emerging from this thesis can be divided over three sub-areas. First, the correspondence between decision-making in a DCE and real-life should be optimized. Methods that account for the influence of choice contexts, hypothetical bias and psychosocial factors in DCEs should be identified, tested and validated in the public health research setting. In this regard, specifically the time-to-think approach deserves further attention. Second, the validity of the preferences measured by DCEs in public health should be investigated in detail. To what extent alternative questioning methods, such as interviews or online interfaces, may improve the ability of immigrants, older respondents and respondents with a lower educational level and/or health literacy to complete a DCE accurately should be verified. Additionally, it is advised to conduct formal external validity studies with respect to DCEs that measure preferences among different populations for different preventive interventions. Since this is difficult in the public health setting, future studies should (at least) try to combine both stated preference and revealed preference (actual behavior) data to obtain insight into the extent to which preferences for hypothetical situations resemble and can predict actual behavior. For instance, if the addition of a new vaccine to the NIP is anticipated in the near future, vaccination decisions of the target population can be combined with and compared to preference of respondents in a vaccination-related DCE. Third, the quality of the data that is gathered by DCE studies should be further secured by increased attention among researchers for the presentation of choice tasks. In addition to in depth investigation of the influence of the opt-out option, the use of graphics, and risk attribute framing, other issues regarding choice task presentation (e.g., color schemes, goal framing, reference point bias, time dependency) should be identified and investigated.

Overall conclusion

When DCEs were applied to public health interventions, all associated measures such as the relative importance of attributes, trade-offs between attributes, willingness-to-pay and potential participation rates could be estimated. Those results have given rise to specific recommendations for program developers and policy makers to facilitate the development, implementation and marketing of effective preventive interventions. The vast majority of the population appeared to be very well capable of completing public health related DCEs and the results could be used to predict engaging behavior.

However, the influence of the choice context, hypothetical bias and psychosocial factors on the decision-making process of respondents is unclear. Some subgroups expressed difficulties with completing DCEs according to the normative theoretical assumptions. Additionally, presenting choice tasks in different ways influenced respondents' decision-making behavior and DCE study outcomes. Fortunately, all these current limitations of the DCE method in public health, might be overcome by conducting additional research and adapting existing guidelines.

In conclusion, DCEs can be used to measure individuals' preferences for public health interventions. There are several options to improve this method specifically for the public health setting. By combining additional research, knowledge from other scientific disciplines, well designed experiments and advanced analytical models, our capacity to accurately measure individuals' preferences and to understand individuals' decision-making behavior will increase even further. Therefore, DCEs are a promising method to contribute to the improvement of the overall public health.

References

1. Clark, M.D., et al., *Discrete Choice Experiments in Health Economics: A Review of the Literature.* Pharmacoeconomics, 2014. 9: p. 9.
2. Marshall, D., et al., *Conjoint Analysis Applications in Health - How are Studies being Designed and Reported?: An Update on Current Practice in the Published Literature between 2005 and 2008.* Patient., 2010. 3(4): p. 249-56.
3. de Vries, H. and J. Brug, *Computer-tailored interventions motivating people to adopt health promoting behaviours: introduction to a new approach.* Patient Educ Couns, 1999. 36(2): p. 99-105.
4. Krebs, P., J.O. Prochaska, and J.S. Rossi, *A meta-analysis of computer-tailored interventions for health behavior change.* Prev Med, 2010. 51(3-4): p. 214-21.
5. Kreuter, M.W., et al., *Understanding how people process health information: a comparison of tailored and nontailored weight-loss materials.* Health Psychol, 1999. 18(5): p. 487-94.
6. Kreuter, M.W., V.J. Strecher, and B. Glassman, *One size does not fit all: the case for tailoring print materials.* Ann Behav Med, 1999. 21(4): p. 276-83.
7. Lloyd, A.J., *Threats to the estimation of benefit: are preference elicitation methods accurate?* Health Econ, 2003. 12(5): p. 393-402.
8. Hensher, D.A., J.M. Rose, and W.H. Greene, *Applied Choice Analysis: A Primer.* 2005, New York: Cambridge University Press.
9. Louviere, J.J., D.A. Hensher, and J.D. Swait, *Stated Choice Methods; analysis and apllication.* 2000, Cambridge: Cambridge University Press.
10. Brownstone, D., D.S. Bunch, and K.E. Train, *Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles.* Transport Res B Meth, 2000. 34(5): p. 315-338.
11. Ozdemir, S., *Improving the Validity of Stated-Preference Data in Health Research: The Potential of the Time-to-Think Approach.* Patient, 2014.
12. Loomis, J., *What's to know about hypothetical bias in stated preference valution studies?* J Econ Surveys, 2011. (25): p. 2.
13. Jacquemet, N., et al., *Do People Always Pay Less Than They Say? Testbed Laboratory Experiments with IV and HG Values* J Public Econ Theory, 2011. 13(5): p. 857-882.
14. List, J.A. and C.A. Gallet, *What Experimental Protocol Influence Disparities between Actual and Hypothetical Stated Values?* . Environ Resour Econ, 2001. 20: p. 241-254.
15. Johnson, F.R., et al., *No time-to-think about benefit-risk preferences: an experiment to test the validity of patients' stated preferences.* Cornell University: Ithaca, NY, 2010.
16. Determann, D., et al., *Acceptance of vaccinations in pandemic outbreaks: a discrete choice experiment.* PLoS One., 2014. 9(7): p. e102505.
17. de Bekker-Grob, E.W., et al., *Men's preferences for prostate cancer screening: a discrete choice experiment.* Br J Cancer., 2013. 108(3): p. 533-41.
18. Struik, M.H., et al., *The preferences of users of electronic medical records in hospitals: quantifying the relative importance of barriers and facilitators of an innovation.* Implement Sci, 2014. 9: p. 69.
19. Ozdemir, S., F.R. Johnson, and A.B. Hauber, *Hypothetical bias, cheap talk, and stated willingness to pay for health care.* J Health Econ, 2009. 28(4): p. 894-901.
20. Stevens, T.H., M. Tabatabaei, and D. Lass, *Oaths and hypothetical bias.* J Environ Manage, 2013. 127: p. 135-41.
21. Hess, S. and A. Stathopoulos, *Linking response quality to survey engagement: A combined random scale and latent variable approach.* J Choice Modelling, 2013. 7: p. 1-12.
22. Hoyos, D., P. Mariel, and S. Hess, *Incorporating environmental attitudes in discrete choice models: An exploration of the utility of the awareness of consequences scale.* Sci Total Environ, 2015. 505: p. 1100-11.
23. Whittington, D., *What have we learned from 20 uears of stated preference research is less-developed countries?* Annu Rev Resour Econ, 2010. 2: p. 209-236.
24. Whittington, D., et al., *Giving respondents time to think in contingent valuation studies: A developing country application* J Environ Econ Management, 1992. 22(3): p. 205-225.

25. Cook, J., et al., *Reliability of stated preferences for cholera and typhoid vaccines with time to think in Hue, Vietnam.* Econ Inq, 2007. 45(1): p. 100-114.
26. Cook, J., et al., *Giving Stated Preference Respondents "Time to Think": Results From Four Countries.* Environ Resour Econ, 2011. 51: p. 473-496.
27. Scott, A., *Identifying and analysing dominant preferences in discrete choice experiments: an application in health care.* J Econ Psychol, 2002. 23(3): p. 383-398.
28. Miguel, F.S., M. Ryan, and M. Amaya-Amaya, *'Irrational' stated preferences: a quantitative and qualitative investigation.* Health Econ, 2005. 14(3): p. 307-22.
29. Kenny, P., et al., *Do participants understand a stated preference health survey? A qualitative approach to assessing validity.* Int J Technol Assess Health Care, 2003. 19(4): p. 664-81.
30. Ozdemir, S., et al., *Who pays attention in stated-choice surveys?* Health Econ., 2010. 19(1): p. 111-8.
31. Alemu, M.H., et al., *Attending to the reasons for attribute non-attendance in choice experiments.* Environ Resour Econ, 2013. 54(3): p. 333-359.
32. Bech, M., T. Kjaer, and J. Lauridsen, *Does the number of choice sets matter? Results from a web survey applying a discrete choice experiment.* Health Econ, 2011. 20(3): p. 273-86.
33. Cheraghi-Sohi, S., et al., *Making sense of patient priorities: applying discrete choice methods in primary care using 'think aloud' technique.* Fam Pract, 2007. 24: p. 276-82.
34. Lagarde, M., *Investigating attribute non-attendance and its consequences in choice experiments with latent class models.* Health Econ, 2013. 22(5): p. 554-67.
35. Ryan, M. and A. Bate, *Testing the assumptions of rationality, continuity and symmetry when applying discrete choice experiments in health care.* Appl Econ Lett, 2001. 8(1): p. 59-63.
36. Ryan, M., V. Watson, and V. Entwistle, *Rationalising the 'irrational': a think aloud study of discrete choice experiment responses.* Health Econ, 2009. 18(3): p. 321-36.
37. Iles, R.A. and J.M. Rose, *Stated Choice design comparison in a developing country: recall and attribute nonattendance.* Health Econ Rev, 2014. 4: p. 25.
38. National Institute for Public Health and the Environment. *Trends in life-expectancy at birth* 2014. Available from: Volksgezondheidenzorg.info.
39. CBS. *Population grouped by origin.* 2014. Available from: http://statline.cbs.nl.
40. Ahlmark, N., et al., *Survey nonresponse among ethnic minorities in a national health survey - a mixed-method study of participation, barriers, and potentials.* Ethn Health, 2014: p. 1-22.
41. Prosser, L.A., J.K. Hammitt, and R. Keren, *Measuring health preferences for use in cost-utility and cost-benefit analyses of interventions in children: theoretical and methodological considerations.* Pharmacoeconomics, 2007. 25(9): p. 713-26.
42. Milte, R., et al., *Cognitive overload? An exploration of the potential impact of cognitive functioning in discrete choice experiments with older people in health care.* Value Health, 2014. 17(5): p. 655-9.
43. Bryan, S. and P. Dolan, *Discrete choice experiments in health economics. For better or for worse?* Eur J Health Econ, 2004. 5(3): p. 199-202.
44. Kahneman, D., J.L. Knetsch, and R.H. Thaler, *The endowment effect, loss aversion and status quo bias.* J Econ Perspect, 1991. 5(1): p. 193-206.
45. Lipkus, I.M., *Numeric, verbal, and visual formats of conveying health risks: suggested best practices and future recommendations.* Med Decis Making., 2007. 27(5): p. 696-713.
46. Luce, M.F., J.W. Payne, and J.R. Bettman, *Emotional trade-off difficulty and choice.* J Market Res, 1999. 36(2): p. 143-159.
47. Timmermans, D.R.M., C.F. Ockhuysen-Vermey, and L. Henneman, *Presenting health risk information in different formats: The effect on participants' cognitive and emotional evaluation and decisions.* Patient Educ Couns, 2008. 73: p. 443-447.
48. Tversky, A. and D. Kahneman, *Judgment under uncertainty: Heuristics and biases.* science, 1974. 185(4157): p. 1124-1131.
49. Tversky, A. and D. Kahneman, *The framing of decisions and the psychology of choice.* Science, 1981. 211(4481): p. 453-8.

50. Brazell, J.D., et al., *The no-choice option and dual response choice designs.* Market Lett, 2006. 17: p. 255-268.

51. McFadden, D., *Conditional Logit Analysis of Qualitative Choice Behavior,* in *Frondtiers in Econometrics,* P. Zarembka, Editor. 1974, Academic Press: New York.

52. McFadden, D., *The Choice Theory Approach to Market Research.* Marketing Science, 1986. 5(4): p. 275–297.

53. Boxall, P., W.L. Adamowicz, and A. Moon, *Complexity in choice experiments: choice of status quo alternative and implications for welfare measurement.* Aust J Arg Resour Ec, 2009. 53: p. 503-519.

54. Dhar, R., *Consumer preference for a no-choice option.* J Cons Res, 1997. 24(2): p. 215-231.

55. Dhar, R. and S. Itamar, *The effect of forced choice on choice.* J Market Res, 2003. 40 (2): p. 146-160.

56. Luce, M.F., *Choosing to avoid: coping with negative emotion-laden consumer desicions.* J Cons Res, 1998. 24(4): p. 409-433.

57. Meyerhoff, J. and U. Liebe, *Status quo effect in choice experiments: emperical evidence on attitude and choice task complexity.* Land Economics, 2009. 85(3): p. 515-528.

58. Nowlis, S.M., B.E. Kahn, and R. Dhar, *Coping with ambivalence: the effect of removing a neutral option on consumer attitude and preference Judgements.* J Cons Res, 2002. 29(3): p. 319-334.

59. Ritov, I. and J. Baron, *Status quo and ommission biases.* J Risk Uncertainty, 1992. 5: p. 49-61.

60. Tversky, A. and E. Shafir, *Choice under conflict: the dinamics of differed discision.* Psychological Schience, 1992. 3(6): p. 358-361.

61. Galesic, M. and R. Garcia-Retamero, *Statistical numeracy for health: a cross-cultural comparison with probabilistic national samples.* Arch Intern Med., 2010. 170(5): p. 462-8.

62. Waters, E.A., et al., *Formats for improving risk communication in medical tradeoff decisions.* J Health Commun., 2006. 11(2): p. 167-82.

63. Harrison, M., et al., *Risk as an attribute in discrete choice experiments: a systematic review of the literature.* Patient, 2014. 7(2): p. 151-70.

64. Bridges, J.F., et al., *Conjoint analysis applications in health--a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force.* Value Health., 2011. 14(4): p. 403-13.

65. Lancsar, E. and J. Louviere, *Conducting discrete choice experiments to inform healthcare decision making: a user's guide.* Pharmacoeconomics, 2008. 26(8): p. 661-77.

SUMMARY

One approach to improve public health is to implement preventive programs that have been proven effective and cost-effective at the population level. For any preventive program to be successful in those terms, it is of paramount importance that a large majority of the target population participates. Unfortunately, it is not self-evident that high participation rates will be attained once preventive programs become available.

Characteristics that influence individuals' participation decisions have to be identified to understand why they decide (not) to participate in effective preventive programs. Besides the well-investigated demographic and psychosocial characteristics and environmental variables, certain program characteristics such as costs or side effects, might also contribute to the low uptake of some preventive programs. Insight into preferences of the target population for specific program characteristics is crucial for the development of new, attractive, and broadly used preventive programs, and for improving preventive programs that already exist. Discrete Choice Experiments (DCEs) are relatively new in the field of public health but have acquired an established position for studying decision-making behavior in other fields, such as marketing and transportation. The DCE methodology is a stated preference method that requires respondents to choose between different sets of program characteristics rather than just ranking or rating a single characteristic which is most common in traditional Likert-scale questionnaires. The method is based on several normative theoretical assumptions that describe how people are assumed to make decisions. DCEs enable researchers to determine the preferred program characteristics, to quantify the importance of those program characteristics relative to each other and to estimate potential participation rates. Therefore, DCEs appear to be a promising method to measure individuals' preferences concerning characteristics of specific preventive programs, which provides a basis for recommendations about program development that contributes to increased participation rates.

This thesis aims to determine whether DCEs can be used to measure individuals' preferences for public health interventions. This aim was further specified in three detailed objectives for three different parts of this thesis:

1. To measure individuals' preferences for specific public health interventions with a DCE;
2. To study the validity of the preferences that are measured with DCEs in public health;
3. To assess how the presentation of choice tasks influences respondents' decision-making behavior and the outcomes of a DCE.

Part 1: Measuring individuals' preferences for public health interventions with a DCE

DCEs are developed to measure individuals' preferences for specific product characteristics and to inform product developers about the optimal design to maximize the attraction of a particular product. In the light of this marketing purpose, it is useful to test whether people will have clear and meaningful preferences for specific characteristics of preventive programs, and whether the results of a public health related DCE will provide clear recommendations for the 'marketing' of preventive programs. In the first part of this thesis, we describe three DCEs within the public heath setting.

In **Chapter 1**, we describe a DCE that aims to identify the preferences of type 2 diabetes mellitus (T2DM) patients aged 35-65 years for different characteristics of lifestyle programs, to assess their willingness-to-pay (WTP) and to estimate the potential participation rate for different lifestyle programs. We showed that the out-of-pocket costs were most important for T2DM patients when deciding about participating in a lifestyle program, followed by the consultation structure, the expected outcome (in terms of weight loss) and the included physical activity (PA) schedule. Patients were willing to pay €21.00 for switching from a very flexible to a more general PA schedule. For a lifestyle program organized via individual consultation instead of consultation with a group of 10 other patients, they were willing to pay an extra €127.80 per year. For 10 kg anticipated weight loss respondents were willing to pay €96.80 per year. Finally, the estimated potential participation rates ranged between 48.5% and 62.4% depending on the specific composition of the lifestyle program. In conclusion, it is advised that lifestyle programs directed at T2DM patients use an individual consultation structure, communicate clearly about the expected outcomes of the program (in terms of weight loss) and keep the out-of-pocket costs as limited as possible.

In **Chapter 2**, we describe a DCE that aims to determine parental preferences with regard to rotavirus vaccination for their six-weeks old newborn and to estimate the potential vaccination coverage for different vaccine scenarios and different ways to implement the vaccine. We showed that, the vaccine effectiveness, the frequency of severe side-effects, the protection duration and the out-of-pocket costs all influenced parents' decision about rotavirus vaccination. The costs of the vaccine were most important. Parents were willing to trade 20.2%-points vaccine effectiveness for a switch in the frequency of severe side effects from 1 in 10,000 to 1 in 1,000,000, or 20.8%-points for a protection duration of three years instead of one year. Potential vaccination coverage ranged between 22.7% and 86.2%, depending on the vaccine scenario and the implementation strategy. In conclusion, when deciding about vaccination against rotavirus, parents were mostly driven by the out-of-pocket costs. The highest vaccination coverage was expected for a vaccine that is very effective, has a long protection duration and is implemented within the current National Immunization Program (implying no out-of-pocket costs). Implementation of a rotavirus vaccine in the free market will result in lowest coverage rates.

In **Chapter 3**, we describe a DCE that aims to explore the preferences of the general population aged 55-65 years concerning genetic screening for colorectal cancer (CRC) within a population-based CRC screening program and to estimate whether they are willing to participate in genetic screening for (1) Lynch syndrome, (2) familial adenomatous polyposis (FAP), and (3) familial colorectal cancer (FCC). We showed that the probability of being genetically predisposed, the probability of developing CRC, the colonoscopy frequency and the survival all significantly influenced their decision to participate in genetic screening. Higher educated respondents, respondents who had no experience with genetic testing and respondents who expressed no anxiety and worries about being genetically predisposed found survival to be most important, while lower educated respondents, respondents who had experience with genetic testing and respondents who expressed serious anxiety and worries about being genetically predisposed found the colonoscopy frequency most important. These differences in preferences resulted in

opposite preferences for participation in different screening scenarios. In conclusion, the general population is willing to participate in genetic screening for CRC if properly informed. If individuals are suspected of genetic or familial CRC, they should at least be informed about their increased risk of being genetically predisposed and about the importance of participating in all preventive follow-up colonoscopies in order to maximize their survival. Specifically, respondents with a lower educational level and those who express serious worries or anxiety about a genetic predisposition should be informed about the frequency of preventive colonoscopies that is appropriate for the genetic or familial CRC they are diagnosed with.

Part 2: The validity of the preferences that are measured with DCEs in public health

In the second part of this thesis we describe how three validity-aspects regarding the preferences that are measured with a DCE in public health were tested. Compared to everyday decisions, deciding about participation in a preventive program is quite different and probably more difficult. Therefore, some respondents might not be able to complete DCEs in the field of public health as described by the normative theoretical assumptions. In **Chapter 4**, we determined to what extent participants evaluate and complete the choice tasks in DCE according to the underlying methodological assumptions by means of 70 structured interviews. We found that participants were actively involved and willing to provide the response that was required. Most participants understood the attributes and acted in accordance with the monotonicity (i.e. prefer more optimal attribute levels over less optimal attribute levels) and the continuity axiom (i.e. decide based on the majority of the attributes). However, about a third of the participants used simplified decision-strategies such as deciding based on less than three attributes and about 4% of the participants reported inverse preferences concerning attribute levels. These violations of the normative theoretical assumptions were more pronounced among participants with a higher age, a lower educational level and/or a lower health literacy. In conclusion, the majority of the participants seems to complete a DCE as presumed by its methodology. However, based on the participants' age, educational level and health literacy, additional measures should be undertaken to ensure that every participant understands all the choice tasks and completes the DCE as presumed.

The inability of some respondents to act in accordance with (all) the normative theoretical assumptions might not be of immediate concern to the overall study outcomes, unless respondents' decision-making behavior is influenced in such a way that different preferences are identified among different subgroups of the population. Such differences in preferences might then be incorrectly reported as preference heterogeneity. Evidently, this would seriously hamper the validity of the measured preferences. To further study this, we focused on the role of health literacy. The research described in **Chapter 5** was incorporated in the DCE on rotavirus vaccination as described in **Chapter 2**. We found that health literacy is associated with parental preferences concerning rotavirus vaccination. Vaccine effectiveness and the frequency of severe side effects were perceived as more important, while protection duration was perceived as less important by parents with a higher health literacy and educational level compared to less health literate and lower educated parents. These differences in preference between respondents with a lower and a

higher health literacy might not be caused by actual differences in preference structures, but might be better explained by varying levels of understanding. It should be explored if preferences of more and less health literate respondents still differ if they interpret the numerical value of risks as similar as possible. Such research will reveal to what extent a decision-making process is influenced by understanding certain (vaccine) characteristics. The results of this study call for health literacy as an important factor to take into account when studying (vaccination) decision-behavior.

When preferences are prone to misinterpretation, they might not reflect respondents' true preferences and therefore may predict behavior poorly. Since the outcomes of DCEs are increasingly being used as input for policy-making, the extent to which preferences as measured in the DCE can predict actual behavior is of societal interest. In **Chapter 6**, the predictive value of a DCE in public health was assessed by comparing stated preferences and actual behavior. Results at the aggregated population level showed that the stated preferences slightly but not statistically significantly underestimated actual behavior. In 74.1% of the cases, the stated preferences corresponded with actual behavior at individual level, which resulted in a positive predicted value of 0.80 and a negative predictive value of 0.44, a sensitivity of 0.90 and a specificity of 0.35. Although actual behavior could partly be predicted based on the stated preferences elicited by the DCE, not all predictions were correct. More respondents for whom participation was predicted actually opted-out compared to respondents who were predicted to opt-out but actually participated. In conclusion, stated preferences can adequately predict participation behavior in a public health setting. However, it is unclear to what extent DCEs can predict refraining behavior. Future research should assess the predictive value of DCEs using different approaches among different patient groups and in different decision contexts.

Part 3: the influence of the presentation of choice tasks on respondents' choice behavior and on the outcomes of a DCE

Within every DCE project, researchers decide about how they will present their choice tasks to respondents. The possible influence of choice task presenstation on respondents' decision-making behavior was, until now, largely unknown and evidence about the extent to which the presentation of choice tasks affects DCE study outcomes was lacking. In the third part of this thesis, we focused on three aspects of choice task presentation.

In **Chapter 7**, we determined to what extent the inclusion of an opt-out option influences respondents' choice behavior, the attribute level estimates, the relative importance of the attributes and the calculated trade-offs. This research was incorporated in the DCE on lifestyle programs as described in **Chapter 1**. We showed increased opt out behavior when respondents were offered this option immediately as compared to respondents that answered forced choice tasks first. Moreover, respondents with a lower educational level opted out significantly more often. The attribute estimates of the forced-choice and opt-out dataset differed. Although there were no notable differences in the relative order of the attributes (as compared to each other), the importance weights of the attributes did differ between the datasets. This caused a significant

difference in the WTP of patients for a general PA schedule (i.e. respectively 44 and 19 euro per year). In conclusion, based on the learning effect that was shown in this study, future DCEs that include an opt-out option may want to incorporate multiple forced choice warm-up exercises or use a dual response design. Additionally, future research should empirically explore how choice tasks should be presented to make them as easy as possible, to minimize the proportion of respondents that chooses to opt-out because they find the choice tasks too complex or difficult. Finally, additional research that uses debriefing of respondents should be conducted to explore the reasons for choosing the opt-out alternative in depth.

In **Chapter 8**, we determined whether presenting attribute levels in words or graphics generates different results with respect to attribute level interpretation, the relative importance of the attributes and the estimated potential participation rates. This research was incorporated in the DCE on rotavirus vaccination as described in **Chapter 2**. Respondents reported that the use of words results in significantly more clear and easy choice tasks compared to the use of graphics. Using words led to significantly more consistent answering patterns and more accurate attribute level interpretation. In the dataset from the DCE in words, out-of-pocket costs were most important while the frequency of severe side effects was most important in the dataset from the DCE in graphics. Although the potential participation rates based on both models showed a high correlation (ICC=0.94), the estimated participation rates based on the graphics DCE were generally lower compared to the participation rates based on the words DCE. In conclusion, results differ when words or graphics are used to present attribute levels in choice tasks. In the absence of any guidelines on how to present a choice task, future research on this topic is recommended in order to establish guidelines about valid presentation of attribute levels in the choice tasks of a DCE. Until then, it is advised not to use only graphics to present the attribute levels. Additionally, if researchers decide to use graphics, attribute level presentation should be incorporated in the focus group phase of the designing stage of a DCE and should be extensively pilot tested among the target population.

In **Chapter 9**, we determined whether framing a risk attribute either positively or negatively influences respondents' decision-making behavior and their preferences. This research was incorporated in the DCE on genetic screening for CRC as described in **Chapter 3**. Based on the positively framed DCE, a significantly higher proportion of respondents always chose the scenario with the highest level of the survival attribute compared to the proportion of respondents who always chose the scenario with the lowest level of the mortality attribute in the negatively framed DCE. Respondents who answered the negatively framed DCE significantly more often opted-out. Based on the positive framing, survival was most important, while colonoscopy frequency was most important based on the negative framing. In conclusion, risk framing affects how respondents value the presented risk thereby affecting the outcomes of the DCE study. Positive risk framing led to increased dominant decision-making behavior while negative risk framing led to risk seeking behavior. Future research on this topic is recommended in order to establish guidelines on how to overcome or otherwise deal with framing effects in DCEs. Until such guidelines are developed, risks should be framed carefully. Attribute framing should have a prominent part in the expert

interviews and focus group interviews that are conducted to select the attributes and attribute levels of a DCE. Preferably, different frames should be used in a pilot version of the DCE as well as in the actual DCE to enable researchers to estimate the magnitude of the effect of choosing different frames.

General discussion and conclusion

When applying DCEs to public health interventions, all associated measures could be estimated. In contrast to the more traditional Likert-scale questionnaires, we were able to actually quantify respondents' preferences by calculating measures such as the relative importance of the attributes, the trade-offs between the attributes and the willingness-to-pay. Moreover, we were able to estimate potential participation rates for different hypothetical preventive programs. Each of these results provided specific recommendations for program developers and policy makers to facilitate the development, implementation and marketing of effective preventive interventions. The vast majority of the population was very well capable of completing public health related DCEs and the results could be used to predict engaging behavior.

However, there are some remaining issues regarding DCEs in public health that should be considered. First, respondents of a DCE are faced with hypothetical situations in a highly controlled situation. Therefore, the choice context, hypothetical bias and psychosocial factors might influence the decisions that respondents make and may hamper the comparability between their answers in a DCE and their real-life decisions. However, to date it is largely unclear to what extent these issues influence the decision-making process of respondents. Different approaches exist that might help to account for this. Cheap-talk or decision-making under oath may reduce hypothetical bias, and analyzing the data using hybrid modelling techniques might capture the influence of psychosocial factors. A time-to-think approach may be used to tackle the influence of hypothetical bias and psychosocial factors at once. Either of these methods should be extensively tested and validated in the public health research setting. Second, some subgroups expressed difficulties with completing DCEs according to the normative theoretical assumptions. This is of particular interest given that the population at large is aging. The increasing number of elderly people and their susceptibility to several (chronic) diseases make them a suitable target population for public health initiatives. However, it is uncertain whether DCEs, as currently practiced, can accurately measure the preferences of this population and thus assist in developing effective and attractive preventive interventions. To what extent alternative questioning methods, such as interviews or online interfaces, may improve the ability of specifically older respondents and respondents with a lower educational level and/or health literacy to complete a DCE accurately, should be verified. Third, presenting choice tasks in different ways influenced respondents' decision-making behavior and DCE study outcomes. In addition to in depth investigation of the influence of the opt-out option, the use of graphics, and risk attribute framing, other issues regarding choice task presentation (e.g., color schemes, goal framing, reference point bias, time dependency) should be identified and investigated. Learning from those experiences, adapting the DCE method according to the findings, and complementing existing guidelines for best practice DCEs will contribute to the quality of the data and the validity of the results of DCE studies. This is of vital importance, since

poor data quality cannot be compensated for by using advanced designs and profound analytical modelling techniques. Improved transparency about choice task presentation in individual study papers and the adaptation of DCE best practice guidelines will ensure that the DCE method remains or even becomes a more valid and high quality method to measure preferences, also within the public health research setting.

In conclusion, DCEs can be used to measure individuals' preferences for public health interventions and to provide recommendations for program development that might contribute to increased participation rates. There are several options to improve this method specifically for the public health setting. By combining additional research, knowledge from other scientific disciplines, well-designed experiments and advanced analytical models, our capacity to measure individuals' preferences accurately and to understand individuals' decision-making behavior will increase even further. Therefore, DCEs are a promising method to contribute to the improvement of the overall public health.

SAMENVATTING IN HET NEDERLANDS

Het implementeren van preventieve programma's die op populatie niveau effectief en kosteneffectief zijn gebleken, is één van de manieren om de volksgezondheid te verbeteren. Om een preventief programma succesvol te laten zijn is het van groot belang dat de meerderheid van de doelgroep waarvoor dit programma bedoeld is ook deelneemt aan het programma. Helaas is het niet vanzelfsprekend dat een hoge deelname wordt bereikt wanneer een preventief programma wordt ingevoerd.

Om te begrijpen waarom mensen beslissen om (niet) deel te nemen aan effectieve preventieve programma's, is het belangrijk om te weten welke factoren hun beslissingen over deelname beïnvloeden. Naast de veel onderzochte demografische en psychosociale karakteristieken en omgevingsfactoren, dragen bepaalde specifieke programmakenmerken, zoals kosten en bijwerkingen wellicht ook bij aan de lage deelnamecijfers van sommige preventieve programma's. Deze informatie is niet alleen cruciaal voor het ontwikkelen van nieuwe, aantrekkelijke en veel gebruikte preventieve programma's, maar ook voor het verbeteren van programma's die al bestaan. Discrete Keuze Experimenten (DCEs) hebben een erkende positie in het onderzoek naar beslisgedrag in de marketing en transportsector, maar op het terrein van de volksgezondheid is deze methode nog relatief nieuw. Terwijl klassieke Likertschaal vragenlijsten sociaal wenselijke antwoorden in de hand werken door mensen elk individueel programmakenmerk apart te laten waarderen, worden respondenten in een DCE gevraagd om te kiezen tussen verschillende sets van programmakenmerken ('stated preference' methode). Deze methode is gebaseerd op verschillende normatieve theoretische aannames die beschrijven hoe mensen worden verondersteld een keuze te maken. DCEs stellen onderzoekers in staat om te meten welke programmakenmerken de voorkeur hebben binnen de doelgroep, te bepalen hoe belangrijk die kenmerken zijn ten opzichte van elkaar en te schatten wat de potentiele deelnamebereidheid van een bepaald programma zal zijn. Kortom, DCEs lijken een veelbelovende methode voor het meten van de voorkeuren voor specifieke kenmerken van preventieve programma's, die vervolgens de basis vormen voor aanbevelingen over programmaontwikkeling en zo kunnen bijdragen aan het verhogen van de deelnamebereidheid.

Dit proefschrift heeft als doel om te bepalen of DCEs gebruikt kunnen worden om voorkeuren van mensen te meten voor volksgezondheid gerelateerde interventies. Dit doel is verder gespecificeerd in drie gedetailleerde doelen voor drie verschillende onderdelen van dit proefschrift:

1. Het meten van de preferenties van mensen voor specifieke volksgezondheid gerelateerde interventies met een DCE;
2. Het bestuderen van de validiteit van de preferenties die worden gemeten met een DCE in de volksgezondheid;
3. Het bepalen van de invloed van de presentie van keuzesets op het beslisgedrag van respondenten en de uitkomsten van een DCE.

Deel 1: Het meten van de preferenties van mensen voor specifieke volksgezondheid-gerelateerde interventies met een DCE

DCEs zijn ontwikkeld om voorkeuren van mensen voor specifieke productkenmerken te meten en daarmee productontwikkelaars te informeren over de optimale vormgeving van een aantrekkelijk product. Gezien dit duidelijke marketingdoel van DCEs is het interessant om te onderzoeken of mensen ook duidelijke voorkeuren hebben voor de kenmerken van specifieke preventieve programma's en of de resultaten van een volksgezondheid-gerelateerde DCE duidelijke aanknopingspunten en aanbevelingen kan verschaffen voor de 'marketing' van preventieve programma's. In het eerste deel van dit proefschrift beschrijven we drie DCEs op het terrein van volksgezondheid.

In **Hoofdstuk 1** beschrijven we een DCE die is uitgevoerd om inzicht te krijgen in de voorkeuren van diabetes mellitus type 2 (DM2) patiënten in de leeftijd van 35-65 jaar voor verschillende kenmerken van leefstijlprogramma's, hun bereidheid tot betalen en de potentiele deelnamebereidheid aan verschillende leefstijlprogramma's. We laten zien dat de kosten de meest belangrijke factor waren voor de beslissing van DM2 patiënten om deel te nemen aan een leefstijlprogramma, gevolgd door de structuur van de consulten, de te verwachten hoeveelheid gewichtsverlies en het sport-schema. Patiënten waren bereid om €21,00 te betalen voor een switch van een zeer flexibel naar een meer algemeen sport-schema. Voor een programma met individuele consulten in plaats van consulten in groepjes met 10 andere DM2 patiënten waren ze bereid om €127,80 per jaar extra te betalen. Voor een gewichtsverlies van 10 kg waren de respondenten bereid van €96,80 per jaar te betalen. De potentiele deelnamebereidheid varieerde tussen de 48,5% en 62,4% en was afhankelijk van de inhoud en de kosten van het leefstijlprogramma. Concluderend, leefstijlprogramma's voor DM2 patiënten zouden bij voorkeur individuele consulten aan moeten bieden, duidelijk moeten communiceren over de te verwachten uitkomsten van het programma wat betreft gewichtsverlies en de kosten van het programma zouden zo laag mogelijk moeten zijn.

In **Hoofdstuk 2** beschrijven we een DCE die is uitgevoerd om de voorkeuren van ouders voor het vaccineren van hun zes weken oude baby tegen rotavirusinfecties te meten en om de potentiele vaccinatiegraad (deelnamebereidheid) te schatten. We laten zien dat de effectiviteit van het vaccin, de frequentie van ernstige bijwerkingen, de beschermingsduur en de kosten allemaal van invloed zijn op de beslissing van ouders over rotavirus vaccinatie. De kosten van het vaccin bleken het meest belangrijk. Ouders waren bereid om 20,2%-punt vaccin effectiviteit op te geven als daarmee de frequentie van ernstige bijwerkingen verlaagde van 1 op 10.000 naar 1 op 1.000.000 en 20,8%-punt als daarmee de beschermingsduur werd verlengd van één jaar naar drie jaar. De potentiele vaccinatiegraad lag tussen de 22,7% en 86,2% en was afhankelijk van het vaccinatie scenario en de implementatiestrategie. Concluderend, als ouders beslissen over vaccinatie tegen rotavirus zijn de kosten van het vaccin het meest belangrijk. De hoogste vaccinatiegraad kan worden verwacht voor een vaccin met een hoge effectiviteit en een lange beschermingsduur dat wordt geïmplementeerd binnen het huidige rijksvaccinatieprogramma (dit betekent dat het vaccin kosteloos is). Het aanbieden van een rotavirus vaccin op de vrije markt zal op basis van deze resulteren in de laagste vaccinatiegraad.

In **Hoofdstuk 3** hebben we een DCE beschreven waarin de voorkeuren van de algemene bevolking in de leeftijd van 55-65 jaar voor genetische screening naar darmkanker binnen een landelijk darmkanker screeningsprogramma zijn gemeten en is bepaald of zij bereid zijn deel te nemen aan genetische screening voor (1) Lynch syndroom, (2) familiaire en erfelijke CRC met adenomateuze polyposis (FAP) en (3) familiaire en erfelijke CRC (FCC). We laten zien dat de kans om genetisch belast te zijn, de kans om darmkanker te ontwikkelen, de frequentie van preventieve coloscopieën en de kans om te overleven allemaal belangrijk waren voor de beslissing over deelname aan genetische screening. Mensen met een hoge opleiding, mensen zonder ervaring met genetische testen en mensen die aangaven zich geen ernstige zorgen te maken over een mogelijke genetische belasting vonden de kans om te overleven het meest belangrijk, terwijl mensen met een lage opleiding, mensen met ervaring met genetische testen en mensen die aangaven bang te zijn of zich ernstige te zorgen maken over een mogelijke genetische belasting, de frequentie van preventieve coloscopieën het meest belangrijk vonden. Deze verschillen tussen de groepen zorgden voor tegenovergestelde voorkeuren voor verschillende screening scenario's. Concluderend, de algemene bevolking is bereid om deel te nemen aan genetische screening naar darmkanker als zij op juiste wijze geïnformeerd worden. Mensen die mogelijk belast zijn met familiaire of erfelijke darmkanker zouden moeten worden geïnformeerd over de kans die zij hebben om genetisch belast te zijn, maar zeker ook over het belang van het deelnemen aan alle preventieve vervolg-coloscopieën, aangezien dit de kans om darmkanker te overleven maximaliseert. Mensen met een lage opleiding en mensen die zich ernstige zorgen maken over een mogelijke genetische belasting moeten in het bijzonder worden geïnformeerd over de frequentie van preventieve vervolg-coloscopieën die zij zouden krijgen gegeven de familiaire of erfelijke vorm van darmkanker waarmee zij zijn gediagnosticeerd.

Deel 2: De validiteit van de preferenties die worden gemeten met een DCE in de volksgezondheid

In het tweede deel van dit proefschrift beschrijven we hoe drie aspecten van de validiteit van de voorkeuren die zijn gemeten met een DCE op het terrein van volksgezondheid zijn getest. Vergeleken met meer alledaagse beslissingen, zijn beslissingen over deelname aan een preventief programma heel anders en waarschijnlijk complexer. Daarom kan het zo zijn dat sommige respondenten niet in staat zijn om een DCE over preventieve programma's in te vullen volgens de normatieve theoretische assumpties die ten grondslag liggen aan deze methode. In **Hoofdstuk 4** hebben we, door middel van 70 gestructureerde interviews, onderzocht in welke mate mensen de keuzesets in een DCE evalueren en beantwoorden volgens de onderliggende methodologische assumpties. We lieten zien dat de deelnemers actief betrokken waren en bereid waren om de benodigde antwoorden te geven. De meeste deelnemers begrepen de attributen en handelden in overeenstemming met het monotoniciteit- (de voorkeur geven aan betere ten opzichte van slechtere attribuutlevels) en het continuïteit axioma (een keuze maken op basis van de meerderheid van de attributen). Ongeveer een derde van de respondenten gebruikte gesimplificeerde beslisstrategieën en nam bijvoorbeeld een beslissing op basis van minder dan drie attributen of rapporteerde tegenovergestelde voorkeuren voor de attribuutlevels. Deze schendingen van de normatieve theoretische assumpties zagen we vooral onder ouderen, mensen met een lagere

opleiding en/of onvoldoende gezondheidsvaardigheden. Concluderend, de meerderheid van de deelnemers vult een DCE in zoals wordt verondersteld vanuit de methodologie. Echter, gebaseerd op de leeftijd, het opleidingsniveau en de gezondheidsvaardigheden van respondenten zullen extra maatregelen moeten worden genomen om ervoor te zorgen dan elke deelnemer de keuzesets begrijpt en de DCE invult zoals dat wordt verondersteld.

Dat sommige respondenten niet geheel in overeenstemming met alle normatieve theoretische assumpties handelen is niet meteen een grote zorg voor de algemene studie-uitkomsten, tenzij het beslisgedrag van respondenten zodanig wordt beïnvloed dat er verschillende voorkeuren worden gevonden tussen bepaalde subgroepen in de populatie. Dergelijke verschillen in voorkeuren kunnen dan ten onrechte worden toegeschreven aan heterogeniteit in preferenties. Uiteraard heeft dit grote gevolgen voor de validiteit van de gemeten voorkeuren. Om dit verder te bestuderen hebben wij in ons onderzoek verder gefocust op de rol van gezondheidsvaardigheden. Het onderzoek dat wordt beschreven in **Hoofdstuk 5** was onderdeel van de DCE over rotavirus vaccinatie zoals beschreven in **Hoofdstuk 2**. We lieten hier zien dat gezondheidsvaardigheden samenhingen met de preferenties van ouders voor rotavirus vaccinatie. De effectiviteit van het vaccin en de frequentie van ernstige bijwerkingen waren belangrijker en de beschermduur was minder belangrijk voor ouders met een hoge opleiding en goede gezondheidsvaardigheden vergeleken met ouders met een lage opleiding en minder goede gezondheidsvaardigheden. Deze verschillen in preferenties tussen mensen met goede en minder goede gezondheidsvaardigheden wordt wellicht niet veroorzaakt door daadwerkelijke verschillen in voorkeuren, maar mogelijk door verschillen in kennis. Het zal verder moeten worden onderzocht of de preferenties van mensen met goede en minder goede gezondheidsvaardigheden nog steeds verschillen wanneer respondenten de numerieke waarde van risico's hetzelfde interpreteren. Dergelijk onderzoek zal uitwijzen in welke mate een beslisproces wordt beïnvloed door het begrip van bepaalde (vaccin) kenmerken. De resultaten van deze studie benadrukken het belang van gezondheidsvaardigheden in studies naar (vaccinatie gerelateerd) beslisgedrag.

Wanneer respondenten programmakenmerken niet goed interpreteren zullen de gemeten voorkeuren de daadwerkelijke keuzes van mensen mogelijk niet goed voorspellen. Echter de mate waarin DCEs gedrag kunnen voorspellen is van maatschappelijk belang aangezien de resultaten van DCEs steeds vaker worden gebruikt als input voor beleid. In **Hoofdstuk 6** hebben we de voorspellende waarde van DCEs in de volksgezondheid bepaald door de voorkeuren van mensen in een DCE, te vergelijken met hun daadwerkelijke gedrag. Op populatieniveau was er een kleine maar niet significante onderschatting van daadwerkelijk gedrag. In 74,1% van de gevallen kwamen de voorspellingen op basis van de DCE overeen met het daadwerkelijke gedrag van de deelnemers, dit resulteerde in een positieve voorspellende waarde van 0,80 en een negatief voorspellende waarde van 0,40, een sensitiviteit van 0,90 en een specificiteit van 0,35. Hoewel gedrag voor een deel voorspeld kon worden op basis van de voorkeuren die waren gemeten met de DCE, waren niet alle voorspellingen correct. Het kwam vaker voor dat er werd voorspeld dat mensen mee zouden doen terwijl ze dat uiteindelijk niet deden dan andersom. Concluderend, preferenties gemeten met een DCE kunnen worden gebruikt om deelname aan een preventief programma te

voorspellen. Het is echter nog onduidelijk in welke mate DCEs ook geschikt zijn om te voorspellen wie er af zal zien van deelname. Toekomstig onderzoek is nodig om de voorspellende waarde van DCEs te testen onder verschillende patiënten groepen en binnen verschillende besliscontexten.

Deel 3: De invloed van de presentatie van keuzesets op het keuzegedrag van respondenten en de uitkomsten van een DCE

In elk DCE project moeten onderzoekers beslissen over de manier waarop ze de keuzesets van hun DCE aan respondenten presenteren. De mogelijke invloed van de presentatie van keuzesets op het beslisgedrag van respondenten was tot nu toe grotendeels onbekend en er was geen wetenschappelijk bewijs voor de mate waarin de presentatie van keuzesets de uitkomsten van een DCE beïnvloedt. In het derde deel van dit proefschrift wordt daarom aandacht besteed aan drie aspecten van de presentatie van keuzesets.

In **Hoofdstuk 7** hebben we bepaald in welke mate het includeren van een opt-out optie invloed heeft op het keuzegedrag van respondenten, de coëfficiënten van de attributen, het relatieve belang van de attributen en de bereidheid tot betalen onder respondenten. Het onderzoek dat wordt beschreven in dit hoofdstuk was onderdeel van de DCE over leefstijl programma's zoals beschreven in **Hoofdstuk 1**. Respondenten kozen vaker voor de opt-out optie wanneer ze hier meteen mee werden geconfronteerd in tegenstelling tot de mensen die eerst werden gedwongen een keuze te maken. Daarnaast kozen respondenten met een lagere opleiding vaker voor de opt-out. De coëfficiënten in de dataset met en zonder opt-out optie verschilden. Hoewel dezelfde attributen het meest en minst belangrijk waren was er wel een duidelijk verschil in de relatieve mate van belangrijkheid van de attributen. Dit zorgde voor een significant verschil in de bereidheid tot betalen onder respondenten voor een algemeen sport-schema (respectievelijk 44 en 19 euro per jaar). Concluderend, op basis van het leereffect dat we in deze studie zagen, zouden toekomstige DCEs het gebruik van meerdere voorbeeld keuzesets of een tweetraps-response design in overweging kunnen nemen. Verder onderzoek zal ook in kaart moeten brengen hoe keuzesets zo makkelijk mogelijk kunnen worden gepresenteerd. Op die manier kan het aantal mensen dat de opt-out optie kiest omdat ze de keuze te moeilijk vinden wellicht worden verkleind. Tot slot is er extra onderzoek nodig dat bij mensen nagaat waarom zij voor de opt-out optie kiezen.

In **Hoofdstuk 8** hebben we onderzocht of het presenteren van attribuutlevels in woorden of plaatjes ervoor zorgt dat er andere resultaten worden gevonden wat betreft de interpretatie van de attribuutlevels, het relatieve belang van de attributen en de geschatte deelnamebereidheid. Het onderzoek dat wordt beschreven in dit hoofdstuk was onderdeel van de DCE over rotavirus vaccinatie zoals beschreven in **Hoofdstuk 2**. Respondenten gaven aan dat ze het gebruik van woorden in plaats van plaatjes significant makkelijker en duidelijker vonden. Het gebruik van woorden leverde ook significant meer consistente antwoordpatronen op. Daarnaast interpreteerden respondenten de attribuutlevels significant beter wanneer deze in woorden waren omschreven. In de dataset over de DCE in woorden bleken de kosten het meest belangrijke attribuut terwijl de frequentie van ernstige bijwerkingen het meest belangrijk bleek in de dataset over de DCE in plaatjes. Hoewel

de berekende deelnamebereidheid op basis van beide datasets sterk gecorreleerd was (ICC=0,94), waren de cijfers op basis van de plaatjes DCE over het algemeen lager dan die op basis van de woorden DCE. Concluderend, resultaten verschillen wanneer woorden of plaatjes worden gebruikt voor het presenteren van attribuutlevels in de keuzesets van een DCE. Gezien het feit dat er op dit moment geen richtlijnen zijn wat betreft de presentatie van keuzesets, is er verder onderzoek nodig om richtlijnen te ontwikkelen over valide presentatiemethoden voor attribuutlevels in de keuzesets van een DCE. Tot die tijd wordt aanbevolen om niet enkel plaatjes te gebruiken voor het presenteren van attribuutlevels. Als onderzoekers toch besluiten om plaatjes te gebruiken zal de presentatie van de attribuutlevels moeten worden meegenomen in de focusgroep fase van de ontwikkeling van de DCE en zijn uitgebreide pilot-testen onder de doelpopulatie noodzakelijk.

In **Hoofdstuk 9** hebben we onderzocht of het positief of negatief verwoorden (framen) van een risicoattribuut invloed heeft op het beslisgedrag van respondenten en hun voorkeuren. Het onderzoek dat wordt beschreven in dit hoofdstuk was onderdeel van de DCE over genetische screening naar darmkanker zoals beschreven in **Hoofdstuk 3**. In de positief verwoorde DCE kozen significant meer mensen altijd voor het scenario met de hoogste overlevingskans vergeleken met het aantal mensen dat altijd koos voor het scenario met de laagste sterftekans in de negatief verwoorde DCE. Respondenten die de negatief verwoorde DCE hadden beantwoord kozen significant vaker voor de opt-out optie. In de positief verwoorde DCE vonden respondenten de overlevingskans het meest belangrijk terwijl respondenten in de negatief verwoorde DCE juist het aantal vervolg-coloscopieën het meest belangrijk vonden. Concluderend, het verwoorden of framen van een risicoattribuut beïnvloedt de waardering van dat risico door respondenten en daarmee de uitkomsten van een DCE. Positieve verwoording leidt tot verhoogd dominant beslisgedrag terwijl negatieve verwoording leidt tot verhoogd risico-zoekend gedrag. Er is verder onderzoek nodig om richtlijnen te kunnen ontwikkelen met betrekking tot de verwoording van risicoattributen en de effecten daarvan op beslisgedrag en de uitkomsten van een DCE. Tot die tijd moet er voorzichtig worden omgegaan met het verwoorden van risico's. Hoe risico's worden verwoord moet een duidelijke rol krijgen in de expertinterviews en de focusgroep interviews die voorafgaand aan een DCE worden gehouden om de attributen en de attribuutlevels voor de DCE te selecteren. Bij voorkeur worden er zelfs verschillend verwoorde risico's gebruikt in de pilot versie van de DCE of in de daadwerkelijk DCE, om onderzoekers in staat te stellen om de omvang van het effect van framing in kaart te brengen.

Algemene discussie en conclusie

Tijdens de verschillende toepassingen van de DCE methode in de volksgezondheid konden alle uitkomstmaten worden bepaald. In tegenstelling tot klassieke Likertschaal vragenlijsten konden wij de voorkeuren van mensen kwantificeren door het berekenen van het relatieve belang van attributen, de afwegingen tussen attributen en de bereidheid tot betalen. Daarnaast konden we de potentiele deelnamebereidheid voor verschillende hypothetische preventieve programma's schatten. Die resultaten leidden tot specifieke aanbevelingen voor programmaontwikkelaars en beleidsmedewerkers om op die manier bij te dragen aan de ontwikkeling, de implementatie en de marketing van effectieve preventieve interventies. De grote meerderheid van de populatie

bleek goed in staat om een volksgezondheid-gerelateerde DCE in te vullen en de resultaten van een dergelijke DCE konden worden gebruikt om deelname aan een preventief programma te voorspellen.

Er zijn echter enkele discussiepunten wat betreft DCEs in de volksgezondheid die nog aandacht verdienen. Allereerst krijgen de respondenten van een DCE hypothetische scenario's voorgelegd in een sterk gecontroleerde context. Daarom hebben de keuze context, hypothese bias en psychosociale factoren misschien een invloed op het beslisgedrag van respondenten en verkleinen ze op die manier de vergelijkbaarheid tussen de antwoorden in de DCE en daadwerkelijk gedrag. Het is tot nu toe onduidelijk of en hoe deze zaken invloed hebben op beslisgedrag van respondenten, maar er bestaan wel verschillende manieren om ermee om te gaan. 'Cheap-talk' of respondenten laten beslissen onder ede kunnen het effect van hypothese bias verkleinen en het analyseren van de data aan de hand van hybride modellen kan de invloed van psychosociale factoren in kaart brengen. De 'Time-to-Think' methode kan worden gebruikt om te controleren voor zowel hypothese bias als de invloed van psychosociale factoren. Al deze methoden moeten uitgebreid getest en gevalideerd worden op het terrein van volksgezondheid. Ten tweede, sommige subgroepen, waaronder ouderen hadden moeite met het invullen van een DCE volgens de normatieve theoretische assumpties. Dit is vooral interessant gezien het feit dat de algemene populatie steeds ouder wordt. De toenemende hoeveelheid ouderen en de vatbaarheid van ouderen voor verschillende (chronische) aandoeningen maakt hen een geschikte doelgroep voor verschillende preventieve interventies. Het is echter onduidelijk of DCEs, op de manier waarop ze nu worden uitgevoerd, de preferenties in deze populatie kunnen meten en daarmee een bijdrage kunnen leveren aan het ontwikkelen van effectieve en aantrekkelijke interventies. In welke mate alternatieve meetmethoden, zoals interviews of online vragenlijsten, kunnen bijdragen aan een verbetering van de capaciteit van ouderen en respondenten met een lage opleiding en/of minder goede gezondheidsvaardigheden om een DCE in te vullen, moet worden getest. Ten derde, de manier waarop keuzesets aan respondenten werden gepresenteerd beïnvloede hun beslisgedrag en de uitkomsten van de DCE. Naast een uitgebreid onderzoek naar de invloed van de opt-out optie, het gebruik van plaatjes en de verwoording (framing) van risicoattributen, zouden ook andere vormen van presentatie (bijvoorbeeld kleurenschema's, doel framing, referentiepunt bias, tijdsafhankelijkheid) moeten worden onderzocht. Van de resultaten van dergelijk onderzoek kan worden geleerd, kan de DCE methode worden aangepast en kunnen huidige richtlijnen voor 'best-practice' DCEs worden aangevuld. Dit alles zal bijdragen aan de kwaliteit van de data en de validiteit van de resultaten uit een DCE. Dit is van cruciaal belang omdat een slechte datakwaliteit niet kan worden gecompenseerd door geavanceerde designs of state-of-the-art statistische analysetechnieken. Verbeterde transparantie betreffende de presentatie van keuzesets in artikelen over DCEs en de aanpassing van DCE 'best-pratice' richtlijnen zullen ervoor zorgen dat de DCE methode valide en kwalitatief goed blijft voor wat betreft het meten van preferenties, ook in de volksgezondheid.

Concluderend, DCEs kunnen worden gebruikt voor het meten van preferenties voor volksgezondheid-gerelateerde interventies en om aanbevelingen te doen over programmaontwikkeling die bij kunnen dragen aan het verhogen van de deelnamebereidheid. Er zijn verschillende opties om deze methode verder te verbeteren specifiek op het terrein van de volksgezondheid. Door het combineren van verder onderzoek, kennis uit andere wetenschappelijke disciplines, goed ontworpen experimenten en geavanceerde analytische modellen zal onze capaciteit om preferenties correct te meten en beslisgedrag van mensen te begrijpen verbeteren. Daarmee kan de DCE methode een bijdrage leveren aan de verbetering van de algehele volksgezondheid.

DANKWOORD

Vier jaar geleden had ik enorm veel zin om aan dit project te beginnen en de jaren lijken voorbij te zijn gevlogen. Nu is het zover, mijn proefschrift is af en ik kijk met heel veel plezier terug op de afgelopen jaren vol uitdagingen en de daarbij behorende leerzame, leuke en interessante momenten!

Het onderzoek dat ik heb gedaan en de bijbehorende artikelen die ik heb geschreven zijn zeker niet alleen mijn verdienste. Gelukkig had ik een grote groep van fijne mensen om me heen die me hierbij heeft geholpen. Ik wil dan ook graag iedereen bedanken die op enige manier heeft bijgedragen aan de totstandkoming van dit proefschrift.

Allereerst zijn er natuurlijk mijn promotor en copromotoren. We zijn met zijn vieren aan een project over DCEs begonnen en eigenlijk had niemand van ons extreem veel expertise op dat terrein. Dit heeft er echter niet voor gezorgd dat we minder gemotiveerd waren, maar maakte het project juist tot een uitdaging voor ons allemaal. Hoewel we door de jaren heen tegen verschillende 'problemen' aan zijn gelopen, wisten we voor alles een oplossing te vinden.

Beste Jet, bedankt dat je je vanaf het eerste moment vol vertrouwen en goede moed in mijn project hebt gestort. Je enthousiasme tijdens onze afspraken gaf mij steeds opnieuw weer een motivatie-boost. Daarnaast heb jij de artikelen begrijpelijk weten te houden en telkens de rode lijn van het verhaal naar voren gehaald wanneer ik weer eens verdwaalde in 'jargonesiteit' en details; bedankt voor al je tips & tricks!

Beste Ardine, allereerst bedankt voor al je vertrouwen. Tijdens het gehele project heb je me heel vrij gelaten over de invulling van verschillende deelprojecten en gaf je me de ruimte om mijn eigen ideeën uit te werken. Het behoeft denk ik geen verdere uitleg om te begrijpen dat dat enorm motiveert. Je kritische houding, je relaxte instelling en je pragmatische aanpak waardeer ik enorm en maakte het heel prettig om met je samen te werken. Heel erg bedankt voor alle kansen die je me hebt gegeven!

Beste Mattijs, de moeilijke en ingewikkelde situaties vond je het meest interessant, een goede discussie ging je nooit uit de weg en brainstormen bleek jouw expertise. Dat alles heeft me gedurende dit project scherp gehouden en menig discussiepunt opgelost. Bedankt daarvoor, maar zeker ook voor de tijd die je de afgelopen jaren voor me had. Te allen tijde kon ik bij je terecht voor vragen en kon ik je teksten sturen waar ik nog onzeker over was. Je inzet en bijna onuitputtelijke bron van ideeën zijn bewonderenswaardig.

Alle coauteurs van de artikelen uit dit proefschrift wil ik uiteraard ook graag bedanken. Zonder jullie inhoudelijke kennis over leefstijlinterventies, diabetes, rotavirus infecties, vaccinatie gedrag, darmkanker, genetische screening en health literacy had ik deze artikelen nooit kunnen schrijven.

Paul van Gils, Jeroen Struijs, Karolien van de Brekel-Dijkstra, Rabin Neslo, Patricia Bruijning-Verhagen, Frank Kallenberg, Evelien Dekker, Annelien Bredenoord, Henk van Kranen, Iris van der Heide, Ellen Uiters, Jany Rademakers en Jantine Schuit, bedankt. Ook op het gebied van DCEs heb ik de nodige hulp gehad, vooral bij het schrijven van de methodologische artikelen. Janine van Til, Brigitte Essers, Carmen Dirksen en Ida Korfage bedankt voor de fijne samenwerking! Esther de Bekker-Grob, dank dat je me zo vroeg in mijn promotie-traject 'onder je vleugels' hebt genomen, me wegwijs hebt gemaakt in het DCE onderzoek en ook later in mijn project altijd mijn vragen hebt beantwoord. Karin Groothuis-Oudshoorn dank voor je hulp bij mijn 'data crisis'. Super dat ik toen, maar ook daarna, bij je terecht kon met mijn statistische vragen.

Hanneke Wanders en Benjamin Salampessy, ook jullie wil ik graag bedanken voor de fijne samenwerking. Als master studenten begonnen jullie aan één van mijn projecten en jullie inzet heeft niet alleen geresulteerd in jullie afstuderen, maar ook in twee mooie publicaties!

Mijn collega's van de afdeling HTA op het Julius Centrum en VPZ & KZG op het RIVM wil ik ook graag bedanken. Ik heb het werken op deze beide locaties altijd als een voordeel gezien, vooral omdat ik zoveel collega's had en dat is top als het zulke fijne mensen zijn! Bedankt voor jullie interesse, ondersteuning, samenwerking en collegialiteit de afgelopen jaren.

Een special bedankje gaat uit naar mijn Julius collega's en mijn kamergenoten uit 6.118 (later 5.08): Anoukh, Maarten, Giske, Mirjam, Judith, Mart, Marloes, Christiana, Carla en Marieke, maar zeker ook naar mijn (oud)RIVM collega's: Jeroen, Ellen, Eline, Claudia, Marga, Iris, Arthur, Saskia, Gerben, Nina, Tessa, Laura en Koen voor de gezellige lunches, wandelingen, koffiemomenten en congres bezoeken. Tot slot wil ik Domino graag bedanken, het was heel prettig om na een jaar eindelijk een 'DCE-mattie' te hebben. Bedankt voor de fijne samenwerking (dat ging als vanzelf), de gezelligheid, en natuurlijk voor de oprichting van onze DCE-Weetjes ☺. Super dat je ook mijn paranimf wilt zijn!

Mijn vrienden wil ik graag bedanken, gewoon omdat ze er altijd en voor alles zijn! Jorien & Elwin, jullie verdienen een speciaal plaatsje in mijn hart, jullie zijn geweldig! Suus & Roy, Marjo, Sanne, Inge, Jolien & Erik, Frans & Elly bedankt. In het bijzonder wil ik ook Malou bedanken, wat fijn om jou als vriendin te hebben en super dat je ook mijn paranimf wilt zijn!

Dan is er nog mijn lieve familie en schoonfamilie. Bedankt voor alle gezellig momenten, jullie interesse en support. Papa & Resy, Mama & Wim, dank voor de steun die ik altijd en onvoorwaardelijk van jullie krijg. Niet alleen voor het schrijven van dit proefschrift, maar vooral op de momenten die er echt toe doen. Mijn oma's wil ik ook graag bedanken voor hun eindeloze interesse en betrokkenheid. Lieve Rick, er is teveel om jou voor te bedanken. Uiteindelijk is het belangrijkste dat wij elkaar 'echt' hebben ontmoet, dat zegt toch genoeg ☺

Lieve Devin, jouw onstuitbare enthousiasme, optimisme en vrolijkheid zijn een ware bron van energie en inspiratie. Lieve Kyan, jouw enorme interesse en je bijzonder pientere instelling houden me scherp. Blijf beiden wie je bent, want jullie zijn geweldig!

Lieve Gerard, Geertje, bevlogenheid en passie zijn jouw mantra en die straal je uit. Passie heb jij niet alleen in overvloed voor je werk, maar zeker ook voor alles daarom heen, ik ben dankbaar dat ik daar onderdeel van mag zijn. Met jou is het leven zoals het hoort te zijn, elke dag opnieuw!

Jorien

PUBLICATIONS

PUBLICATIONS AS PART OF THIS THESIS

Veldwijk, J. M.S. Lambooij, P.F. van Gils, J.N. Struijs, H.A. Smit and G.A. de Wit, *Type 2 diabetes patients' preferences and willingness to pay for lifestyle programs: a discrete choice experiment.* BMC: Public Health, 2013, 13: p. 1099.

Veldwijk, J., M.S. Lambooij, P. Bruijning-Verhagen, H.A. Smit and G.A. de Wit, *Parental preferences for rotavirus vaccination in young children: a discrete choice experiment.* Vaccine. 2014, 32: p. 6277-83.

Veldwijk, J., M.S. Lambooij, E.W. de Bekker-Grob, H.A. Smit and G.A. de Wit, *The effect of including an opt-out option in discrete choice experiments.* PlosOne, 2014, 9: e111805.

Salampessy B.H. **J. Veldwijk**, A.J. Schuit, K. van den Brekel-Dijkstra, R.E.J. Neslo, G.A. de Wit and M.S. Lambooij, *The predictive value of discrete choice experiments in public health: An explorative application.* Patient, 2015, In press.

OTHER PUBLICATIONS

Veldwijk, J., S. Scholtens, G. Hornstra and W.J.E. Bemelmans, *Body Mass Index and cognitive ability of young children.* Obesity Facts, 2011, 4: p. 264–69.

Veldwijk, J., M.C.E. Fries, A. Haveman-Nies, W.J.E. Bemelmans, H.A. Smit and A.H. Wijga, *Overweight and school performance among primary school children: the PIAMA birth cohort study.* Obesity, 2012, 20: p. 590-96.

Harmsen, I.A, M.S Lambooij, R.A.C. Ruiter, L. Mollema, **J. Veldwijk**, Y. van Weert, G. Kok, T. Paulussen, G.A. de Wit and H. de Melker, *Psychosocial correlates of parents' intention to vaccinate their newborn child against hepatitis B in the Netherlands.* Vaccine, 2012, 30: p. 4771-7.

Veldwijk, J., C. Hoving, B.M. van Gelder and T.L. Feenstra, *Potential reach of effective smoking prevention programs in Dutch vocational schools: determinants of school director's intention to adopt.* Public Health, 2012, 126: p. 338-42.

Veldwijk, J., K.I. Proper, H. Hoeven-Mulder and W.J.E. Bemelmans, *The prevalence of physical, sexual and mental abuse among overweight adolescents.* BMC: Public Health, 2012, 4: p. 840.

Barte, J.C.M., **J. Veldwijk**, P.J. Teixeira, F.M. Sacks and W.J.E. Bemelmans, *Differences in weight loss across different BMI classes: a meta-analysis of the effects of interventions with diet and exercise.* Int J Behav Med. 2014, 21: p. 784-93.

Struik, M.H., F. Koster, A.J. Schuit, R. Nugteren, **J. Veldwijk** and M.S. Lambooij, *The preferences of users of electronic medical records in hospitals: quantifying the relative importance of barriers and facilitators of an innovation*. Implement Sci, 2014, 9: p. 69.

Wanders, J.O.P., **J. Veldwijk**, G.A. de Wit, B.E. Hart, P.F. van Gils and M.S. Lambooij, *The effect of positive and negative financial incentives in a discrete choice experiment: an application to lifestyle programs.* BMC: Public Health, 2014, 14: p. 870.

Lambooij, M.S., I.A. Harmsen, **J. Veldwijk**, H. de Melker, L. Mollema, J.W.M. van Weert and G.A. de Wit, *Linking stated and revealed preferences: combining a Discrete Choice Experiment and a behavioral experiment on vaccination behavior.* BMC: Med Res Methodol, 2015, In Press.

ABOUT THE AUTHOR

Jorien Veldwijk was born on the 7th of January 1987 in Heino, the Netherlands. After completing her secondary school education at the Carmel College Salland in 2005, she studied Health Sciences at Maastricht University and obtained her Bachelor of Science degree in 2008. She started her Master of Health Sciences with a major in Health Education and Promotion at Maastricht University, during which she followed an internship at the National Institute for Public Health and the Environment. During this internship, she examined the potential reach of effective smoking prevention programs in Dutch vocational schools. She graduated and obtained her Master of Science degree in 2009. Between 2009 and 2011, she worked as a junior researcher at the National Institute for Public Health and the Environment, where her research mainly focused on obesity among children and adolescents. In March 2011, she started her PhD-project at the Julius Center for Health Sciences and Primary Care in collaboration with the National Institute for Public Health and the Environment. She completed her PhD-project under the supervision of Prof. dr. Henriëtte (Jet) Smit, Dr. Ardine de Wit and Dr. Mattijs Lambooij. Parallel to her PhD-project she followed a postgraduate Master program Clinical Epidemiology at the University Medical Center Utrecht and graduated in 2013.