Valuing and Refining Outcome Measures for Economic Evaluations in Health Care

Liesbet Lawerman–van de Wetering

Valuing and Refining Outcome Measures for Economic Evaluations in Health Care

Liesbet Lawerman-van de Wetering

ISBN: 978-94-6169-846-9

Layout and printing: Optima Grafische Communicatie, Rotterdam, The Netherlands

Valuing and Refining Outcome Measures for Economic Evaluations in Health Care

Het waarderen en verfijnen van uitkomstmaten voor economische evaluaties in de gezondheidszorg

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof.dr. H.A.P. Pols

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op

donderdag 12 mei 2016 om 11:30 uur

door

Elisabeth Josine Lawerman-van de Wetering
geboren te 's-Gravenhage

Erasmus University Rotterdam

PROMOTIECOMMISSIE

Promotor
Prof.dr. W.B.F. Brouwer

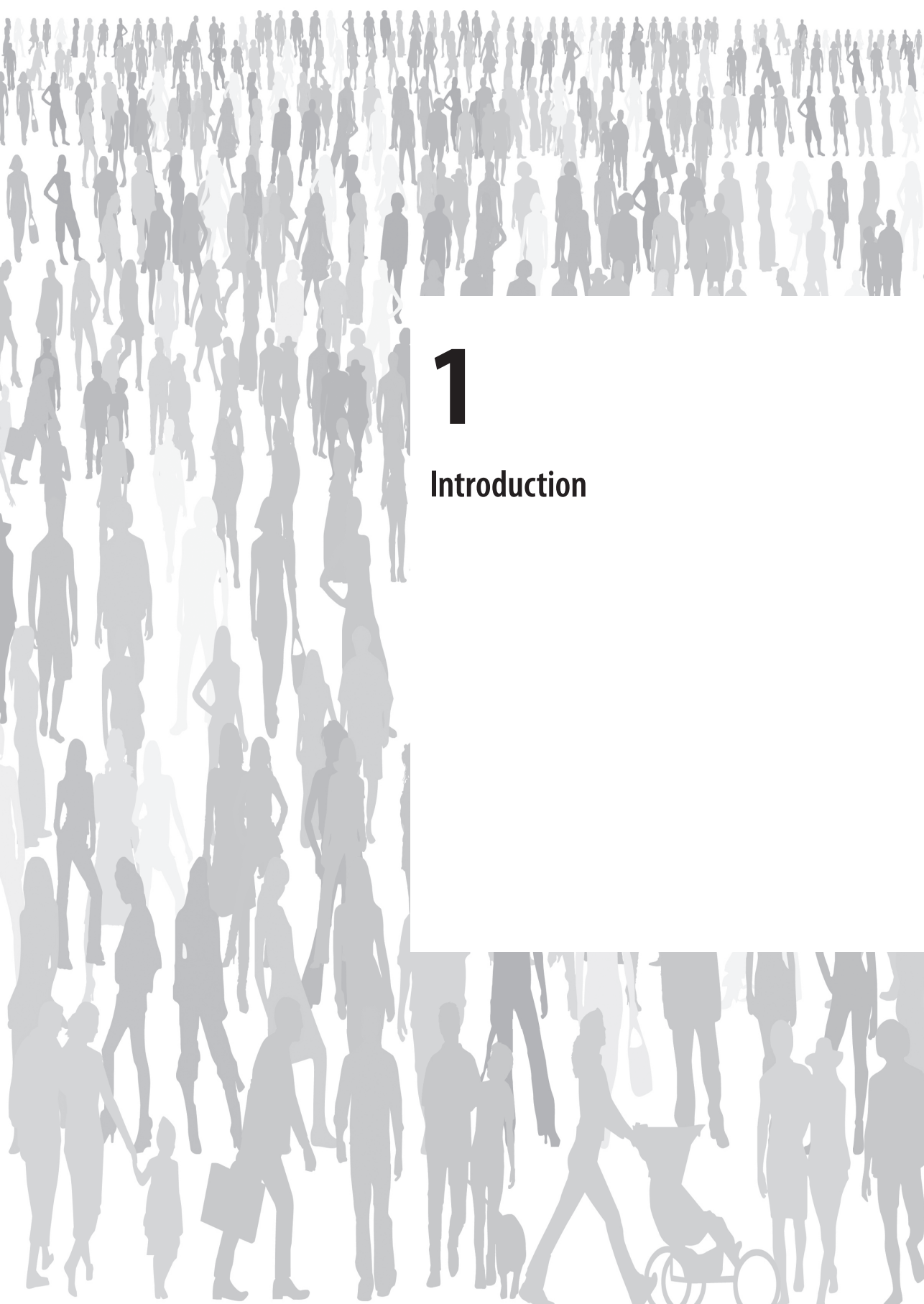
Overige leden
Prof.dr. M.P.M.H. Rutten-van Mölken
Prof.dr. C.D. Dirksen
Prof.dr. J.J. van Busschbach

Copromotor
Dr. N.J.A. van Exel

CONTENTS

TABLES

FIGURES



1

Introduction

1.1. ALLOCATING SCARCE RESOURCES IN HEALTH CARE

Due to new expensive technologies and aging populations, the pressure on health care budgets is increasing rapidly. To ensure an efficient use of the limited health care resources priorities must be set and choices regarding access to health care become inevitable [1]. These choices are a sensitive topic in society since their outcomes necessarily imply that some health care interventions will not be funded, withholding these interventions from patients in need. At this moment most allocation decisions are still rather implicit. However, the growing pressure on health care budgets and the social impact of allocation decisions ask for a more consistent and transparent decision making process, if possible based on explicit decision rules. In that context, economic evaluations, which provide insight into the costs and (health) benefits of new technologies, are increasingly used to inform decisions regarding the allocation of scarce health care resources [2, 3]. The use of economic evaluations should contribute to a more systematic and transparent decision making process with the aim to stimulate overall health and welfare [4].

In order to contribute to these goals, economic evaluations need to be methodologically sound. Although the additional value of economic evaluations seems to be generally accepted there are still many methodological challenges to deal with. This thesis focusses on two important methodological issues related to measuring and valuing the benefits of new interventions. The first relates to the differentiation of the (monetary) value of a QALY according to distributional preferences in society. The second issue concerns the scope of the outcome measure, since happiness as a broader outcome measure than health may be better capable to capture all benefits of an intervention. Before addressing these issues further, economic evaluations are introduced.

1.2. ECONOMIC EVALUATIONS IN HEALTH CARE

Economic evaluations are rooted in welfare economics and traditionally take the form of a cost-benefit analysis. In such an analysis, two treatment strategies are compared in terms of their costs and benefits. The costs and benefits are both expressed in monetary terms which makes the decision rule fairly straightforward: reimbursing the technology is considered welfare improving if the additional benefits of the new technology outweigh its additional costs, compared to the old technology.

In health care, cost-benefit analyses are less common, primarily because expressing the value of health gains in monetary terms is difficult and contentious [5]. Hence, economic evaluations in health care often take the form of cost-effectiveness analysis (CEA) or

cost-utility analysis (CUA). These differ from cost-benefit analysis in the way they express health benefits. In CEA health benefits are expressed in relevant 'natural units', for instance life-years gained, hip fractures avoided or percentage blood pressure reduced. While this relates well to clinical practice, decision making at societal level is hampered by the incomparability of outcomes between different diseases, also in terms of their value. CUA was developed to overcome this problem. In a CUA the health benefits of a technology are typically expressed in Quality-Adjusted Life-Years (QALYs). This is a preference based measure that combines length and quality of life by weighing each year for the quality of life during that year, with 1 representing perfect health and 0 represents the state 'dead'. One perfectly healthy life-year counts as 1 QALY. Imperfect health states receive a weight between 0 and 1, although states worse than dead (i.e., receiving negative weights) are possible and observed [6].

The outcome of a CUA is an incremental cost effectiveness ratio (ICER) and reflects the additional cost per gained QALY of a technology compared to the best alternative [1, 7]. As in cost-benefit analysis, a technology can be considered welfare improving when the value of the additional health gain in QALYs (i.e. the benefits) exceeds the additional cost of the new technology compared to the old one, as represented by the following equation [7]:

Eq. 1.1 $V_{QALY} * \Delta QALY - \Delta costs > 0$

In this equation V_{QALY} is the monetary value of a QALY, ΔQALY denotes the incremental QALY gain and Δcosts represent the associated incremental costs. The health benefits of the technologies are thus reflected by the term $V_{QALY} * \Delta QALY$.

The equation can be reformulated to demonstrate the common decision rule of economic evaluations:

Eq. 1.2 $$\frac{\Delta \text{ costs}}{\Delta \text{ QALY}} < V_{QALY}$$

This equation clearly shows that the ICER (i.e. Δcosts/ΔQALY), 'the price of a QALY', should be lower than the value of that QALY (V_{QALY}) to be welfare improving. Here, the V_{QALY} represents the social value of QALY, what the society is willing to pay to gain a QALY. It should be emphasized that without an estimate of this social value of a QALY, an economic evaluation will not be complete. A threshold value is necessary to judge whether the technology can be considered good value for money and should be reimbursed. At this moment, the threshold to be used in decision making is often rather implicit.

Regardless of the exact height of the threshold, an important question is whether this threshold should be equal for all health gains. Often, it is (implicitly) assumed that all health gains are valued equally, so that one fixed threshold, or V_{QALY}, can be used for all interventions. This assumption is based on the well-known rule that 'a QALY is a QALY is a QALY' [6] which implies that specific characteristics of, for instance, the recipient (e.g. age) or the disease (e.g. severity) are not considered in traditional economic evaluations [8]. Hence, priority will be given to interventions that generate most health given a unit investment [4], regardless of the 'context' of these health gains.

However, with the increasing use of economic evaluations to inform health care allocation decisions, the discrepancies between recommendations based on economic evaluations and actual or socially desired decisions become more evident [4, 9]. The results of traditional economic evaluations do not (always) align with societal preferences with respect to the distribution of scarce health care resources. This may partly explain the still relatively limited role of economic evaluations in health care allocation decisions [9]. In order to become more useful for policy making, somehow the outcomes of economic evaluations have to reconcile with existing societal preferences for the distribution of health and health care.

As indicated, this thesis deals with two distinct issues related to the benefit side of economic evaluations. The first is directly related to V_{QALY} in Equation 1.2, the monetary value of a QALY. This value is commonly considered to be equal for all QALY gains. However, an increasing body of research suggests that V_{QALY} should differ according to distributional preferences in society.

The second issue concerns the ΔQALY in Equation 1.2, which represents the benefits of an intervention expressed in QALYs. It has been increasingly recognized that the narrow focus on health may not (always) capture all benefits of an intervention [10-12]. Some health care interventions, e.g. in long term care, do not necessarily aim to improve health but rather wellbeing. Therefore, adequate measurement of benefits requires outcome measures capturing outcomes beyond the QALY. This may for instance be done through measuring happiness. In this thesis we explore the support among the general public for happiness as an outcome measure in economic evaluation and the relationship between health and happiness. In the following sections, the two methodological issues addressed in this thesis will be further introduced. Addressing these issues will contribute to the refinement of the methodology and understanding of economic evaluations in health care and reconciliation of recommendations based on economic evaluations with societal preferences for the distributions of health and health care.

1.3. DISTRIBUTIONAL CONCERNS IN ECONOMIC EVALUATIONS

There is increasing recognition that, contrary to conventional economic evaluations, society seems to differentiate between QALY gains. This differentiation appears to be partly related to the wish to come to a more equitable distribution of health and health care [13]. Society may, for instance, prefer to treat relatively severely ill patients even when the potential QALY gains in these patients are relatively small [2, 14]. These equity considerations can be incorporated more systematically in health care allocation decisions by weighting QALY gains according to some equity principle. After weighting them, they can be compared to the relevant, fixed threshold value.

The same differentiation can also be achieved by using a flexible threshold. Then, QALY gains are not weighted, but ICERs are judged against a flexible threshold, which varies with the equity characteristics of the QALY gains. For instance, it could be higher for younger patients or more severely ill patients. This makes it explicit that some QALY gains carry more social value than others. In principle, both procedures are equivalent. The decision rule of an economic evaluation than becomes:

$$\text{Eq. 1.3} \quad \frac{\Delta \text{ costs}}{\Delta \text{ QALY}_i} < V_{QALYi}$$

Here, the *i* reflects the specific equity characteristic of the QALY gain and V_{QALYi} the social willingness to pay (WTP) for this QALY.

In deriving relative social values for QALY gains a first important and normative choice is which (combination of) equity principle(s) is considered an appropriate base for QALY differentiation. Two principles are often proposed in the literature: fair innings and severity of illness. Roughly speaking, fair innings strives to equalize lifetime health [15], while severity of illness severity of illness focusses on people's health expectations, irrespective of already experienced health [16]. Some support has been found for both of these principles, which may imply that people to some extent adhere to both [2, 14, 17, 18]. Therefore, the Netherlands proposed the equity principle "proportional shortfall". This principle is based on the proportion of remaining lifetime health lost due to disease and could therefore also be seen as an intermediate principle between fair innings and severity of illness [19-21]. This concept may be viewed as a first, pragmatic attempt to operationalize an equity principle which could be used in the allocation of scarce health care resources. Therefore, it can be interesting for other countries wishing to balance efficiency and equity considerations in health care allocation decisions.

Besides learning from the first experiences of the operationalization of an equity concept in practice, empirical studies can be designed to evaluate which principle better

reflects the distributional preferences in society by eliciting relative QALY weights. These empirical studies may also provide important information to better understand societal preferences and thus help to refine the decision making process. For instance, besides severity and age, other characteristics may be considered relevant by the public in health care priority setting, such as culpability or rarity of the disease.

For health care allocation decisions, QALY weights do not provide full information. They must be accompanied by knowledge on the value of the threshold, i.e., the value of a QALY. To be precise, in using equity weights, at least one reference case is needed for which both the equity weight and the social value is known. When using a flexible threshold, the different monetary valuations of QALY gains need to be elicited in different equity contexts. So far, most studies that explored public preferences for QALY gains in relation to a variety of equity principles and characteristics of the beneficiary or the disease have not addressed the monetary valuation [22]. Studies aimed to derive the WTP per QALY normally did not take a societal viewpoint in eliciting values and (therefore) did not differentiate according to the equity context [23]. In this thesis, such estimates of the social WTP for QALY gains in different equity subgroups are presented. The experiments were specifically designed to learn how respondents solve the dilemmas they are confronted with, and to better understand societal support for differentiating QALY values between groups of patients.

It should be noted that in light of these societal preferences, the decision framework and policy instruments also require attention. A direct reason for this is the increasing use of "conditional reimbursement" as a policy instrument. Since allocation decisions usually have to be made close to the launch of a new treatment, the available evidence to inform health care decision makers is often insufficient. However, denying access to the market on those grounds would delay the availability of new interventions undesirably. Conditional reimbursement allows to include a new treatment in the basic benefits package for a given period of time until more scientific evidence about (cost) effectiveness becomes available [24, 25]. After that, a permanent decision regarding reimbursement is made. Importantly, the feasibility of ending temporary reimbursement may be reduced by the fact that ending prior reimbursement resembles a loss, which normally carries more weight than an equally sized gain [26]. Loss aversion may thus make it more difficult to take something out of the basic benefits package than not adding the same intervention to the package in the first place. In other words, people may attach more value to interventions already in the package than those not yet reimbursed, ceteris paribus.

It is important to stress here that the normative questions about which equity principle should justify and underlie QALY differentiation, and whether or not prior reimburse-

ment status *should* matter in societal decisions, may not necessarily be answered best (or even appropriately) through positive research. In other words, not all preferences of the general public in the context of an 'equitable' distribution of health care, may be considered equitable (or desirable) from a normative viewpoint.

1.4. BROADER OUTCOME MEASURE

The second issue addressed in this thesis concerns the outcome measure of economic evaluations, in Equation 1.2 represented by ΔQALY. Traditionally, the analysis of benefits of an intervention solely focused on health effects, often expressed in QALYs, reflecting *health-related utility*. However, this narrow focus on health may not always capture the full benefits of an intervention. For example in long term care, interventions are not always (primarily) aimed at improving 'health' but also at contributing to other aspects of general wellbeing, such as autonomy [4]. Therefore, whether broader outcome measures capturing general wellbeing would be more appropriate to capture outcomes beyond the QALY, is a rising issue of debate [27].

Happiness measurement is an interesting and popular candidate in this context [28]. The interest in happiness as an outcome measure for economic evaluations is in line with the growing interest, also in economics, in happiness research. Like in the early days of economics, happiness can be interpreted as a proxy for utility or overall welfare [5]. Happiness then represents the ultimate goal in life and hence becomes an interesting outcome measure of social welfare. Interventions, also in health care, could then be assessed in terms of their effects on happiness of individuals [29, 30].

In the literature, the concepts 'happiness', 'well-being' and 'life-satisfaction' are often used interchangeably [31]. In this thesis we will define happiness as the degree to which an individual judges the overall quality of his life favourably [31]. Since happiness is a subjective measure, there have been many discussions in the literature about its measurability. At this moment, there seems to be consensus for quantifying happiness by questions like: 'On a scale from 0 to 10 how satisfied are you with your life?' where 0 represents completely dissatisfied and 10 completely satisfied. Research has shown the validity of using self-reports and a cardinal interpretation of happiness measures [32].

Within health care, the outcome measure happiness may more fully capture the overall welfare effects on individuals. This implies that happiness may provide additional information about the value of interventions that is relevant to evaluate those interventions and to inform health care allocation decisions. However, at present it is unclear how information on changes in happiness should be incorporated in such allocation decisions.

It also raises new questions, such as: Who should be the source of happiness changes? Should happiness replace the QALY outcome or complement it? If both measures are considered, how does that influence the allocation decision? How can health and happiness measures be traded off?

Given the increasing interest in happiness as an outcome measure in health care it is also important to improve the understanding of what constitutes happiness and in particular about the relationship between health and happiness. Happiness is a multidimensional concept and its determinants seems to differ across individuals, life stages, and life domains [33-35]. More detailed information of what constitutes happiness contributes to the understanding of the role of health and health care in the achievement of overall happiness.

Of particular importance in the context of this thesis is the nature of the relationship between health and happiness. Most studies use subjective measures of health and happiness which makes it difficult to disentangle the association between the two. Are healthy people happier or is it the other way around? Reverse causality as well as correlations of subjective health and happiness with other unobserved factors, such as personality, may bias the estimations of the effects of health on happiness [36, 37]. For instance, a relatively unhappy individual may experience and value her health worse than her true health status. Therefore, more knowledge on the relationship between health and happiness remains warranted.

1.5. OBJECTIVE

The overall aim of this thesis is to contribute to the methodology of outcome measurement in economic evaluations in health care. This thesis focusses on two distinct issues and is structured in two sections. Part A investigates the distributional concerns of the general public in the context of health care allocation decisions. Part B focuses on happiness as a broader outcome measure for economic evaluations in health care. Addressing these two issues will contribute to improvement of the methodology of economic evaluations in health care, reconciliation of recommendations based on economic evaluations with societal preferences for the distributions of health and healthcare, and ultimately to a more influential role of economic evaluations in health care decision making.

To contribute to the overall aim of this thesis the following research questions are formulated:

Part A

a) How is the equity efficiency trade-off operationalized in the decision making framework in the Netherlands?
b) What is the effect of providing information about severity of illness and fair innings on health care allocation decisions?
c) What is the social WTP for a QALY gain at different levels of proportional shortfall and in different age groups?
d) Does the social WTP per QALY depend on severity of illness and age characteristics of beneficiaries?
e) How does the current reimbursement status of an intervention affect the relative social value of a QALY?

Part B

f) Do respondents focus on health, happiness or both in prioritizing patients in health care?
g) What are the determinants of happiness and the role of health therein?
h) Is there a causal effect of health on happiness?

1.6. OUTLINE

Part A of this thesis addresses the issue of incorporating distributional concerns in economic evaluations. In chapter 2 of this thesis we evaluate how efficiency and equity considerations are combined in the Dutch decision making framework for health care allocation decisions. The operationalization of proportional shortfall appears to be a complex process with many methodological and normative questions. This analysis is directly relevant for the Dutch policy context. However, given the universality of some of the issues addressed, these analysis may also be relevant for other countries currently struggling with the formalization of equity concerns for priority setting in healthcare. Chapter 3 investigates public preferences with respect to severity of illness, fair innings and other characteristics of the beneficiaries (i.e. culpability, rarity of the disease and having dependents). Relative weights are obtained using a discrete choice experiment. Chapter 4 and 5 present the results of two studies deriving the social willingness to pay for a QALY considering different equity concepts. Chapter 4 uses a discrete choice experiment while chapter 5 is based on contingent valuation. Chapter 6 investigates the effect of the current reimbursement status of an intervention on the relative social value of a QALY. Additional characteristics potentially relevant in the decision making process are also considered.

Part B addresses the scope of the outcome measures in economic evaluations. Chapter 7 explores whether the general public considers happiness in addition to health in health care allocations decisions and how they trade-off health and happiness gains in treatment allocation decisions. Chapter 8 explores the determinants of happiness more generally, including the role of health therein. Chapter 9 investigates the causal relation between health and happiness, considering possible endogeneity bias resulting from the subjective nature of health and happiness.

Finally, chapter 10 discusses the main findings of this thesis, addresses its limitations and implications, and ends with some policy recommendations.

Note that the chapters of this thesis are based on papers published (or submitted for publication) in international peer reviewed journals. Therefore, these chapters can be read independently and some overlap may exist between chapters.

Part A

2

Balancing equity and efficiency in the Dutch basic benefits package using the principle of proportional shortfall

Based on:
Van de Wetering, E. J., Stok, E.A., Van Exel N.J.A. & Brouwer W.B.F. (2013). Balancing equity and efficiency in the Dutch basic benefits package using the principle of proportional shortfall. The European journal of health economics, 14(1), 107-115

ABSTRACT

Economic evaluations are increasingly used to inform decisions regarding the allocation of scarce health care resources. To systematically incorporate societal preferences into these evaluations, QALY gains could be weighted according to some equity principle, the most suitable of which is a matter of frequent debate. While many countries still struggle with equity concerns for priority setting in healthcare, the Netherlands has reached a broad consensus to use the concept of proportional shortfall. Our study evaluates the concept and its support in the Dutch health care context. We discuss arguments in the Netherlands for using proportional shortfall and difficulties in transitioning from principle to practice. In doing so, we address universal issues leading to a systematic consideration of equity concerns for priority setting in health care. The article thus has relevance to all countries struggling with the formalization of equity concerns for priority setting.

2.1. INTRODUCTION

Economic evaluations are increasingly used to inform decisions regarding the allocation of scarce health care resources. They generally take the form of cost-utility analysis, in which incremental costs per gained QALY (Quality Adjusted Life Year) are evaluated against some threshold to ascertain the intervention's value for money. In this procedure QALY gains are (implicitly) valued equally irrespective of, for instance, the beneficiary or disease. Such practice has been an issue of debate, however, because accumulating evidence shows that the public may prefer some QALY gains over others (e.g., young over old) [2, 3, 14], often relating to a more equitable distribution of health and health care. Such notions of equity are normally not captured in economic evaluations where QALYs are typically weighted equally and often remain implicit in subsequent policy decisions, if included at all.

Almost three decades ago researchers recognized that equity concerns could be incorporated into allocation decisions in the health care sector by weighting QALY gains according to some agreed upon equity principle [38], such as giving more weight to gains in the severely ill. Nonetheless, explicit QALY weighting is still uncommon. It seems that little has changed since Schwappach's 2002 assertion that equity weighting, if considered at all, was at a developmental stage [14]. However, given the growing pressure on health care budgets – partly due to (expensive) new technologies and increased demand stemming from demographic changes – we can expect that the process of allocating scarce health care resources will require increasing attention. Since decision makers increasingly use economic evaluations to inform their decisions [39], the discrepancies between recommendations based on economic evaluation outcomes and actual or publicly-desired decisions may become more evident, as experienced in the UK, Netherlands, and Australia [40-43]. One explanation for such discrepancies is the presence of equity concerns that are insufficiently reflected in current economic evaluations [44]. Therefore, we need a more explicit and systematic incorporation of equity weights to obtain sustainable decisions.

Recent developments in the UK with respect to the funding of costly life-prolonging cancer drugs illustrate the attention to equity concerns [45, 46]. In general, the National Institute for Health and Clinical Excellence (NICE) would like to adhere to the principle that 'a QALY is a QALY', implying that all QALY gains should receive equal weight [8], but under the rule it was difficult to come to sustainable decisions in the context of costly life-prolonging cancer drugs. The appraisal committees thus now explicitly consider the "magnitude of the additional weight that would need to be assigned to the QALY benefits … for the cost-effectiveness of the technology to fall within the current threshold range" [47]. The rule, specific in terms of applicable interventions and open in terms of

what weights might be considered appropriate, might be a first step in defining more general rules regarding equity weights.

In response to the equity problem the Dutch have developed a decision-making framework that defines an equity-weighted basic benefits package. Its primary criteria (proposed in 1991) are necessity, effectiveness, and efficiency [48]. The first refers to a notion of equity based on the need for medical intervention; the latter two refer to the merits of the intervention itself [20]. The three criteria inform the decision for including an intervention in the benefits package. Equity weights thus are intended to be part rather than exceptions to the rule. In this context, broad consensus appears to exist for using a particular operationalization of necessity as basis for equity weighting, i.e., the concept of *proportional shortfall* [19-21, 49]. Although choosing a specific equity principle may lead to a more systematic and transparent way of using equity weights to set priorities in the Dutch healthcare system, it also requires justification.

The Dutch and UK experiences suggest that including equity concerns in practice is not straightforward. Arriving at a systematic consideration raises two important questions: (i) which equity principle(s) are used to base QALY weights on, and (ii) how can we derive practically applicable QALY weights that are in line with the chosen principle(s). These difficult and inherently normative questions complicate the formalization of equity concerns.

To our knowledge, the Netherlands is the first country where decision makers and health economists have been involved in a joint effort to formulate an equity principle and develop a model for putting it into practice. Although the concept of proportional shortfall is not yet firmly implemented, it does appear to be the ultimate goal. Evaluating the proportional shortfall concept and the support for it in more detail is therefore directly relevant to the Dutch situation as well as other countries currently struggling with the formalization of equity concerns for priority setting in healthcare.

In this article we will evaluate the proportional shortfall concept as used in the Dutch health care context, with a focus on the abovementioned research questions on QALY weighting within the context of a chosen principle.

2.2. ECONOMIC EVALUATION AND EQUITY

Equity weights are integrated into economic evaluations to adequately consider costs and benefits. The common decision rule for economic evaluations is shown in Equation 2.1 [7]:

Eq. 2.1 $V_{QALYi} * \Delta QALY_i - \Delta costs > 0$

where V_{QALYi} denotes the monetary value society attaches to a QALY of type *i* and the subscript *i* is used to distinguish QALY gains according to some equity principle. $\Delta QALY_i$ represents the number of type *i* QALYs gained; Δcosts are the associated costs. Both are relative to some relevant comparator. Accordingly, the term $V_{QALYi} * \Delta QALY_i$ reflects the benefits related to the intervention. Put simply, Equation 2.1 indicates that incremental benefits of the intervention need to outweigh its incremental costs to be eligible for funding.

In common economic evaluations, however, the monetary value component in the benefits is not included in the equation. Rather the focus is only on $\Delta QALY_i$ (commonly without any distinction between QALYs, following 'a QALY is a QALY no matter who gets it') and Δcosts. This means that common economic evaluations in the field of health care do not directly address V_{QALYi}. As such, these analyses are basically partial economic evaluations. This is easily demonstrated by reordering Equation 2.1:

Eq. 2.2 $$\frac{\Delta \text{ costs}}{\Delta \text{ QALY}_i} < V_{QALYi}$$

where the costs per QALY gained of type *i* have to be lower than the societal value attached to that particular type of QALY in order to be eligible for funding. Only then a common cost-effectiveness or cost-utility analysis becomes a full economic evaluation as noted in Equation 2.1. (Note that irrespective of how QALYs, costs and potentially the value of QALYs are derived, the equivalence between a common cost-benefit analysis and the decision rule based on a cost-utility analysis may still be questioned [50, 51]).

Commonly, one threshold is set for all QALYs, regardless of the context in which they are gained. However, in line with Equation 2.2, different cost-effectiveness ratios may be acceptable for different types of QALYs gained *if* the value of a QALY is allowed to vary, for instance, on the basis of the disease or beneficiary characteristics. If so, rather than having one threshold value for all QALYs, we have a range whose endpoints are defined by the lowest and highest possible values attached to a gained QALY in a particular context. These different contexts may well refer to notions of equity. Note that such a practice is equal to keeping the threshold value constant but attaching 'equity weights' to the QALY gains on the left hand side of Equation 2.2. This implicitly ensures the use of an appropriate threshold value since the equity weights may simply be regarded as the relative values of different QALYs. In the Dutch context the former approach is taken; i.e., different threshold values are used when the burden of illness is high (e.g., acute life-threatening diseases) or low (e.g., toenail fungus). The latter approach appears to have

been adopted for cancer drugs in the UK, i.e., keeping the threshold fixed but weighting QALY gains.

To illustrate that one can either vary the threshold values or attach equity weights to QALYs, we use a hypothetical example in which the QALY value is allowed to vary between young and old. For children the value is highest, say, 1QALY = €100,000. For people age 90 and over, the value is lowest, say, 1QALY = €5,000. Let the reference value of a QALY be that of a 40-year-old, or €40,000. One way to judge a CE ratio of gains in children is to compare the CE ratio to the threshold line running from 5,000 to 100,000, where for this intervention the high endpoint is relevant. Alternatively, 'equity weights' can be used to adjust the CE ratio itself, which can be judged against the common threshold of 40,000. In this example, the equity weight of QALYs for children relative to the 40-year-old reference group is 2.5 ($v_{children}$ / $v_{40\ year\ olds}$ = 100,000 / 40,000). Thus, in order to use equity weights appropriately, we (implicitly) compare the relative values of different QALY gains to some reference group with the standard threshold. Subsequently, such equity weights can be multiplied with the QALY gains in the CE ratio (i.e. $\Delta C/2.5*\Delta E$), which indirectly corrects the threshold value used.

The above demonstrates how using a flexible threshold is basically equal to attaching appropriate equity weights to different QALYs and comparing them to a single, relevant threshold. In that sense, the Dutch and UK approaches are similar, albeit the UK approach currently seems to be used in only a few specific circumstances. A crucial question, however, is which equity principle should guide the derivation or evaluation of equity weights (research question (i)). To this end, we next discuss the well- known principles of fair innings and prospective health, followed by discussion of the principle selected in the Dutch context, proportional shortfall.

2.3. FAIR INNINGS AND PROSPECTIVE HEALTH

Which (combination of) equity concept(s) is most suitable for equity weighting? Several have been proposed even more equity concepts are imaginable [15, 52]. A problem is that improved equality with one particular definition of equity may be (necessarily) accompanied by greater inequalities in the context of a different definition [53]. Selecting one (or more) principle(s) to guide the derivation or evaluation of equity weights is thus important and not straightforward. Two important principles are fair innings and prospective health. Fair innings, roughly speaking, strives for equity in lifetime health, while prospective health is more concerned with people's health expectations, regardless of experienced health. An aspect both principles share is that their basis is found within the health domain, i.e. they both focus on health characteristics of beneficiaries,

not on aspects like gender or income. The principles differ, however, in that one strives to equalize lifetime health and the other prospective health.

The fair innings approach, advocated by Alan Williams (1997), is based on the assumption that everyone is entitled to some 'normal' span of life or health achievement. Anyone failing to achieve this has in some sense been disadvantaged in terms of lifetime health, whilst anyone getting more than this is living on 'borrowed time' [15]. This assumption implies that QALY gains in people who have had their fair innings should be valued lower than QALY gains in people who are expected to get less than their fair innings. Thus, the equity weights depend on the expected lifetime QALY total and age is a key element (as proxy for lifetime health achievement), resulting in higher weights for QALY gains in relatively young persons and lower ones for those in relatively older persons.

In contrast, the principle of prospective health bases equity weights on the expected QALY profile of a person in the case of no treatment. This aligns with an alternative definition of need, namely, expected ill health over the remaining years of life [16]. Prospective health considers the expected health (including death) in future years in the case of non-intervention and distributes QALY gains initially to those with the worst prognosis if left untreated [16]. The approach appears to be related to the Rule of Rescue, which implies that rescuing identifiable individuals facing avoidable death should have priority over other types of care [54, 55]. While prospective health incorporates non-identifiable individuals and non-life threatening conditions, both prioritize people with poor health prospects.

The different perspectives of fair innings and prospective health obviously result in different equity weights. As seen in Table 2.1, group A faces immediate death and group B has one remaining QALY. However, group B consists of younger persons who, consequently, have enjoyed fewer QALYs than persons in group A. Adhering to the equity principle of prospective health, group A would have priority because they face immediate death. The fair innings principle, on the other hand, would prioritize group B since it comprises younger people, or, put more precisely, it has a lower lifetime quality-adjusted life expectancy [13, 16, 56].

Table 2.1: Illustrating fair innings and prospective health

Patient group	QALY consumed	QALYs remaining (prospective health)	Expected QALY total (fair innings)
A (imm. death)	60	0	60
B (younger)	40	1	41

A number of empirical studies have found at least some public support for both the principles of fair innings and prospective health [2, 14, 17], although it may depend on the context of the decision [57]. As mentioned, age is important to the fair innings principle. Both Dolan and colleagues (2005) and Schwappach (2002) have found in their reviews that the majority of studies reveal support for giving less weight to health gains in older people, but Schwappach argues that age preferences vary across countries, study designs and context. Additionally, both reviews note that age weighting may reflect underlying rationales other than the fair innings principle. For example, people may prefer health gains in young people because they expect them to last longer. Separation of the different rationales places specific requirements on the design of studies, and they are not always met. Furthermore, Shah's (2009) recent review indicates that the public prefers to prioritize individuals in poorer health rather than those in better health without treatment, even if it results in lower overall health gains [2, 14, 17]. Shah (2009), however, emphasized that the strength of the support should be estimated more precisely to gain a true reflection of it.

There appears to be little evidence showing that either equity principle reflects the distributional preferences of society completely, or that one fully lacks support. In that regard, we should note that people may in fact adhere to both principles: someone may feel that (holding other things constant, i.e., health prospects) young people should receive priority over older people and, *at the same time*, feel that (holding other things constant, i.e., age) people with worse health prospects should receive priority over those with better health prospects. In the Dutch context, therefore, the equity principle proposed was an intermediate principle between fair innings and prospective health, or "proportional shortfall". We should note that the Dutch decision, while deliberatively taken and explicitly justified, should be seen as a first, pragmatic attempt to find an equity principle that is practically applicable and supported by the public.

2.4. THE PRINCIPLE OF PROPORTIONAL SHORTFALL

The concept of proportional shortfall adopts the normative viewpoint that priority should be given to those patients who lose the greatest *proportion* of their remaining health expectancy due to some illness if the illness remains untreated. In other words, measurements of inequalities in health should concentrate on the fraction of QALYs lost due to illness, relative to remaining life expectancy. Proportional shortfall (PS) can be measured on a scale from 0 (no health loss) to 1 (complete loss of remaining health) using the following formula:

$$\text{Eq. 2.3} \quad PS = \frac{\text{Disease related QALY loss}}{\text{Remaining QALY expectation in absence of the disease}}$$

where the denominator reflects the remaining QALY expectation in normal health, which could for example be determined on the basis of age and gender. The numerator presents the QALY loss, which is determined by deducting a patient's QALY expectancy given the disease without treatment from the remaining QALY expectancy in absence of the disease. The proportional shortfall is 1 for all patients who face a threat of immediate death, irrespective of their age. Since they will lose 100% of their remaining life expectancy they all receive equal weight. Likewise, if a young patient with a normal QALY expectation of 40 loses 20 QALYs, he or she will get the same equity weight as an older patient with a QALY expectation of 2 who stands to lose 1 QALY: both patients lose 50% of their remaining life expectancy. Since proportional shortfall is a relative measure, both younger and older individuals can experience a low or high proportional shortfall. For another example, a 30- year-old losing 1 of 40 remaining QALYs would receive low treatment priority (1/40 = 0.025) while a 70-year-old losing 1 of 5 remaining QALYs would receive higher priority (1/5 = 0.2).

Whereas fair innings and prospective health equalize absolute health outcomes in terms of total and future health, proportional shortfall proposes to equalize relative attainments. By doing so, the concept combines elements of both fair innings and prospective health. In accordance with fair innings, proportional shortfall is concerned with disease-related QALY loss; at the same time, in accordance with prospective health, it takes the remaining QALY expectation without treatment into account [18]. Therefore, the principle may be perceived as an intermediate position between fair innings and prospective health.

Obviously, this does not necessarily mean that proportional shortfall is in any sense 'better' than the other two. Its use must be justified, both normatively and empirically. A limitation to normative justification is that unlike, say, fair innings, the concept is not derived from a particular theory about distributive justice. Instead, normative arguments for both prospective health and fair innings are assumed to be compelling; thus, given that both principles result in different prioritizations, a reconciliation of or trade-off between the two is required. While balancing two principles that appear to have some normative and empirical support may intuitively make sense, whether the resulting combination has the same (or even a better) moral status than the individual principles themselves remains questionable [58].

One convincing argument for this (and thus proportional shortfall) may be if the latter reflects societal preferences better than either of its underlying principles. Evidence is thus far inconclusive, as is the only head-to-head comparison of all three concepts

in which Stolk and colleagues (2005) performed an experiment in a sample of Dutch health policy makers, researchers, and students to explore support for the individual systems. Observed rank orders of ten conditions were compared with rank orders based on the three equity principles. While both fair innings and proportional shortfall were highly correlated with the observed rank order, rank orderings based on prospective health showed only a moderate correlation coefficient with the respondents' ranking. Fair innings seemed to slightly outperform proportional shortfall, but neither offered a fully accurate description of societal preferences: there were always cases where the predicted rank ordering diverged from the observed. It was therefore concluded that more (elaborate) research was required before firm conclusions could be drawn about which concept best reflects societal preferences.

Despite the limited empirical and theoretical evidence to support proportional shortfall, broad consensus exists in the Netherlands to use it for equity weighting. The choice was partly pragmatic, subject to possible adjustment according to incoming evidence and ongoing experience. From the outset it was obvious that the adoption of any equity principle would not be without problems. To avoid perfection becoming the enemy of the good, however, efforts were made to operationalize proportional shortfall for use in practice. We highlight this in the following section, while addressing the issue of how practically-applicable QALY weights can be derived in line with the chosen principle and how they can subsequently be used in decision making on the allocation of scarce health care resources.

2.5. THE PRACTICE OF PROPORTIONAL SHORTFALL

As indicated, three criteria have been particularly important in the delineation of the basic benefits package in the Dutch context: necessity, effectiveness, and efficiency. While proposed and well received about two decades ago [48], adopting them for practical use has proven difficult and controversial. Effectiveness, which has always been a dominant criterion in the health care sector, was least controversial but not efficiency, which is nonetheless increasingly used and accepted. It has been operationalized primarily through cost-effectiveness and cost-utility analyses, and most widely applied in the context of new pharmaceuticals. 'Necessity', while intuitively important, eluded definition and measurement and thus remained barely applied as a (systematic) selection criterion in practice until 2002. Attempts to exclude services based on the criterion of necessity commonly provoked political or societal debate. Proportional shortfall was introduced to provide a more systematic and quantitative definition of necessity, which also solved its problem of dichotomy: interventions were deemed either necessary or

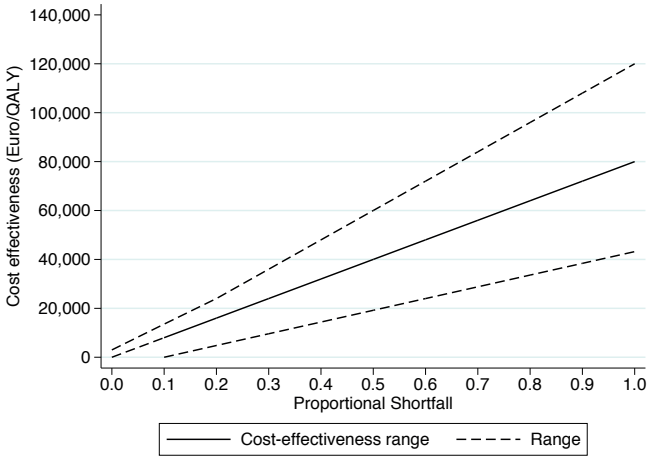


Figure 2.1: Decision making framework

unnecessary with no in-between. This proved virtually impossible as a proper definition of necessity and a clear and universal cut-off point between unnecessary and necessary (or, for that matter, cost-effective and cost-ineffective) care was lacking. Mostly, there would be (groups of) patients for whom it would be difficult to label the intervention as unnecessary or cost-ineffective [59].

The introduction of proportional shortfall made it possible to quantify the necessity criterion and to integrate it with or relate it to results regarding (cost-) effectiveness in the decision-making framework. Basically, the threshold of cost per QALY (i.e., the value of the QALY) was allowed to vary with the necessity of the intervention, creating a decision framework in line with Equation 2.2. The idea was that society is willing to pay more for an intervention, given the underlying proportional shortfall, considered more necessary. Put differently, a less favourable cost-effectiveness ratio is acceptable for an intervention in the context of a greater proportional shortfall, i.e., when the treatment is deemed to be more necessary, and unacceptable when the associated proportional shortfall is low. This decision- making framework is illustrated in Figure 2.1.

While the framework is increasingly supported [19, 49], it is unclear how it has affected decisions and it is likely that it is currently used predominantly as a conceptual framework. One reason may be that it is unclear how the threshold of costs per QALY gained varies with proportional shortfall. The equity weights placed on the different QALY gains (the value of gained QALYs relative to the underlying proportional shortfall) remain uncertain. In that sense, the threshold and margins in Figure 2.1 are largely tentative. Recently a maximum threshold height of some €80,000 has been suggested [49, 60], although the choice lacks sound basis [61]. Furthermore, it has been suggested that social

willingness to pay for QALY gains in the case of a very low burden of disease (proportional shortfall less than 10%) might not exist at all in the context of a collectively-funded health insurance scheme (which explains the dotted threshold line near the origin of the graph in Figure 2.1).[1] The shape of the curve would ultimately be based on relative valuations of QALY gains in different proportional shortfall contexts, information that is currently lacking. A linear relationship has thus been proposed as a pragmatic starting point [21] but, even under this assumption, we need to be able to adequately judge the proportional shortfall in different circumstances to use the framework in practice.

The framework is thus currently more conceptual than prescriptive in assessing health care interventions. Supporting this may be the fact that the shortcomings of proportional shortfall have barely been discussed. As highlighted by the fair innings and prospective health principles, not all consequences of the proportional shortfall principle may be in line with common conceptions of an equitable distribution of health care. Perhaps the most counterintuitive implication of the principle is that anyone facing imminent death should receive the maximum necessity score of 1, since all remaining health will be lost. This seems hard to defend when comparing between patients that differ substantially in age. Indeed, proportional shortfall assigns a necessity score of 1 when all remaining health is lost, regardless of the absolute number of life-years lost. The principle is indifferent to whether a three year-old is losing 80 years or an 80 year-old is losing three years, but in practice many people judge intervention to be more necessary in the former situation [61], making it conceivable that the result conflicts with society's equity principles. Likewise, since women have a higher life expectancy than men, an absolute QALY loss at a certain age will have more weight for men than women (for instance when comparing breast cancer and prostate cancer in certain age groups) and it is unclear whether this would be judged equitable [62-64].

Moreover, it became apparent that the operationalization of proportional shortfall required numerous normative choices that have important impacts on final outcomes. Coming to practically-applicable proportional shortfall scores that are in line with public preferences is therefore not straightforward [62]. For instance, the calculation of proportional shortfall in preventive treatments requires clear normative choices. Considering that many who receive preventive treatment will never experience the negative health effects precluded by the intervention, which group is relevant? While the treated group may be relevant in a cost-utility analysis of a preventive intervention, it may not be when calculating the proportional shortfall of the underlying disease. Calculating proportional

[1] Asserting that a disease has a low necessity of treatment is in itself difficult. A relatively small health loss may be due to something severe during a small period of time or something relatively mild but chronic. Such profiles may be evaluated differently, as discussed later in the text.

shortfall over the entire group would result in a very low *average* proportional shortfall since only a small percentage of the treated group would actually experience a health loss. This in turn results in low priority for (primary) preventive action, but the very *aim* of the intervention is to avoid health loss in those who would experience it *without* the preventive intervention. Then, it seems reasonable to calculate the proportional shortfall in the subgroup only, resulting in a higher proportional shortfall and threshold value. (The latter position is taken in the Netherlands.) Consequently, normative choices are necessary in defining the relevant group in which to determine proportional shortfall, which does not necessarily coincide with the population involved in the economic evaluation.

The timeframe for calculating proportional shortfall is another issue. Should the onset of preventive treatment be the starting point from which to calculate proportional shortfall? Or should it be the moment at which the negative health effects would have actually occurred? Obviously, the shorter timeframe will increase the proportional shortfall. Consider, for example, that a preventive intervention reduces a risk factor that, left untreated, results in death 20 years hence, reducing lifespan by 10 years. Until the moment of death, patients are without health loss. Calculating proportional shortfall starting from treatment time means that the first 20 years are in health and only the final 10 years are lost, resulting in a proportional shortfall of 33%. Calculating from the moment of illness the proportional shortfall is 100% (since then 10 of the remaining 10 life-years are lost). In the Netherlands, it has been argued that since society is likely to feel quite different about acute death than about a predicted (or certain) death in 20 years, calculating proportional shortfall from the moment of treatment would be more appropriate. Again, this is a normative choice, with substantial influence on results.

An additional problem exists in handling episodic diseases. An average severity of illness of 0.04 can result from a stable yet mild condition that causes a loss of 0.04 per day as well as from a disease that is primarily latent (no health loss for 350 days per year), but leaves the patient in agony during the episode (a loss of 1.0 during the remainder of the year). In the current operationalization of proportional shortfall, such episodes are simply averaged over the full year (as in normal QALY calculations). However, the appropriateness and justification of such a simple method of transforming health profiles into proportional shortfall scores can be disputed. Can we really conclude on such a basis that a certain illness has a modest severity? It appears that the variation over time may be important here too; yet, how could or should this be included in the calculation of proportional shortfall?

The above illustrates that not just the choice of an equity principle is normative; putting whatever it is into practice requires additional normative choices. Whatever the chosen principle, it appears inevitable that counterintuitive prioritizations may result in certain

circumstances. Clearly, therefore, decision makers should be aware of additional and potentially conflicting equity considerations.

2.6. DISCUSSION

Explicit inclusion of equity weights in the decision-making framework for allocation decisions in the health care sector has become increasingly important. A pragmatic start has been made in the Netherlands with the principle of proportional shortfall, which adopts the normative viewpoint that when people stand to lose relatively more of their remaining health, a higher cost per QALY threshold is appropriate. It thus quantifies the criterion of necessity in the Dutch decision-making framework. The higher the proportional shortfall, the more necessary the intervention.

This article highlights that the approach is not without problems. Both the normative basis and empirical support warrant further study. Of particular concern are situations where the consequences of proportional shortfall diverge from public preferences. For example, it seems hard to defend that avoiding a full loss of all remaining health would be equally important when the choice concerns either a very large or small *absolute* QALY loss, i.e., young and old people, respectively. Whether proportional shortfall adequately reflects societal preferences in such cases is uncertain and information on the circumstances of misalignment is pivotal in refining the principle and its employment.

We also highlighted that operationalizing proportional shortfall (or of any equity principle for that matter) involves normative choices that can have a profound effect on outcomes, like in the case of preventive interventions. It is crucial that these normative choices are as widely discussed as those embedded in cost-effectiveness analyses.

The highlighted shortcomings of proportional shortfall clearly should not be misinterpreted as a plea to replace it with a different equity concept such as fair innings. Indeed, whatever principle is chosen, similar shortcomings and normative choices will arise in transitioning from principle to practice. Since different notions of equity – all of which have some support in some instances – will always conflict in certain circumstances [53], conflicts with societal preferences will be inevitable. At this moment, there is no conclusive evidence that another equity principle reflects the distributional preferences of society better than proportional shortfall. It seems more appropriate, given the relatively strong (political) support for the equity concept of proportional shortfall in the Dutch context, to further test, develop, and refine the principle and its operationalization. For example, one might consider incorporating age weights in proportional shortfall in

order to better reflect societal preferences.[2] Obviously, this requires more and detailed research on relevant societal preferences as well as public debate. Improving the quantification of necessity will most likely be a lengthy and difficult process.

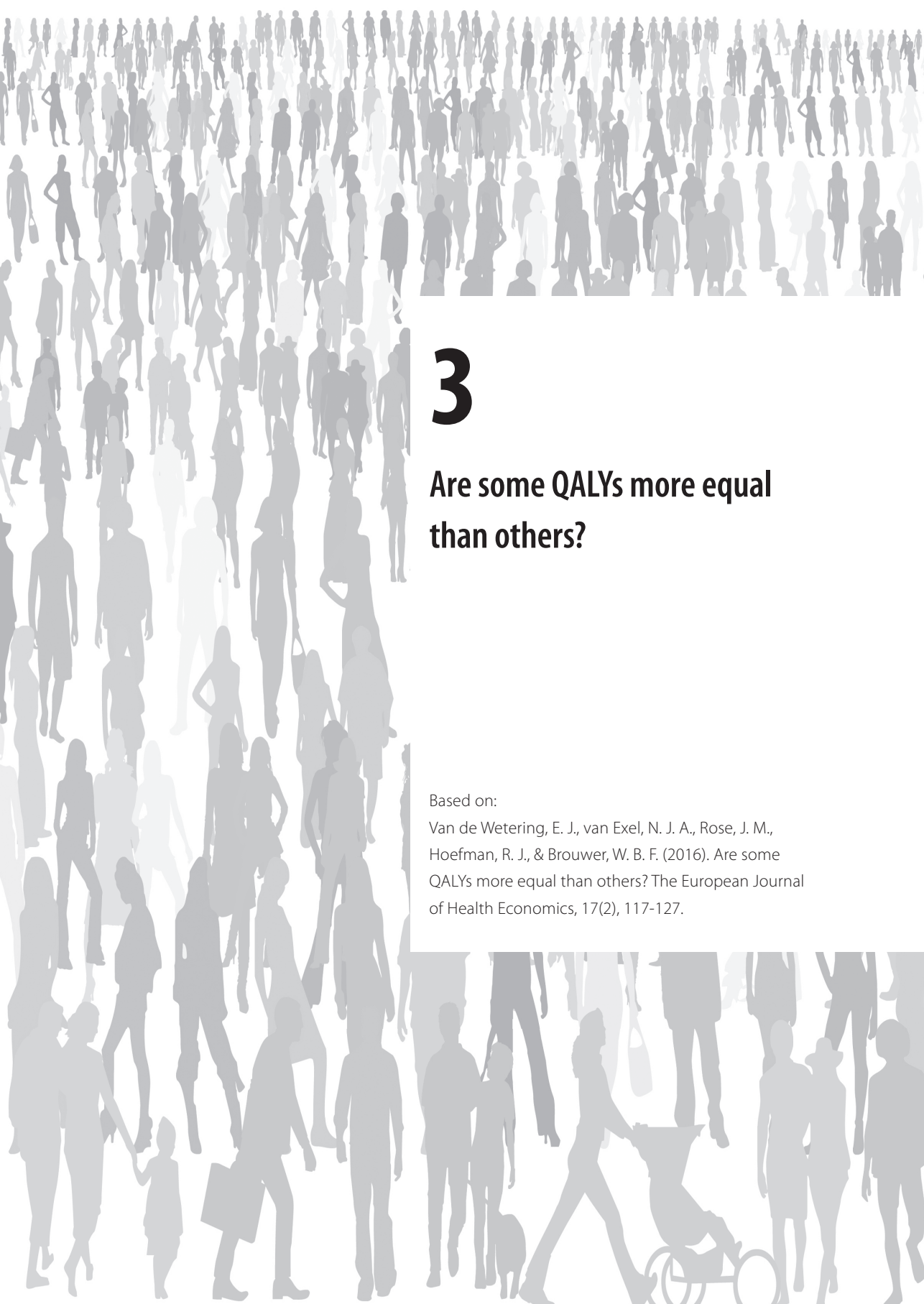
The current situation, while perhaps far from perfect, may be seen as an important and perhaps essential step in the development of an appropriate set of equity weights in the Netherlands. The associated quantification of necessity should improve consistency and transparency in the decision-making process. Meanwhile, experience from the systematic use of proportional shortfall will potentially improve the principle and its practical use.

Besides proportional shortfall or health profiles of beneficiaries in general, many other factors may be relevant for decision making in relation to equity considerations. Reviews by Dolan et al. (2005) and Schwappach (2002) have identified numerous factors besides health attainments or prospects that appear to influence the relative valuation of QALYs, such as prior health consumption, culpability, age, having dependents, and socioeconomic status. More recent studies have added to this field [17, 65], but so far most involve small and unrepresentative samples, the studies are quite context-specific, and findings are sometimes contradictory. It seems difficult at this stage to be conclusive regarding the relative weights given to these considerations in an empirical sense. While it may be interesting and helpful to analyse such additional equity concerns in relation to proportional shortfall, such empirical work should coincide with normative debates regarding whether such additional (or alternative) concerns should be included in the decision-making process. For instance, even if the public (on average) considers culpability important in fair allocation of health care resources, wanting to institutionalize such sentiments is questionable if only because Dutch legislation prohibits it.

Another challenge in the Dutch context is further quantification of the decision model. Currently, QALY value is unclear, let alone how it varies with different proportional shortfall percentages and which equity weights should be placed on various QALY gains. To use the decision-making framework in practice, the (relative) values assigned to QALY gains for different levels of proportional shortfall have to be elicited, for instance, from the public. Different methods have been used such as willingness to pay, person trade-off, or discrete choice analysis [3]. Which technique best captures the preferences of society may depend on the research question and whether relative weights of various equity concerns will be investigated simultaneously in combined trade-offs.

[2] Obviously, this also depends on whether one wishes to consider societal preferences to be a good guide for normative choices.

In conclusion, although proportional shortfall provides important information for decision makers by acceptably quantifying the necessity of treatment in the Dutch context, it clearly does not perfectly capture societal preferences. Sufficient room should be left in the decision-making framework and process to judge whether the equity weights accurately reflect the public preferences in particular circumstances and to improve on principles and practice if so indicated. More generally, the Dutch experience with equity considerations in relation to economic evaluations has given insight into the difficulties related to the choice for and operationalization of an equity concept for the allocation of scarce health care resources. Although the Dutch experiences are based on the concept of proportional shortfall, similar issues are likely to occur when opting for other equity concepts. Therefore, the Dutch experiences can provide helpful lessons for countries currently struggling with the important issue of formalization of equity concerns in priority setting in health care.



3

Are some QALYs more equal than others?

Based on:

Van de Wetering, E. J., van Exel, N. J. A., Rose, J. M., Hoefman, R. J., & Brouwer, W. B. F. (2016). Are some QALYs more equal than others? The European Journal of Health Economics, 17(2), 117-127.

ABSTRACT

Including societal preferences in allocation decisions is an important challenge for the health care sector. Here, we present results of a phased discrete choice experiment investigating the impact of various attributes on respondents' preferences for distribution of health and health care. In addition to the renowned equity principles severity of illness (operationalized as initial health) and fair innings (operationalized as age), some characteristics of beneficiaries (culpability and having dependents) and the disease (rarity) were included in the choice experiment. We used a nested logit model to analyse the data. We found that all the selected attributes significantly influenced respondents' choices. The phased inclusion showed that additional attributes affected respondents' preferences for previously included attributes and reduced unobserved variance. Although not all these attributes may be considered relevant for decision making from a normative perspective, including them in choice experiments contributes to our understanding of societal preferences for each single attribute.

3.1. INTRODUCTION

With the increasing use of economic evaluations to inform decisions regarding the allocation of scare resources in health care, it has become evident that allocation recommendations directly based on traditional economic evaluations do not (always) align with societal preferences. This is importantly related to the fact that traditional economic evaluations are based on the well-known rule that 'a QALY is a QALY is a QALY', while accumulating evidence shows that, from a societal perspective, some QALY gains are considered to be more valuable than others. Concerns for an equitable distribution of health and healthcare importantly underlie such preferences. This recognition has fuelled the debate regarding how to reconcile the outcomes of traditional economic evaluations with the existing distributional preferences in the general public [2, 3, 14].

This debate is not only academic. In actual decision making, differentiation between QALY gains of different types or to different beneficiaries is becoming more common as well. For instance, in the UK, in specific circumstances, such differentiation appears to be accepted as exception to the conventional rule that a QALY is a QALY [47], while in the Netherlands this differentiation has been put forward as the new rule for decision making rather than an exception [21, 49, 66]. The way in which the differentiation is addressed can differ as well. It sometimes is discussed in terms of 'equity weights', referring to a process of weighting QALY gains according to some equity principle in order to recognize that some gains are more important ('weighty') than others. Weighted incremental QALY gains are subsequently related to the incremental costs of producing them and judged against some fixed threshold. Differentiation can also be discussed in terms of a flexible threshold. Then, QALY gains are not weighted, but the incremental costs per QALY are compared to a differentiated threshold value, that varies with the equity context in which specific QALYs are gained. While the latter approach makes it more explicit that some QALY gains carry more (social) value than others, both procedures are equivalent [66, 67]. The attention for both appropriate decision rules [7, 68] and equity weights [2, 3] is increasing.

In deriving equity weights a first important choice is the basis on which equity classes are differentiated, i.e. the (combination of) equity principle(s) considered to be the appropriate base for weighting QALY gains. It should be emphasized that the normative question about which equity principle should justify and underlie equity weights (or flexible thresholds), may not necessarily be answered best (or even appropriately) through positive research. For example, not all characteristics the general public may consider relevant in the context of an 'equitable' distribution of health care, may be considered equitable (or desirable) from a normative viewpoint.

The equity principles 'severity of illness' and 'fair innings' have been regularly proposed as suitable candidates to conceptualize important distributional concerns in society. The principle of severity of illness bases equity weights on current and future health profiles. This approach can take into account severity of illness at the time of intervention, but also expected severity – including death – in future years in case of non-intervention [56, 69]. Proportional shortfall, the equity principle used in the Netherlands, emphasizes the proportion of health lost due to some disease [62] and could therefore also be seen as a measure of severity of illness [66]. The fair innings approach, advocated by Alan Williams (1997), is based on the assumption that everyone is entitled to some 'normal' span of life or health achievement. Equity weights then depend on this expected lifetime QALY total, resulting in higher weights for those who fall short of this norm and lower weights for those who exceed the norm. Lifetime health achievement is the key element in the argumentation of fair innings, especially in comparison to severity of illness [15, 56]. Although obviously not without problems, age is often taken as a proxy for lifetime health achievement. So far, there is no conclusive evidence that either severity of illness or fair innings better reflect the distributional preferences of society. Both principles rely on justified normative arguments and empirical studies have shown at least some public support for both of them [66].

Empirical studies furthermore highlight that the public may consider other characteristics, even some outside the health domain, to be relevant in allocation decisions on health and health care. People may for instance consider culpability (i.e. responsibility for own disease), social economic status, rarity of the disease and having dependents relevant characteristics of beneficiaries and their diseases in this context [2, 14]. However, the results of these studies are inconclusive as the (relative) importance of these variables and methodological differences hamper their comparability. Furthermore, it should be noted that many of these studies consider characteristics (e.g., age or severity) in isolation. This may not lead to an optimal insight in true distributional preferences, when people would normally consider multiple characteristics simultaneously. Moreover, there are indications that the results from empirical studies are sensitive to the framing of questions, inclusion of characteristics and the analysis [70]. To date, only few studies have considered multiple characteristics simultaneously and have attempted to quantify relative equity weights for QALY's based on several characteristics [14, 70, 71].

Another important issue is how to validly quantify equity weights in empirical research. Obviously, such quantification is essential, if the objective is to use equity weights in actual allocation decisions [18, 61, 64, 70]. More information on how elicited distributional preferences are affected by the information provided in an elicitation exercise remains crucial. This can also reveal how people weight different equity characteristics of beneficiaries of health programs and their diseases relative to each other.

This study aims to contribute to the literature in this important area. It does so by considering both policy relevant aspects (which characteristics play a role in distributional preferences and to what extent) as well as methodological aspects (how does including more equity characteristics in a preference elicitation exercise affect observed distributional preferences). We tested the hypotheses that preferences of the general public for allocation of health and health care are influenced by:

i. the equity principle *severity of illness* (hypothesis 1);
ii. the equity principle *fair innings* (hypothesis 2) and;
iii. culpability, rarity of the disease and having dependents (hypothesis 3).

In addition we hypothesized that the preferences for severity of illness and fair innings remain stable when additional information regarding the decision context (e.g., the beneficiaries and their health) is provided (hypothesis 4).

To test these hypotheses, as well as to allow investigation of the impact of offering additional information on respondents' preferences, we used a 'phased discrete choice experiment' in which attributes were added stepwise to the choice sets. By using a nested logit (NL) model to analyse the data we were able to investigate how additional information affected the elicited preferences in previous phases of the experiment, also in terms of unobserved variance. In this way, this study provides insight in the relevance of the context of health care allocation decisions, which is essential for the understanding of distributional preferences and can help to explain some of the (contradictory) findings in the literature. To our knowledge, this is the first study that systematically considers this impact of additional information of the beneficiaries or the disease on distributional preferences with respect to health care allocation decisions in a large sample from the general public.

The structure of this paper is as follows. In section 2 we provide details of the methods used, including the identification of attributes and levels (section 2.1), the design of the questionnaire (section 2.2) and choice sets (section 2.3), and the analyses of the data (section 2.4). The results of these analyses are presented in section 3. Section 4 concludes the paper with an interpretation of the results in light of previous literature and some comments and suggestions in terms of policy and future research.

3.2. METHODS

Relative weights of the different attributes were elicited from a representative sample from the Dutch general public using a discrete choice experiment (DCE). DCEs can be used to identify the effect of individual characteristics of some good on preferences for that good independently of all other characteristics, which makes it a useful tool

to investigate societal preferences for health (care) allocation [72]. DCEs are based on random utility theory, which assumes that a respondent, confronted with a choice between two scenarios, always chooses the alternative with the highest utility value for him or her.

Let U_{nsj} denote the utility of alternative j perceived by respondent n in choice situation s. U_{nsj} may be partitioned into two separate components, an observed component of utility, V_{nsj} and a residual unobserved (and un-modelled) component, ε_{nsj}, such that:

Eq 3.1 $$U_{nsj} = V_{nsj} + \varepsilon_{nsj}$$

The observed component of utility is typically assumed to be a linear relationship of observed attribute levels, x, of each alternative j and their corresponding weights (parameters), β, such that:

Eq 3.2 $$U_{nsj} = \lambda_j \sum_{k=1}^{k} \beta_k x_{njk} + \varepsilon_{nsj}$$

where β_{nk} represents the marginal utility or parameter weight associated with attribute k for respondent n and the unobserved component, ε_{nsj}, is assumed to be independently and identically (IID) extreme value type 1 (EV1) distributed. This implies that for all pairs of alternatives the unobserved component has the same influence and consequently the same scale parameter λ. In other words, scale is arbitrary set to 1.0 in the multinomial logit model [73].

3.2.1. Identification of attributes and levels

An extensive review of the literature was conducted to identify potential relevant aspects for distributional preferences for health (care) [66]. In this process, the recent systematic reviews of Bobinac et al. (2012), Dolan et al. (2005) and Schwappach (2002) were important sources of information. In order to keep the experiment feasible for respondents we limited the design to six attributes: fair innings, severity of illness, size of health gain from treatment, culpability with respect to having the disease, having dependents and rarity of the disease. Severity of illness can be operationalized in different ways. In this experiment we used a simple definition, also to reduce the cognitive burden on respondents, describing severity of illness as quality of life prior to treatment. This also assured a uniform interpretation of severity of illness. Fair innings was operationalized as age of the patients. Appropriate levels for the attributes were selected and both the attributes and levels were tested in a pilot study with 100 respondents. This resulted in adjustment of the level ranges of the attributes quality of life before treatment and age for the main study (Table 3.1).

Table 3.1: Overview of attributes and levels

Attributes	Levels
Quality of life before treatment* (QOL) scale 0-100	5, 30, 55, 80
Health gain from treatment for one year (HG) scale 0-100	5, 10, 15, 20
Age* (AG) In years	10, 40, 70, 90
Culpability (CULP)	Probably not responsible Probably partly responsible Probably largely responsible Probably fully responsible
Rarity of the disease (RAR)	Prevalent Rare
Have dependents (HD)	Yes No

Note:* The range of levels was expanded after the pilot study. Original levels for quality of life before treatment were 25,40,55,50 and for age were 20,40,60 and 80 years. These original levels showed insignificant coefficients in the pilot study.

3.2.2. Questionnaire

To be able to test the four hypotheses, the questionnaire was divided into three blocks of choice sets in which attributes were added stepwise to the choice sets. Each block started with a detailed description of the attributes. The first hypothesis was addressed in the first block of the questionnaire, which purely focused on severity of illness and the scenarios thus consisted only of the attributes 'initial quality of life' and 'health gain from treatment'. The fair innings principle (hypothesis 2) was considered in the second block of the questionnaire by adding 'age' to the choice sets; respondents thus had information on 'initial quality of life', 'health gain from treatment' and 'age' in the second block. Block three addressed the third hypothesis, by exploring whether other characteristics of the illness and the recipients are deemed relevant for the allocation of scarce health care resources by the general public. Hence, in the third block 'culpability', 'rarity' and 'having dependents' were added to the scenarios. The scenarios and the information presented to the respondents were kept constant over the three blocks.

Each choice set consisted of two scenarios describing different groups of patients both in need of treatment. Respondents were asked to put themselves in the position of a health care decision maker with a limited budget. Respondents were then asked which of the two groups they would prefer to treat from each choice set (Figure 3.1). They were informed that any other aspects not described in the experiment (including number of patients and costs of treatment) were identical between the two groups. The alternatives were unlabelled and the choice sets were randomized within the different blocks of

You, as a decision maker, can only treat one of the two groups of patients. Which of the groups below would you prefer?

Group A	Group B
Quality of life is 55	Quality of life is 30
Gains for one year 10 units quality of life	Gains for one year 15 units quality of life
40 years of age	70 years of age
Probably fully responsible	Probably not responsible
Prevalent	Rare
No dependents	Dependents

As a decision maker I would prefer…

Group A ❑ Group B ❑

Figure 3.1: Example of choice set (block 3)

the questionnaire to avoid order biases in the results. The questionnaire was presented online and we obtained a sample of 1,001 respondents through an internet panel. The study sample was representative for the adult (18+) population of the Netherlands with respect to age, gender and level of education. As an incentive for participation, upon completing the questionnaire, respondents could select a charity to receive a small monetary donation.

3.2.3. Design choice sets

The program Ngene 1.1 was used to generate efficient multinomial logit designs for the main study. An efficient design minimizes the predicted standard errors of the parameters in order to optimize the information from each choice set. The efficiency of the designs was determined by the D-error, which is the most widely used measure of efficiency [74]. The results of the pilot study were used as prior estimates for the parameters. Since the levels of quality of life before treatment and age were adjusted after the pilot study we had to interpolate prior values for the new levels. The values of these attributes were included as Bayesian priors. Bayesian priors are more robust to misspecification because they optimize on prior distributions instead of fixed parameters [74]. Interaction effects between quality of life, health gain and age were included and for each block of the experiment we used 1,000 Halton draws. Halton sequence draws are based on a set of values for simulation which are well spaced in the interval and therefore more effective than random draws [73].

For the first two blocks of the questionnaire, designs with 20 pairwise choice sets were obtained and for the third block a design with 40 pairwise choice sets was generated. The choice sets were divided over 5 versions using a blocking variable. This resulted in 5 questionnaires with each 16 choice tasks, 4 choice sets each for blocks 1 and 2 and 8 choice sets for block 3.

3.2.4. Analyses

The basic multinomial logit (MNL) specification of choice modelling relies on the assumption that the unobserved components are independently and identically distributed (IID). However, the questionnaire in this study consists of three blocks which are in fact three different datasets with their own designs. Hence, it would be naïve to assume that the variances of the unobserved components of different datasets are identical. Since parameter estimates are confounded with scale, different variances of unobserved components between datasets make it impossible to compare the parameter estimates directly. Any differences in coefficients could therefore either be the result of differences in preference or differences in the unobserved variance. To overcome this issue of scale heterogeneity, the different blocks of the questionnaire were estimated jointly in a two level NL model.

The NL model is the most common model used to date in the literature to account for scale heterogeneity. The NL model is typically set up with a hierarchical tree like structure linking alternatives that share common scale or error variances. Each branch or nest of the model, which sits above the (elemental) alternatives in the tree, also will have its own utility as well as scale. The NL model allows for a (partial) parameterization of scale at each level of the model (after some normalization). The scale parameters within the model are inversely related to the error (co)variances of the common set of alternatives linked to that branch or nest and are multiplicative with the utility of those same alternatives.

Let $\lambda_{(j|b)}$ represent the scale parameter at the top branch level or nest and $\mu_{(j|b)}$ represent the scale at the elemental alternative level of the tree. The utility of an alternative located at the lower level of the tree like structure nested within branch or nest *b* is given as:

Eq 3.3 $$U_{nsj} = \mu_{(j|b)} \sum_{k=1}^{k} \beta_k x_{nsjk} + \varepsilon_{nsj}$$

where $\mu_{(j|b)} = \dfrac{\pi^2}{6\sigma^2_{(j|b)}}$ and where is the variance of the error term for each alternative *j* in branch *b*.

From Equation 3.3, the influence of scale and error variance upon utility can clearly be seen. As the error variance increases, the magnitude of $\mu_{(j|b)}$ decreases and hence the

observed component of utility decreases. Likewise, a decrease in error variance will result in an increase in $\mu_{(j|b)}$ and an increase in the magnitude of the magnitude of the observed component of utility.

The utility at the upper level of the tree structure is linked to the utility of the alternatives contained within the 'nest' below such that:

Eq 3.4 $$\lambda_b \left(\frac{1}{\mu_{(j|b)}} \log \left(\Sigma_{bej} exp \left(\mu_{(j|b)} V_{nsj|b} \right) \right) \right)$$

where $\lambda_b \frac{\pi^2}{6\sigma_b^2}$ represents the scale at the upper branch level.

The NL model remains over-parameterized, requiring the normalization of one or more parameters for model identification. It is typical to normalize either $\mu_{(j|b}$ or λ_b to 1.0 for one or more of the branches or nests. Normalizing $\mu_{(j|b}$=1.0 results in models that are said to be normalized to random utility 1 (RU1) whilst normalizing λb=1.0 produces random utility 2 (RU2) models [75, 76]. In either case, what is actually being estimated in the model is λ_b or $\frac{1}{\mu_{(j|b)}}$ rather than both $\mu_{(j|b)}$ and λ_b separately. The estimated parameters are often referred to as IV parameters within the literature.

The link between the scales contained at each level of the tree structure can best been seen when examining the choice probabilities produced from the NL model. These are calculated using Equation 3.5:

Eq 3.5 $$P_{nsj} = P_{nsj|b} . P_{nsb} =$$

$$\frac{exp(\mu_{(j|b)} V_{nsj|b})}{\Sigma_{i \in jb} exp(\mu_{(i|b)} V_{nsi|b})} \cdot \frac{exp\left(\frac{\lambda_b}{\mu_{(j|b)}} log \left(\Sigma_{b \in jb} exp(\mu_{(j|b)} V_{nsj|b}) \right) \right)}{\Sigma_{b=1}^{B} exp\left(\frac{\lambda_b}{\mu_{(i|b)}} log \left(\Sigma_{i \in jb} exp(\mu_{(i|b)} V_{nsi|b}) \right) \right)}$$

where $P_{njs|b}$ is the conditional probability that respondent *n* will select alternative *j* in choice task *s* given that alternative *j* belongs to branch *b* and P_{nbs} is the probability of respondent *n* choosing branch *b*.

The three blocks of the questionnaire were nested in three groups with their corresponding alternatives. The deterministic components of the elemental alternatives were represented by:

Eq 3.6 Block 1. $V_A/\lambda_{s1} = \beta_0 + \beta_1 QOL_5 + \beta_2 QOL_{30} + \beta_3 QOL_{55} + \beta_4 HG_5 + \beta_5 HG_{10} + \beta_6 HG_{15}$

$V_B/\lambda_{s1} = \beta_1 QOL_5 + \beta_2 QOL_{30} + \beta_3 QOL_{55} + \beta_4 HG_5 + \beta_5 HG_{10} + \beta_6 HG_{15}$

Block 2. $V_A/\lambda_{s2} = \beta_0 + \beta_1 QOL_5 + \beta_2 QOL_{30} + \beta_3 QOL_{55} + \beta_4 HG_5 + \beta_5 HG_{10} + \beta_6 HG_{15} + \beta_7 AG_{10} + \beta_8 AG_{40} + \beta_9 AG_{70}$

$V_B/\lambda_{s2} = \beta_1 QOL_5 + \beta_2 QOL_{30} + \beta_3 QOL_{55} + \beta_4 HG_5 + \beta_5 HG_{10} + \beta_6 HG_{15} + \beta_7 AG_{10} + \beta_8 AG_{40} + \beta_9 AG_{70}$

Block 3. $V_A/\lambda_{s3} = \beta_0 + \beta_1 QOL_5 + \beta_2 QOL_{30} + \beta_3 QOL_{55} + \beta_4 HG_5 + \beta_5 HG_{10} + \beta_6 HG_{15} + \beta_7 AG_{10} + \beta_8 AG_{40} + \beta_9 AG_{70} + \beta_{10} CULP_{partly} + \beta_{11} CULP_{largely} + \beta_{12} CULP_{fully} + \beta_{13} RA + \beta_{14} HD$

$V_B/\lambda_{s3} = \beta_1 QOL_5 + \beta_2 QOL_{30} + \beta_3 QOL_{55} + \beta_4 HG_5 + \beta_5 HG_{10} + \beta_6 HG_{15} + \beta_7 AG_{10} + \beta_8 AG_{40} + \beta_9 AG_{70} + \beta_{10} CULP_{partly} + \beta_{11} CULP_{largely} + \beta_{12} CULP_{fully} + \beta_{13} RA + \beta_{14} HD$

where V is the deterministic component of the random utility function, λ_s is the scale parameter and β are the parameters to be estimated. The coding of the corresponding variable is reported in Table 3.1. Alternative specific constant terms for each block of the questionnaire were included to account for potential left and right biases.

An important requirement to isolate scale is that at least one of the attributes has to be homogenous (generic) across the different blocks of the questionnaire. This implies that the preferences for this attribute are assumed to be homogenous. Quality of life before treatment and health gain from treatment were the only attributes included in each block of the questionnaire. Since quality of life before treatment was part of the main focus of this experiment, health gain was selected as the generic attribute. To justify this assumption we tested for possible interaction effects between health gain and the additional attributes in block 2 and 3. No significant interaction effects were found. Besides, we compared the Log Likelihoods of the independent MNL models with the NL Model holding health gain fixed. These models did not significantly differ. In order to estimate scale ratios for blocks 1 and 2, the scale parameter of the block 3 was normalized to 1.0.

After accounting for scale, we were able to test the fourth hypothesis. T tests were conducted to test for statistical differences between the parameters estimates across the three phases. Since we expected nonlinearities, all attributes were entered as dummy variables. Furthermore, we tested for interaction effects between quality of life before treatment, health gain and age. The statistical analyses were performed in the program Nlogit 4.0 (Econometric Software Inc.).

3.3. RESULTS

The dataset included 1,001 respondents representative of the Dutch population in terms of age, gender and educational level. 51.2% of the sample was female, the mean age was 47.9 years, 28.0% of the respondents was lower educated and 27.7% was higher educated. Most respondents reported to feel healthy with a mean score of 71.4 on a scale from 0-100 (with endpoints defined as worst conceivable health state and best conceivable health state) (Table 3.2). The results of the NL model are presented in Table 3.3. The model had a good overall model fit with a McFadden's Pseudo R^2 of 0.65.

Table 3.2: Demographic variables (n=1,001)

Variable	Category	Fraction	Mean	SD	Min	Max
Gender	Female	51.2				
Age			47.9	16.3	18	88
	18 – 34	24.5				
	35 – 49	26.2				
	50 – 64	28.8				
	>64	20.5				
Education status	Elementary school	28.0				
	High school	44.3				
	University	27.7				

The first block of the questionnaire was used to test hypothesis 1 and aimed to elicit the relative values of initial quality of life (as proxy for severity of illness) and health gain from treatment. The results show that the respondents indeed considered initial quality of life relevant in their allocation decisions, as all levels had a significant influence on the choice outcome. Relative to an initial quality of life score before treatment of 80, respondents preferred to treat groups of patients with an initial quality of life of 30 or 55, but the highest severity of illness (i.e., an initial health state of 5) reduced the probability to receive treatment. As shown in Table 3.3, as expected, compared to a health gain of 20 units, smaller health gains significantly reduced the probability to receive treatment.

In the second block of the questionnaire, the attribute age was added to the choice sets in order to test hypothesis 2, i.e., the influence of fair innings on distributional preferences for health care. The coefficient for age 70 was not significantly different from that for age 90, but the coefficients for ages 10 and 40 were. The similar magnitudes of the latter coefficients indicate that respondents had a relatively strong preference to treat younger (10 years and 40 years) patients relative to older patients (70 and 90 years), all else equal, but did not differentiate within these broader age groups. Notably, with

Table 3.3: Results of the NL model (n=1,001)

Attribute	Level	Block 1		Block 2		Block 3	
		Beta	s.e.	Beta	s.e.	Beta	s.e.
Quality of life before treatment	5	-0.106*	0.034	-0.020	0.034	0.107*	0.054
	30	0.111**	0.028	0.119	0.073	0.526**	0.083
	55	0.168**	0.034	0.102	0.062	0.446**	0.090
	80	-	-	-	-	-	-
Health gain from treatment for one year	5	-0.303**	0.433	-0.303**	0.433	-0.303**	0.433
	10	-0.182**	0.038	-0.182**	0.038	-0.182**	0.038
	15	-0.094*	0.033	-0.094*	0.033	-0.094*	0.033
	20	-	-	-	-	-	-
Age (in years)	10			0.601**	0.111	1.308**	0.115
	40			0.564**	0.124	1.157**	0.077
	70			0.076	0.069	0.330**	0.092
	90			-	-	-	-
Culpability	Probably not responsible					-	-
	Probably partly responsible					-0.283**	0.049
	Probably largely responsible					-0.463**	0.062
	Probably fully responsible					-0.612**	0.045
Rarity of the disease	Prevalent					-0.077*	0.026
	Rare					-	-
Have dependents	Yes					0.351**	0.034
	No					-	-
Constant		-0.024	0.013	-0.091**	0.026	0.039	0.025
IV parameter		0.378**	0.064	0.567**	0.111	1.000	-
McFadden's Pseudo R^2		0.654					

Note: **=p<0.001; *=p<0.15.

the inclusion of age, respondents became statistically insensitive to variation in severity of illness. However, the signs and sizes of the coefficients were comparable to the first block of the experiment. These results lend support to the second hypothesis with respect to fair innings.

Block 3 of the experiment, which addressed the third hypothesis, contained the full set of attributes (Table 3.1). All attributes significantly influenced the probability to receive treatment. In order words, besides severity of illness and fair innings, respondents in-

deed considered other characteristics of beneficiaries and their disease to be relevant in allocation decisions. The negative coefficients of culpability levels indicate that people gave relatively less priority to treatment of groups of patients who could be held more responsible for their own illness. Rarity of the disease also significantly influenced the probability to receive treatment. Ceteris paribus, respondents preferred to treat rare diseases over more prevalent diseases. However, the size of the coefficient shows that rarity is the least important characteristic in this study. Having dependents also significantly increased the probability of a group to be selected for treatment. Furthermore, all coefficients of initial quality of life before treatment were significant and positive. Relative to an initial quality of life score of 80, a score of 30 showed the largest impact on the probability to receive treatment, all else equal, followed by a score of 55. In line with block 1, the lowest initial quality of life score of 5 received lower priority than scores of 30 and 55, but in contrast to block 1 and 2, the initial quality of life score of 5 did receive a higher priority than that of 80. Age showed a similar pattern as in block 2, suggesting that respondents preferred to treat younger over older groups of patients, all else equal. However, while respondents still did not significantly differentiate between 10 and 40 year olds (although the coefficient for the former was slightly higher), they did now differentiate between 70 and 90 year olds, modestly favouring the former. Given the magnitude of the coefficients, age seems to be a relatively important characteristic when choosing which group of patients to prioritize, irrespective of other attributes.

With respect to the fourth hypothesis, the results of the NL model showed that the scale parameters for block 1 and block 2 were statistically significantly different from 0 and 1.0, the normalized value for block 3. This indicates that considerable scale differences were observed between the different blocks of the questionnaire. Therefore, MNL coefficients of the different blocks could not be compared directly. Any differences in coefficients could either be the result of differences in preference or differences in unobserved variance. The NL model accounts for these differences in the variance of the unobserved effects. Interestingly, the error variances decreased when more attributes were included in the choice sets.

After accounting for scale, a series of *t*-tests was conducted to identify statistical differences in parameter estimates of the attributes quality of life before treatment and age across the three blocks of the questionnaire. Apart from the quality of life estimates between block 1 and block 2, all parameter estimates were significantly different. However, it should be noted that the quality of life estimates in block 2 were not statistically significant. The associated large standard errors might partly cause the insignificance of the difference between blocks 1 and 2. The significant differences between the three blocks indicate that the additional attributes indeed significantly affected the elicited distributional preferences. For example, a different weight is given to initial health state

when additional information (on age, culpability, etc.) is provided. Herewith, hypothesis 4 is rejected.

3.4. DISCUSSION

Including societal preferences for the distribution of health and health care in allocation decisions has received quite some attention in the literature and increasingly appears to play a role in actual decision making as well. Hence, it is important to study such distributional preferences and to increase knowledge as to how to derive these societal weights of QALYs gains of different types in an appropriate way. This study aimed to contribute to the literature by using a phased discrete choice experiment to investigate the influence on these distributional preferences related to the principles severity of illness and fair innings as well as other potential relevant characteristics. Furthermore, our methodological approach provided the opportunity to investigate the impact of additional information on respondents' preferences. Here, we address the implications of our findings, the limitations of our study as well as some areas for future research.

The first distributional concern we investigated was severity of illness, operationalized here as initial quality of life. We found somewhat mixed results for this variable. Severity of illness was a significant predictor in block 1, but in block 2, after adding age as characteristic, it lost its statistical significance (although the signs and coefficients were similar to those in block 1). One explanation for such a finding could be that the step from block 1 to block 2 in the experiment caused respondents to especially focus on the single additional variable (age). In block 3 severity of illness turned out to be a significant predictor of the choices again, next to age, supporting hypothesis 1. Overall, the coefficients indicate its influence to be non-negligible, but also show a remarkable pattern. The lowest initial health state (5) received less priority in block 1 (and block 2, albeit non-significantly) relative to the highest initial health state, while the two intermediate initial health states received more priority. In block 3, the lowest initial health state did receive somewhat more priority compared to the highest level, but still less so than the intermediate levels. This result may be considered counterintuitive, as a higher severity may be expected to be associated with higher priority for treatment. However, this result could be driven by our study design. Note that the maximum health gain in our experiment was 20 units. This was chosen to avoid end states above 100 in the choice sets (since the best initial health state was 80). As a consequence, for the lowest initial health state, i.e., patients starting from 5, the best case scenario after treatment was a quality of life score of 25. This is still relatively low. Hence, the low quality of life *after treatment* may have induced a lower priority for treatment of those with high-

est severity, leading to this counter-intuitive result. There indeed are indications that respondents consider the health status after treatment in their decision which group of patients to prioritize [70]. Therefore, this result needs to be interpreted with clear caution. In future research it would be interesting to combine low initial health states with larger health gains and thus better health states after treatment. In order to do so, a design should be built in which the ranges of attributes can be expanded without creating unrealistic scenarios with quality of life scores which exceed 100 (perfect health). Moreover, since the results for the variable severity of illness were not constant across blocks, more research is required to explore the influence of severity of illness in priority setting, especially in relation to age.

The second distributional concern we investigated was fair innings, operationalized simply as age. Our results show a strong preference for treating younger (10 and 40 year-olds) over older patients (70 and 90 year-olds). In block 2, respondents did not differentiate *within* these groups (e.g., did not give more priority to 10 than to 40 year olds). In block 3, the results showed a similar pattern. However, there a significant (although modest) differentiation was observed between 90 and 70 year olds in the expected direction. In the literature, contradictory results have been reported regarding the relevance of age in health care decision making. Similar to our study, Winkelhage and Diederich (2012) showed that age was the most important criterion for allocations of health care resources, followed by severity of illness, improvement of health, healthy lifestyle, treatment costs and type of illness. However, a recent study by Lancsar et al. (2011) implied that no weight should be given to age of onset of illness, while earlier studies reported that with an increase in age the value per QALY sharply decreased [70, 77, 78]. Such differences in findings may be the result of framing, other information (e.g., age of death), as well as the levels of the attributes used [71, 79]. For instance, in our main study we used a wide range of levels for age, i.e., 10 to 90 years, and found a significant influence of age. In our pilot, we used a smaller age range, which did not result in significant coefficients. The absent or small differences between 10 and 40 year olds as well as between 70 and 90 year olds emphasize this point. We do emphasize that age does not fully reflect the fair innings argument, which considers the life time health achievement. Age was used as a (simplifying) proxy, to keep the choice task understandable. However, especially in relation to severity, the age variable here cannot be simply interpreted as representing fair innings.

In block 3, we added three more variables: rarity, culpability and having dependents. All three aspects proved to significantly affect the choices. We stress that such evidence need not imply that these aspects should also be used in actual decisions. Normative and positive science may conflict here. Rarity, for instance, has been frequently discussed in the literature, especially in the context of expensive orphan drugs, and opinions differ

regarding whether it would be a reason for equity weighting. Some have argued that rarity is a reason to attach higher weights to QALY gains (i.e., allowing higher cost-per-QALY ratios), while others suggest that this would be unduly and inequitably disadvantage people with common diseases [80, 81]. In this experiment, rarity was simply labelled as either rare or prevalent. Respondents proved to be sensitive to this attribute; giving slightly more weight to health gains in the context of rare diseases. It needs emphasis that, while significant, the size of the coefficient was relatively small. For culpability, the distinction between positive and normative science is important as well. Indeed, even when the general public (on average) considers culpability important in fair allocation of health care resources, it is questionable whether this can be justified from a normative perspective.

Overall, our results showed support for the first three hypotheses, as people on average seemed to consider severity of illness, fair innings and additional characteristics of the beneficiaries and diseases simultaneously in the health (care) allocation decisions.

The methodological choices related to our stepwise design also warrant further discussion. Our results demonstrate both scale and preference differences across the three blocks of the questionnaire. Block 1 showed the smallest scale coefficient relative to block 3, indicating that block 1 had the largest unobserved variance. After accounting for scale, respondents showed significantly different preferences for age and initial health state in the different phases of the experiment. One explanation might be that respondents require more 'contextual' information in order to make informed decisions. When respondents do not have sufficient information they may 'fill in the blanks' and create 'plausible scenarios' based on the limited information provided. For example, in block 1, which only provides information about the initial health state and potential health gain, respondents may assume that patients in a very poor initial health state are relatively old (even though they were instructed this was not the case). The problem is that the analyst then does not observe the implicit attributes used by individual respondents, which leads to more unobserved heterogeneity. Lancsar et al. (2011) recently also emphasized the relevance of providing sufficient information in order to ensure an informed response. Here, an intriguing trade-off may be observed: too many characteristics may result in too complex choice sets, which increases heterogeneity as well [82], while our results indicate that something similar happens when too little information is provided. Complex choice sets appear to result in a greater number of errors in respondents' choice process but also affects respondents decision strategy. Therefore, the key seems to be to provide sufficient information to respondents to be able to make an informed choice without making the choice sets to complex: the optimum between too little information and too much information. This is an important area for further research and such a balance could be tested in the pilot phase of future studies. This

finding requires the fourth hypothesis to be rejected: contextual differences do affect the elicited preferences.

This study has several limitations that deserve mentioning. We used the NL model in this study to account for scale heterogeneity. However, the NL model does not account for preference heterogeneity and panel data. Our data contains 16 observations per individual which is likely to result in correlated unobserved components. Furthermore, the results show preference heterogeneity across datasets but there might also be preference heterogeneity within the datasets. Mixed Multinomial Logit (MMNL) models can identify and account for panel and preferences heterogeneity but these models cannot isolate scale, which was essential in this study.

In addition, we tested for interaction effects between initial quality of life, health gain and age. The interaction terms did not behave consistently and were difficult to interpret. This seems partly due to the design of this study. A disadvantage of an efficient design is that not all combinations of the dummy variables are included in the design and that, therefore, the two-way interaction terms might be confounded [73, 83]. However, our tests did reveal some patterns useful for further research. Importantly, low initial health status combined with only a small health gain seemed to reduce the chance to receive treatment. This emphasizes the relevance of quality of life after treatment. Starting low and ending low may be perceived as a treatment with little effect (even though its absolute size is not different from a similar gain high on the scale) as it does not bring a patient to an 'acceptable' health state. Also, being relatively old in combination with a low quality of life or a small health gain seemed to negatively influence the choice outcome. Future research could investigate these issues further, for instance by already specifying such interactions while generating the design. Furthermore, it should be noted that the constant term in block 2 was significant. Although a constant term in an unlabelled experiment has no substantive meaning, it can be an indication of left and right bias. Apparently, respondents systematically preferred group B to A. However, given the modest magnitude of the constant term and the fact that the constant terms of blocks 1 and 3 were not significant, our results do not appear to be systematically biased.

Our design included certain constrains to avoid implausible scenario's, i.e. health states above 100 and a maximum age of 90. Still, some combinations of attributes may be considered less plausible (e.g. a 10 year old who is fully responsible of her/his health problems). We did not observe any signs of protest answers for such scenarios. The complexity and plausibility of the questionnaire was tested in the pilot study with additional questions about respondents' experiences with the DCE questions. It appeared that more than half of the respondents thought it was difficult to opt for one of the

groups in the choice sets. However, this was mainly due to the fact that people prefer not to choose between groups of patients at all. Only 16.2% of all respondents thought that the questionnaire itself was complicated.

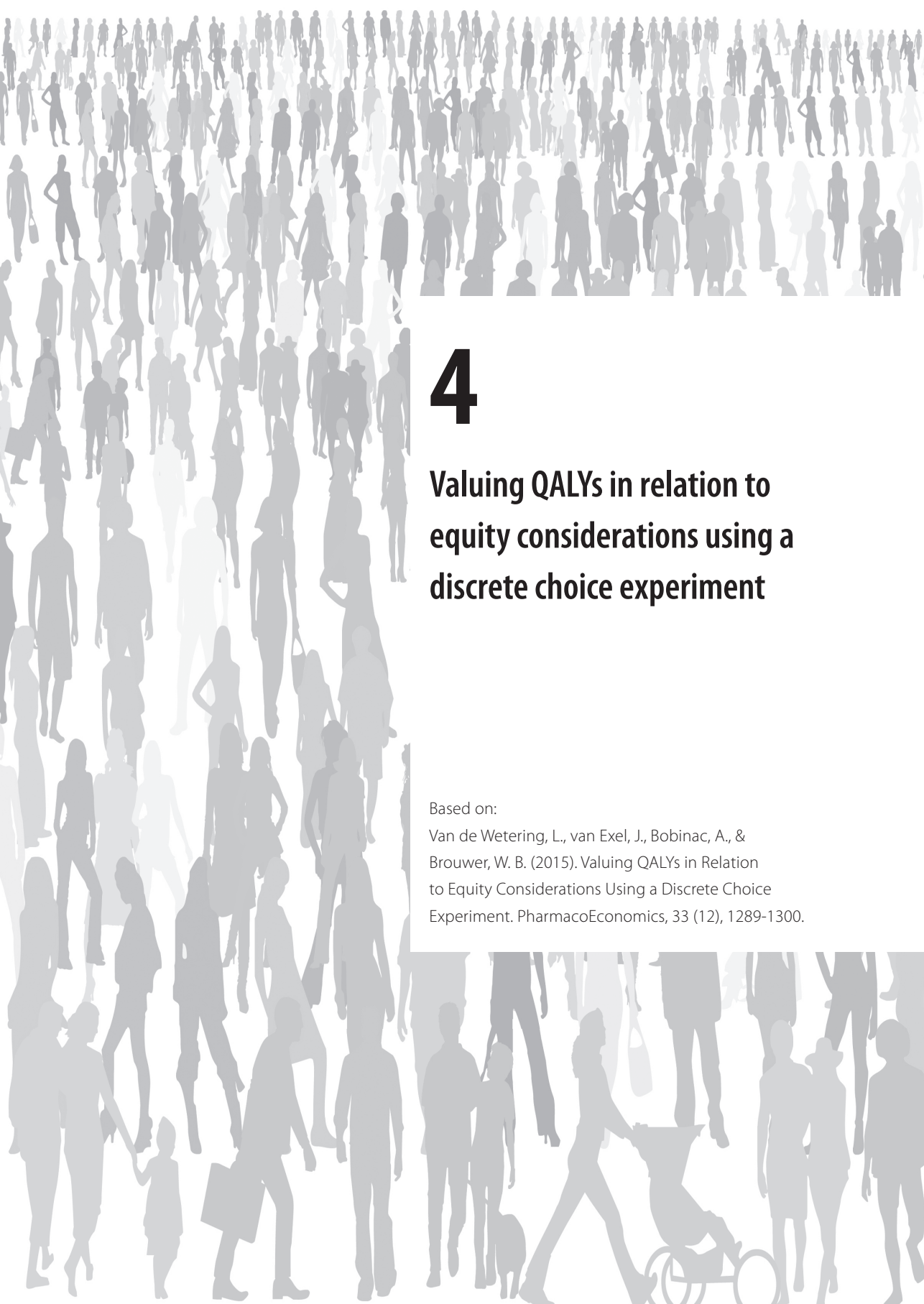
Despites the limitations, this study has important implications for the further investigation of distributional preferences in the context of health care decision making. First of all, in addition to severity of illness, age, as being a proxy operationalization of the fair innings argument, appeared to significantly influence public preferences for the distribution of health and health care and, important as well, significantly influenced the preferences regarding severity of illness. This may indicate that distributional preferences are indeed based on both severity of illness and fair innings. In that sense, this study suggests that while the fair innings and severity of illness approaches may have been contrasted in the literature, empirically derived distributional preferences may combine both principles. Past health performance (age) is important next to health prospects (initial health state). Such insights may feed normative debates, the development of new normative equity principles and future studies.

Furthermore, we found that all additional characteristics included in this experiment significantly influenced respondents choices and, at the same time, reduced the unobserved variance. As a result of our stepwise approach we were able to demonstrate that the inclusion of age, culpability, rarity and having dependents influenced the preferences for severity of illness. This suggests that describing the 'broader context', i.e., other characteristics of beneficiaries of health programs and their diseases, influences the elicited distributional preferences, also in relative terms. This implies that it would be preferable to include all relevant items in choice experiments. Omitting relevant information from the choice sets results in an increase of unobserved variation. Note that more relevant items may be identified than those included in this study and future research should be aimed at doing so. Then, as noted before, it is important to strike a balance between adding more relevant variables while not cognitively overburdening respondents. Providing more information does not imply that all attributes used in a study should also be considered by policymakers in the real life decision making. The normative questions of which aspects should be considered in distributing health and health care may not be best answered through positive research.

Finally, it should be noted that the relative (equity) weights of attributes derived in any study are useless without some reference case, described by its equity characteristics, and a corresponding cost-per-QALY threshold [67]. Relative to that reference case equity weights can be applied (or thresholds can be varied). Investigating the social willingness to pay for QALY gains in different equity contexts could be another interesting way of deriving equity weights.

Acknowledgements

This study was financially supported by Zorgonderzoek Nederland (ZonMW), Netherlands Organization for Health Research and Development (project number 1520020011). The usual disclaimer applies.



4

Valuing QALYs in relation to equity considerations using a discrete choice experiment

Based on:
Van de Wetering, L., van Exel, J., Bobinac, A., & Brouwer, W. B. (2015). Valuing QALYs in Relation to Equity Considerations Using a Discrete Choice Experiment. PharmacoEconomics, 33 (12), 1289-1300.

ABSTRACT

Background: To judge whether an intervention offers value for money, the incremental costs per gained quality-adjusted-life-year (QALY) need to be compared with some relevant threshold, which ideally reflects the monetary value of health gains. Literature suggests that this value may depend on the equity context in which health gains are produced, but the value of a QALY in relation to equity considerations has remained largely unexplored.

Objective: The objective of this study was to estimate the social marginal willingness to pay (MWTP) for QALY gains in different equity subgroups, using a discrete choice experiment (DCE). Both severity of illness (operationalized as proportional shortfall) and fair innings (operationalized as age) were considered as grounds for differentiating the value of health gains.

Methods: We obtained a sample of 1,205 respondents, representative of the adult population of the Netherlands. The data was analysed using panel mixed logit and latent class models.

Results: The mixed logit models showed counterintuitive results, with more severe health states reducing the probability of receiving treatment. The latent class models revealed distinct preference patterns in the data. MWTP per QALY was sensitive to severity of disease among a substantial proportion of the public, but not to the age of care recipients.

Conclusion: These findings emphasize the importance of accounting for preference heterogeneity among the public on value-laden issues such as prioritizing health care, both in research and decision-making. This study emphasizes the need to further explore the monetary value of a QALY in relation to equity considerations.

4.1. INTRODUCTION

Cost-utility analysis is increasingly used to inform allocation decisions about scarce healthcare resources. To evaluate whether an intervention yields good value for money the incremental costs per gained QALY (quality-adjusted life-year) must be judged against some monetary threshold value. The nature of this threshold is a matter of debate. One stream of literature considers it as the opportunity costs of spending within a fixed healthcare budget, while the other considers it to represent the consumption value of health gains [68]. Here, we take the latter view and, more precisely, consider the appropriate threshold to reflect the *social* willingness to pay for a QALY gain [84-86]. In other words, the threshold expresses the maximum acceptable cost to society for a QALY gained through an intervention. Without such a threshold the results of a cost utility analysis are of limited value to healthcare decision makers. Somehow, they must judge whether a treatment with a cost-per-QALY ratio of, say €50,000, offers value for money and should be reimbursed [87]. It need not surprise that this threshold has generated much debate. Societally, the idea of using a threshold expressing the value of health in monetary terms to decide about funding treatments has been contested [84]. Scientifically, the debate is especially about *how* to set a threshold, and whether there should be a fixed threshold or one that could vary with societal preferences for QALYs.

Regarding the latter issue, it is important to acknowledge that accumulating evidence suggests that the public prefers some QALY gains over others (e.g., those in young children over those in elderly) [2, 14, 88, 89]. This suggests that *the* social value of a QALY does not exist [90] but that this value may vary with, for example, characteristics of the disease and the beneficiaries of treatment [91]. The use of a single threshold in judging the results from economic evaluations would therefore not align with societal preferences. The distributional preferences of society can be incorporated in the decision framework by applying a more flexible threshold or, under a fixed threshold, by applying equity weights to QALYs [23, 66, 87].

Although in most countries the threshold is still rather implicit, differentiation between QALY gains of different types or to different beneficiaries already exists in actual decision making. In the UK, the National Institute for Health and Care Excellence (NICE) recently formulated a decision rule explicitly giving higher value to costly life-prolonging end-of-life drugs. Under the assumption that all QALY gains should be valued equally, these interventions would probably have exceeded the threshold range. The new decisions rule explicitly considers the "magnitude of the additional weight that would need to be assigned to the QALY benefits … for the cost-effectiveness of the technology to fall within the current threshold range" [47]. This exception may prove to represent a first step in defining more general rules using a flexible threshold, depending on the context

in which QALYs are gained [92]. The Netherlands has developed a decision-making framework in which the relationship between equity considerations and the value of a QALY has been made more explicit. The value of a QALY increases with the severity of illness in the target population, the latter being expressed using the concept of *proportional shortfall* [62, 66].

A fundamental question in the development of a decision framework using a flexible threshold is which equity principle(s) should be the basis for differentiation. In literature, the equity principles 'severity of illness' and 'fair innings' have been regularly proposed as suitable candidates. The principle of severity of illness considers severity at the time of intervention and expected severity – including death – in future years in case of non-intervention [56, 69]. The fair innings approach, advocated by Alan Williams [15], is based on the assumption that everyone is entitled to some 'normal' span of life or lifetime health achievement. As a result, a relatively high priority would be given to those who fall short of this norm and a relatively low priority to those who exceed this norm. Although obviously not without problems, age is often taken as a proxy for lifetime health achievement. Whether severity of illness or fair innings better reflects the distributional preferences of society is still a matter of debate, but both principles rely on justified normative arguments [66]. Proportional shortfall, the equity principle used in the Netherlands, is based on the proportion of remaining lifetime health lost due to some disease [62] and could therefore also be seen as a measure of severity of illness [66, 93]. Proportional shortfall measures the fraction of QALYs lost due to illness relative to remaining life expectancy in absence of the disease, on a scale from 0 (no loss) to 100 (complete loss of remaining health).

Empirical studies show mixed findings with respect to the direction and strength of the preferences for age and severity. These variations might be caused by the framing of the concepts, or by context and methodological differences between studies [22, 93]. Moreover, often only particular aspects of potential value are investigated (e.g. only age or only severity) rather than, arguably more relevant, combinations. This hampers not only definite conclusions about support for specific decision rules, but also about the exact values (weights) attached to different QALY gains.

In that context, it also needs noting that the monetary value of a QALY and equity weights have both received quite some attention in the literature, but typically not jointly in one study [94]. Most WTP studies focus on the individual perspective, asking respondents to value changes in their own health, thus ignoring equity considerations. In the context of healthcare allocation decisions it seems to be more appropriate to consider the *social* value of a QALY, defined by the amount of their own consumption individuals are willing to forego in order to contribute to a health gain achieved in society [23]. Other studies

have explored public preferences for a variety of equity principles and characteristics of the beneficiary or the disease, but these studies have not addressed the monetary valuation [17, 95]. To illustrate; a recent systematic review by Whitty et al. [22] shows an exponential growth in choice based studies to elicit public preferences with respect to healthcare priority setting. However, most of these studies have not translated preferences into equity weights, let alone included the monetary valuation of QALYs for different equity considerations [2, 14, 22, 96].

The objective of the current study is to contribute to the existing literature by estimating the social WTP for QALY gains in different equity subgroups. More precisely, we aim to estimate the marginal WTP for a QALY at different levels of proportional shortfall, in different age groups. The study was framed in such a way that it could be directly helpful in further shaping the (Dutch) decision-making framework and build on previous studies in this area [66, 67, 94]. Public preferences were elicited using a discrete choice experiment (DCE) which is currently the most commonly applied method to elicit public preferences [22]. Respondents were asked to act as social decision makers. We included both the equity principles 'severity of illness' (operationalized as proportional shortfall) and 'fair innings' (operationalized as age) in one experiment. In order to arrive at WTP per QALY estimates, we used the payment vehicle of increases in insurance premiums, which is the common financing mechanism in The Netherlands. In light of the diversity in the literature in terms of methods and results, we need to be modest in our aim. While we want to inform the (Dutch) debates regarding appropriate equity weights and thresholds, the current experiment was especially designed to learn how respondents solve the dilemmas they are confronted with, and to better understand support for differentiating QALY values between groups.

4.2. METHODS

4.2.1. Discrete Choice Experiment

DCEs are based on the assumption that a good can be described by its characteristics and that the relative importance of these characteristics can be identified in isolation. This makes the DCE a valuable method to explore the preferences for healthcare allocation in relation to equity considerations [22, 72, 97]. DCEs are modelled according to random utility theory, which assumes that a respondent asked to choose between multiple options always chooses the alternative with the highest utility for her/him. The utility of an alternative for respondent n, U_n, can be decomposed in an observable component of utility, V_n, which reflects the utility effect of the characteristics of the

alternative, and an unobserved component, ε_n, which reflects the utility not captured by these characteristics, such that:

Eq 4.1 $U_n = \lambda V_n + \varepsilon_n$

where λ is the scale parameter which presents the variance of the unobserved component.

Identification and Presentation of Attributes and Levels

The main objective of this study was to estimate the WTP for a QALY at different levels of proportional shortfall, in different age groups. Therefore, the following attributes were included: quality of life if untreated, age of death if untreated, gain in quality of life, gain in life expectancy and cost of treatment. The quality of life attribute was presented on a scale from 0 to 100, with 0 representing the worst imaginable health state and 100 representing perfect health. The cost attribute was operationalized as an increase in the mandatory health insurance premium for all Dutch adult citizens for a period of 1 year. To be able to explore fair innings (or ageism), we designed three versions of the questionnaire considering different age groups: 10 year olds, 40 year olds and 70 year olds. The levels of the attributes quality of life if untreated, gain in quality of life and costs of treatment were identical for all age groups. However, in order to present a comprehensible and plausible range of proportional shortfall in each of the three age groups to respondents, the levels of the attributes age at death if untreated and gain in life expectancy differed between age groups.

Next, to compensate for the smaller absolute health gains in the older age groups we differentiated the number of people at risk between the age groups. The number of affected people in the Dutch population was 2,000 people in the 10-year-old age group (age group 10), 4,000 people in the 40-year-old age group (age group 40) and 12,000

Table 4.1: Overview of attributes and levels

Attributes		Levels
Quality of life without treatment (scale 0-100)		45, 65, 85
Age at death if untreated (scale 0-80)	- Age group 10	30, 50, 70
	- Age group 40	50, 62, 74
	- Age group 70	73, 76, 79
Gain in quality of life		5, 15, 25, 35
Gain in life expectancy	- Age group 10	5, 10, 15, 20
	- Age group 40	2, 6, 10, 14
	- Age group 70	0.5, 1, 1.5, 2
Increase of health insurance premium (euro)		6, 12, 18, 24

Note: Affected people: 2,000 in age group 10, 4,000 in age group 40 and 12,000 in age group 70.

in the 70-year-old age group (age group 70). An overview of the attributes and levels is presented in Table 4.1 (Note that it has been found that people may prefer larger gains in fewer people over smaller gains in more people, even when the two add up to the same total [9]).

Following the approach adopted by Lancsar et al. [70], we used both words and diagrams to present the choice sets, as shown in Figure 4.1. Each scenario was represented by a graph with 'quality of life' on the vertical axis (on a scale from 0-100) and age on the horizontal axis (on a scale from current age until 80 years old) as shown in Figure 4.1. The light grey shaded area shows the health prospect without treatment, the dark grey shaded area combined with the middle grey shaded area shows the health loss without treatment (proportional shortfall). The middle grey shaded area shows the potential health gain from treatment.[3] Below the graphs, the percentages of remaining health without treatment, potential health gain from treatment and the increase in monthly premium were presented. Given the complexity of the graphs we first showed a step-by-step introduction of the graphs to respondents.

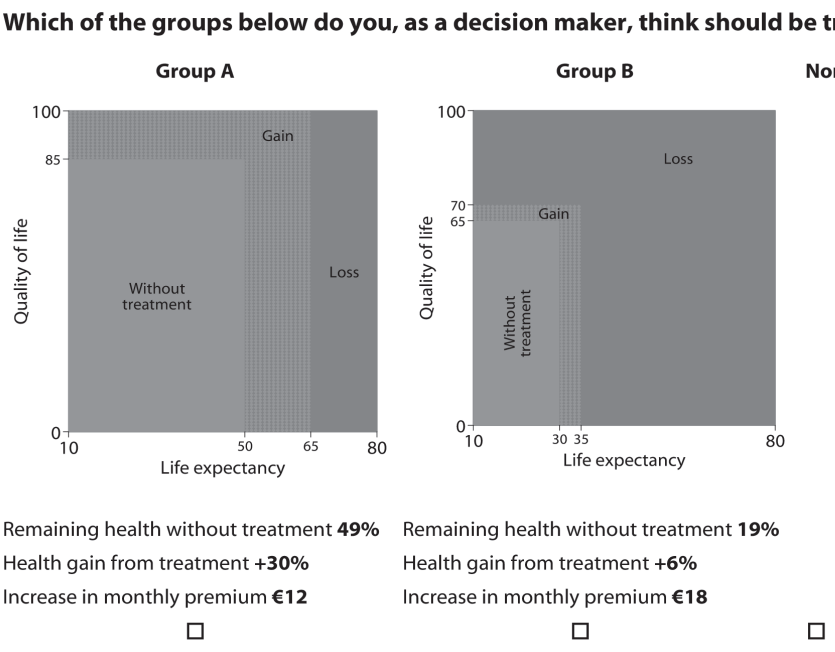


Figure 4.1: Question 1 (Age group 10, version 1, choice set 1)

[3] In the questionnaire the light grey shaded area was colored green, the middle grey shaded area was green-and-red shaded, the dark grey shaded area was colored red.

The attributes, levels and presentation of choice sets were pilot-tested in a small sample of 75 respondents for each age group version. This resulted in adjustment of the level ranges of three attributes: age at death without treatment, gain in life expectancy and costs of treatment. In addition, to improve the clarity of the graphs we added the colours green for remaining health without treatment, red for health loss and shaded green-and-red for potential health gain instead of the blue colours of Lancsar et al. [70].

4.2.2. Questionnaire

Respondents were instructed to imagine themselves being in the position of a decision maker facing allocation decisions in health care. They were then asked to imagine that tomorrow an illness will strike two groups of people from the Dutch population that would have otherwise lived in perfect health until death at 80 years of age. The demographic characteristics of the groups were the same, but the illness and the treatment could affect the groups differently, and the costs of treatment could also differ between the groups. The illness would reduce the length and quality of life of the groups of people. There was a treatment available for each group, which would restore some, or all, of the health loss due to the illness. However, the treatment was not yet included in the basic benefit package. Therefore, it would have to be financed through an increase in the mandatory health insurance premium for all Dutch adult citizens for the period of 1 year. The respondents were asked which of the two groups of people they, as decision makers in the healthcare sector, would prefer to treat. An opt-out option was included in order to get valid WTP values [98].

The program Ngene 1.1 was used to generate efficient multinomial logit designs for the main study. An efficient design minimizes the predicted standard errors of the parameters in order to optimize the information obtained from each choice set. The efficiency of the designs was determined by the D-error, which is the most widely used measure of efficiency [99]. Since the levels of the attributes were adjusted after the pilot study we could not use the estimates of the pilot study as Bayesian priors for the main study, but only the signs of the estimates. Bayesian priors are more robust to misspecification because they optimize on prior distributions instead of on fixed parameters [99].

Since certain combinations of levels of attributes resulted in implausible scenarios, we imposed some constrains in the design (e.g., the gain in life expectancy added to the age at death if untreated could not exceed the maximum age of 80 years). Furthermore, interaction effects between quality of life if untreated and age at death if untreated were included to be able to consider the additional effect of proportional shortfall. For each age group we used 1,000 Halton sequence draws [73].

For each age group, designs with 24 choice sets were generated. The choice sets were divided over three versions using a blocking variable. This resulted in a total of nine blocks (and versions of the questionnaire) each with each eight choice tasks. The alternatives were unlabelled, meaning that the scenarios only varied by the included attributes, and the choice sets were randomized within blocks to avoid order biases in the results. Two control questions were added to each block to detect inconsistent respondents: one dominant choice set was presented as first choice-set in all blocks. In a dominant choice set, the attribute levels of one scenario (the dominant scenario) are superior to the levels of the other scenario (the dominated scenario) on each attribute. Therefore, respondents who carefully consider the choice set may be expected to opt for the dominant scenario. Furthermore, the fifth choice set was repeated as the tenth choice set, but now left and right scenarios reversed. Respondents carefully considering the choice sets are expected to choose the same scenario in both questions, independent of its positioning left or right. Altogether, each respondent received 10 choice tasks for one age group. If a respondent chose the dominated scenario in the first choice (i.e. the first control question) and reversed preferences in the tenth choice (i.e. the second control question), the respondent was removed from the dataset. Furthermore, based on the distribution of completion times in the pilot study and a quickest possible reading and responding test by three researchers we determined a minimum completion time for the ten choice sets of 150 seconds.

In April 2013, the questionnaire was distributed by a professional Internet survey company to a representative sample of the adult population of the Netherlands in terms of gender, age and level of education. The DCE questions were the first part of a larger questionnaire that also contained three contingent valuation questions (as the second part) and questions about socio-demographic characteristics (as the third part). Each respondent was randomly assigned to one of nine versions of the questionnaire (i.e., three age groups times three blocks of choice sets). For an English copy of the questionnaire we refer to the electronic supplementary material.

4.2.3. Analyses

To be able to estimate the MWTP per QALY gain for different levels of proportional shortfall the initial model included the following parameters: total QALY gain, proportional shortfall and the increase in health insurance premium. These parameters were calculated from the original attributes using the following equations:

Eq 4.2 $$Total\ QALY\ gain = (QG * (AD - AO)) + (YG * (QOL + QG))$$

where *QG* represents the gain in quality of life, *AD* represents age of death without treatment, *AO* is age of onset, *YG* is life years gained, *QOL* the quality of life before treatment. Proportional shortfall was calculated using the following formula:

Eq 4.3 $$Proportional\ shortfall = \frac{((MQ - QOL) * (AD - AO) + ((MY - AD) * * 100))}{(MY - AO)}$$

where *MQ* represents the maximum quality of life (100) and *MY* the maximum life expectancy, which was set at 80 years of age.

To determine the social WTP, the QALY gains were multiplied by the size of the risk group and the increase in monthly premium was multiplied by 12 monthly instalments and the number of health insurance payers in the Netherlands (i.e. 13,260,000). The deterministic components of the alternatives in a choice-set were represented by:

Eq 4.4
$$V_A/\lambda_s = \beta_1 QALYGAIN + \beta_2 PS + \beta_3 COST$$
$$V_B/\lambda_s = \beta_1 QALYGAIN + \beta_2 PS + \beta_3 COST$$
$$V_C/\lambda_s = \beta_0$$

where V is the observed component of the random utility function for alternative A, B or C (opt-out), λ_s is the scale parameter and β are the parameters to be estimated. The constant terms represent the expected utility for no treatment over treatment. Likelihood ratio tests were used to test different specifications of the utility functions (categorical or numerical attribute levels and interaction effects between QALY gain and proportional shortfall).

In our attempt to find appropriate explanations for the observed patterns in the data, we estimated numerous models. To allow for preference heterogeneity among the population, panel mixed logit models with correlated coefficients were used to analyse the data. All parameters were included as random parameters. MWTP per QALY values were computed as:

Eq 4.5 $$MWTP_a = \frac{\beta_a}{\beta_{cost}}$$

However, including the cost parameter as a random parameter in mixed logit models may cause problems with respect to the WTP calculations. When a normal distribution for a price coefficient overlaps zero it will result in undefined moments of WTP since dividing by zero is impossible. Furthermore, divisions by numbers arbitrarily close to zero results in very large WTP estimates. Different solutions have been proposed in the literature to tackle this issue, such as WTP space models, mixed models with a fixed parameter for

the cost attribute or constrained distributions like lognormal or triangular distributions [100-102]. All these specifications have been tested for the current models. WTP space models did not fit our data. Different parameter distributions were tested combined with large numbers of Halton draws (i.e. up to 3,000), but we were not able to find a model fit. Therefore, different specifications of the mixed logit model were estimated and compared using Log Likelihood ratio tests and examining the Akaike and Bayesian information criteria. The mixed logit models were estimated with 1,000 Halton draws, the statistical results of this process are presented in Table 4.3. As this table shows, the random parameters with restricted distributions for the costs parameter did not result in better model fits than the specification of a fixed coefficient for the cost attribute. Besides, it should be noted that the specification of a constrained distribution for the cost attribute would still complicate the calculation of the WTP estimates and related confidence intervals. Therefore, in our models cost was specified as a fixed parameter [100-102].

The mixed logit models based on the above mentioned attributes did not behave as expected. As shown later on in the results section, counterintuitive results were found with respect to proportional shortfall (i.e. scenarios with higher proportional shortfall were less likely to be chosen, c.p.). Moreover, all standard deviations of the random parameters were significant which implies a substantial amount of preference heterogeneity within the sample. To further explore these results and understand the preference structure of respondents, we searched for decision patterns within the data. For that reason, we relaxed our assumptions with respect to proportional shortfall and absolute QALY gains to explain respondents' preferences and used the attributes as presented to the respondents instead. Latent class models were estimated to identify different subgroups in the population based on unobserved characteristics that affect their preferences. It is assumed that preferences are homogeneous within the classes but differ between classes [103]. The optimal number of classes was determined by examining the Akaike and Bayesian information criteria of different numbers of classes and the standard errors of the corresponding parameters. The latter is a valid additional argument in this context, because an increasing number of classes may lead to extremely large standard errors of several parameters, complicating the interpretability of the model. Latent class models with four classes showed extremely large standard errors in age groups 10 and 40, and insignificant coefficients -and consequently meaningless WTP estimates- in age group 70. Thus, in all three age groups the number of classes was limited based on the standard errors of the corresponding parameters, despite the fact that accepting more classes would have improved model fit [100, 104].

The results of the latent class models provided additional insights in respondents' preferences compared to the mixed logit models. Therefore, the latent class models were chosen as a starting point for further analyses.

Overall MWTP values were estimated as the weighted average of conditional class MWTPs. Confidence intervals for MWTP estimates were estimated using the Delta method [70, 103], [100].

Analyses were performed in Nlogit 5.0 (Econometric Software Inc.).

Table 4.2: Demographic statistics (n=1,205)

Variable		Mean	SD	Min	Max
Age		45.0	15.0	18	86
Gender (% female)		50.8			
Partner (% yes)		67.0			
Children (% yes)		58.3			
Monthly income (%)	- Group 1 (< €1,000)	23.5			
	- Group 2 (€1,000 - €1,999)	31.5			
	- Group 3 (€2,000 - €3,499)	32.3			
	- Group 4 (≥ €3,500)	12.7			
Education status (%)	- Elementary school	25.5			
	- High school	42.1			
	- University	32.4			
Health status:	-VAS scale (0-100)	80.1	15.0	15	100
Opt-out (%)		10.9			

Note: Number of observations per version of the questionnaire: 411 for age group 10, 410 for age group 40, 384 for age group 70. General population statistics: 45 years of age (18+), 50.9% female (18+) and 33.0% elementary school, 40.2% high school, 26.8% university (15+) http://statline.cbs.nl/Statweb/?LA=en.

4.3. RESULTS

The final dataset included 1,205 respondents, representative of the adult population of the Netherlands with respect to age (mean 45.0 years), gender (50.8 % female) and education level (25.5%, 42.1%, 32.4% had lower, middle, higher education, respectively). Demographic statistics of the sample are presented in Table 4.2. The completion time for the ten DCE questions was, on average, 5.2 minutes.

The results of the panel mixed logit models for the three age groups are presented in Table 4.3. As already briefly discussed in the method section, we strongly questioned whether this model accurately represents respondents' preferences. The results with respect to proportional shortfall were counterintuitive and the standard deviations of

Table 4.3: Results Mixed logit models with QALY gain and proportional shortfall

	Age 10		Age 40		Age 70	
	MNL	MMNL	MNL	MMNL	MNL	MMNL
Health gain from treatment (QALY)	0.092**	0.133**	0.170**	0.243**	0.735***	1.178**
Proportional shortfall	-0.016**	-0.025**	-0.021**	-0.032**	-0.016***	-0.022**
Increase in health insurance premium (€/month)	-0.027**	-0.055**	-0.036**	-0.052**	-0.048***	-0.075**
Constant	-1.251**	-3.324**	-1.693**	-3.875**	-1.554***	-3.053**
SD random parameters						
Health gain from treatment (QALY)		0.007		0.107**		0.810**
Proportional shortfall		0.076**		0.063**		0.066**
Constant		6.366**		5.479**		6.050**
Log-likelihood at convergence	-2723.392	-2244.898	-2590.386	-2193.658	-2466.932	-2054.220

Note: **=p<0.001, *=p<0.1. MNL: multinomial logit model, MMNL: mixed multinomial logit model.

the random parameters were all statistically significant with relatively large coefficients, which suggest a substantial heterogeneity in preferences in the sample.

Table 4.4 presents the results of the mixed logit and latent class models using the attributes as presented to the respondents, that is, health gain as a percentage, remaining health without treatment (%) and the increase in health insurance premium. The mixed logit models were comparable to the mixed logit models of Table 4.3 with respect to preference heterogeneity and counterintuitive results for health state before treatment (that is, an average preference was observed to treat people who already were relatively healthy). Although the mixed logit models had a slightly better model fit than the latent class models, we preferred to use the latent class models since they seem to provide additional insight in the heterogeneous preference structures of the respondents. The results for the selection of number of classes are presented in Appendix 4.1. For all three age groups, the most appropriate model consisted of three classes (as explained in method section).

Respondents belonging to the first latent class of age group 10 had a relatively strong preference *not* to choose between one of the groups of patients as indicated by the positive significant constant term. In case respondents were willing to treat one of the groups of patients, more remaining health without treatment increased the probability to receive treatment. Remarkably, the coefficients of health gain from treatment and remaining health without treatment were comparable in magnitude and sign. This indicates that these respondents did not really differentiate between these two attributes.

Table 4.4: Results mixed logit and latent class models original attributes

	Age 10				Age 40				Age 70			
	MMNL	Latent Class			MMNL	Latent Class			MMNL	Latent Class		
		Class 1	Class 2	Class 3		Class 1	Class 2	Class 3		Class 1	Class 2	Class 3
Health gain (%)	0.110**	0.090**	0.090**	0.051**	0.097**	0.084**	0.087**	0.046**	0.117**	0.075**	0.137**	0.054**
Remaining health without treatment (%)	0.028**	0.070**	-0.035**	0.015*	0.032**	0.082**	-0.013**	0.022**	0.022**	0.052**	-0.046**	-0.001
Increase in premium (€/month)	-0.063**	-0.042**	-0.049**	-0.119**	-0.052**	-0.016*	-0.054**	-0.123**	-0.074**	-0.060**	-0.050**	-0.133**
Constant	-0.494	1.263**	-1.877**	1.286**	-0.752**	1.652**	-1.254**	1.450**	-0.863**	-0.011	-1.646**	0.427
SD Random Parameters												
Health gain (%)	0.059**				0.043**				0.081**			
Remaining health without treatment (%)	0.084**				0.062**				0.066**			
Constant	3.383**				2.931**				2.755**			
Probability of class membership		0.476**	0.407**	0.117**		0.400**	0.496**	0.104**		0.567**	0.304**	0.130**
LL at convergence		-2219.3		-2288.5		-2194.2		-2233.7		-2054.0		-2124.1
WTP/QALY (per class)		€243.635	€208.072	€48.793		€533.015	€160.695	€37.271		€165.784	€366.749	€53.873
WTP/QALY (95% CI)	€197.663	€206.408 (€ 125.993 - € 286.823)			€ 184.893	€296.756 (€ 37.281- €556.233)			€209.588	€212.322 (€ 136.454- €288.190)		

Note: **=p<0.001, *=p<0.1. MMNL mixed multinomial logit model, SD standard deviation, LL log likelihood, WTP willingness to pay.
Appendix 4.1: Selection number of classes in the LCM.

The increase in monthly health insurance premium was the least important attribute in this class. The significant negative constant term in class two of age group 10 indicates a general preference toward treating one of the groups of patients. Respondents belonging to this class were more likely to treat patients with larger health gains and a more severe health states before treatment. Larger increases in monthly health insurance premium decreased the probability to be chosen. Respondents belonging to the third class preferred not to choose between the groups of patients. The increase in health insurance premium had the largest marginal effect on respondents' choice. Probabilities of class membership were 47.6%, 40.7% and 11.7%, respectively.

A similar preference structure was found for age group 40, although the highest probability was to be assigned to class 2 (49.6%), implying a preference to treat patients with more severe health states before treatment.

Somewhat distinct preferences were observed for age group 70. The insignificant constant terms in the first and third classes indicate that respondents did not have a general preference for either choosing between groups of patients, or not. Respondents had a 57% probability to be in first class in which health gain was the most important attribute, followed by the increase in health insurance premium. Respondents belonging to this class preferred to treat patients with a relatively good health state before treatment which is different from what we expected but in line with the other age groups. Respondents had a 30% probability to be in class 2. These respondents were willing to choose between groups of patients and preferred to treat patients with a more severe health state before treatment. Respondents in class 3 seemed to be mainly driven by the increase in health insurance premium in their decision. Remaining health without treatment did not significantly influence respondents' preferences.

The probability weighted MWTP values ranged from € 206,408 in age group 10 to €296,756 in age group 40, but were not significantly different between the age groups. This indicates that we did not find a significant age effect in our data. Interaction effects between health state before treatment and health gain were not significant and therefore not included in the final models. This indicates that, statistically, the value of a health gain was not different for different levels of severity. However, the main effect of severity was significant, which indicates that severity did influence preferences between groups.

4.4. DISCUSSION

It is increasingly recognized that a monetary threshold value against which health gains from an intervention can be evaluated should vary with distributional preferences in society. However, most WTP per QALY studies so far have focused on the individual perspective and have not incorporated such equity considerations. Studies exploring public preferences for QALYs, on the other hand, rarely translate these preferences into equity weights or subgroup-specific QALY values. Therefore, the aim of this study was to contribute to the existing literature by estimating the *social* MWTP for QALY gains in different equity subgroups, considering the equity principles severity of illness (operationalized as proportional shortfall) and fair innings (operationalized as age). Our results show substantial preference heterogeneity among members of the public. As discussed further below, this finding may be helpful in explaining the mixed findings in literature with respect to the value of a QALY in relation to severity of illness and age of care recipients.

Before the results are discussed in more detail, our approach to the data analysis warrants further discussion. A variety of model specifications were tested to analyse the data. Given the aim of this study, levels of proportional shortfall and QALY gains were calculated from the original attributes and included in mixed logit models. Table 4.3 showed substantial preference heterogeneity and counterintuitive results: we found that respondents were less likely to choose patients with higher levels of proportional shortfall. It should be noted that, although counterintuitive, this finding is consistent with Lancsar et al. [70], Dolan and Tsuchiya [105] and Skedgel et al. [103].

In order to better understand how respondents made their decisions, latent class models were estimated with the attributes as presented to respondents. These latent class models demonstrated distinct preference structures in the data, which seem plausible and were helpful in clarifying some of the counterintuitive results we found in the mixed models. It is often suggested that different views exist in society regarding the distribution of health and healthcare [91]. Exploring mean preferences may therefore not be most insightful in the context of such value-laden issues. We suggest that future studies in this area should account for these heterogeneous preferences in society by considering multiple models to explore possible decision patterns underlying the data.

The results of the latent class models (Table 4.4) showed some interesting decision patterns with respect to equity considerations in healthcare allocation decisions, which were more or less consistent across the different age groups.

The first class of each age group showed aforementioned counterintuitive preferences for treating persons who were already in a relatively good health state before treatment

(i.e. less severe diseases). In addition, in the first class of age group 40, respondents reported fairly equal preferences for health state without treatment and health gain (and also in age groups 10 and 70 the differences were relatively small). This might indicate that respondents in this class were driven by the best health state after treatment, irrespective of whether this was a consequence of the health state before treatment or the health gain from treatment. Other studies also have found that respondents consider health state after treatment more important than health state before treatment [22]. However, it is also possible that this finding was (partly) induced by the presentation of the scenarios in our study. A closer look at the graphs of the scenarios (Figure 4.1) shows that the best end state after treatment automatically coincides with the smallest health loss, indicated by the red area in the graph. It is conceivable that some respondents just opted for the smallest health loss (i.e. the smallest red area). Using graphs to clarify the scenarios might thus be helpful in presenting complex choice problems to respondents, but at the same time unintentionally influence their choices. As the use of such graphs is relatively new in this field, this deserves further study, and future studies should be aware of this issue when they consider using graphs to present their attributes to respondents.

The second latent class of all age groups aligned with the principle of proportional shortfall, thus expressing concerns for severity of illness. These respondents were the only ones willing to choose between the groups of patients and, ceteris paribus, preferred to treat patients with a relatively more severe health state without treatment. The probabilities of membership of this class were considerable, which highlights considerable support for considering severity in healthcare priority setting in the general public.

Respondents assigned to class 3, the smallest class of each age group, seemed to consist of individuals with a general aversion to prioritizing patients based on the health characteristics included in the study. The remaining health state without treatment attribute was not significant in age group 70, and only marginally significant in age group 40, suggesting that differences in health state without treatment were not a relevant argument for them to prioritize between different groups of patients. Moreover, the constant term indicated that these respondents generally preferred not to choose between patients, and when they did choose, their decision was mainly driven by the change in monthly health insurance premium.

In other words, in each age group we found two latent classes with a general preference not to choose between patients, and one class that was willing to choose and displayed preferences that aligned with what was expected from the theory of proportional shortfall. The first two classes represented the majority of respondents in all three age groups, but a substantial minority thus supports accounting for severity in priority setting.

Interaction effects between remaining health without treatment and health gain were found not to be significant. This indicates that, statistically, severity did not influence the value of a QALY itself in our sample. Nevertheless, the significant coefficients of the main effects suggest that health state before treatment does influence respondents' choices. However, theoretically, these two cannot be valued separately since a certain health gain is always accompanied by a certain health state before treatment (or proportional shortfall). This suggests that at least indirectly the MWTP for a QALY depends on the health state without treatment. Overall, it seems worthwhile to investigate these preferences with respect to severity in more detail, in particular taking the preference heterogeneity within the general public into consideration.

No clear support was found for the fair innings argument in this study, since the MWTP per QALY estimates did not significantly differ between age groups - although the value in age group 40 appears considerably higher (Table 4.4). The confidence interval of the MWTP estimate of age group 40 was large, which may be due to the low significance of the health insurance premium attribute in the first class. The relatively small coefficient for health insurance premium in this class resulted in a fairly high MWTP for a QALY estimate (€533.015), which in turn (given the substantial probability to be part of group 1) led to a relatively high MWTP estimate for age group 40.

Apart from the common limitations that come with DCEs and online surveys, the following limitations of this study need to be mentioned. First of all, as discussed here, a possible explanation for part of the preference heterogeneity observed in this study might relate to the graphical presentation of the scenarios. Such graphs, also used before by Lancsar et al. [70], Shah et al. [106] and Brazier et al. [107] may unintentionally give room to different interpretations of the scenarios by respondents, and therefore may not be the best way to present the attributes to respondents. How respondents perceive the information contained in such graphs deserves further study, for instance using a think-aloud procedure.

Second, finding that fair innings is of no relevance for the value of a QALY may be a result of framing, since age was part of the scenario description and not an attribute in the choice set. This implies that respondents did not trade age against other characteristics of the recipients, which may have given a different meaning to age in the choices made. In the literature there has been a growing interest in the context and framing of studies in order to improve the consistency and comparability between studies. Our results are in line with those reported by Lancsar et al. [70] and Diederich et al. [108]. It would be interesting for future research to investigate whether a DCE with a fixed level of severity in each scenario and age included as an attribute would result in opposite findings.

Concluding, this study aimed to contribute to the existing literature by bridging the gap between WTP per QALY studies from an individual perspective and the growing literature exploring societal preferences for health and health care. A recent review of Whitty et al. [22] underlined the importance of multi criteria studies and the translation of public preferences into equity weights that can be used for policy making. In this study, we estimated WTP per QALY for different age groups and found no support for the fair innings argument, or for prioritizing based on health characteristics more generally. We did find support for considering severity of illness among a substantial minority of the public, but since interaction terms between health state without treatment and QALY gains were not significant, we cannot say that the WTP per QALY estimates differed statistically significantly for different levels of severity of illness.

While some of our results may be related to the design of our study, including the graphical presentation of the scenarios, they are insightful and, most of all, highlight the importance of accounting for heterogeneity in preferences among the public on value-laden issues such as prioritizing health care, both in research and in decision making.

Acknowledgements

We would like to thank Steef Baeten for preparing the graphical presentation of the scenarios. This study was financially supported by Zorgonderzoek Nederland (ZonMW), Netherlands Organization for Health Research and Development (project number 152002038). The usual disclaimer applies.

APPENDIX 4.1: SELECTION NUMBER OF CLASSES IN THE LCM

Table A. 4.1: Selection number of classes in the LCM

	Number of classes	Log Likelihood	AIC	BIC
Age group 10	1	-2723.39212	5454.8	5479.2
	2	-2473.78587	4965.6	5020.5
	3	-2288.50549	4605.0	4690.4
	4	-2271.26754	4580.5	4696.4
	5	-2225.79214	4499.6	4645.9
	6	-2202.77079	4463.5	4640.4
	7	-2184.51031	4437.0	4644.4
	8	-2178.68425	4435.4	4673.2
	9	-2178.71372	4445.4	4713.7
	10	-	-	-
Age group 40	1	-2590.38614	5188.8	5213.2
	2	-2342.04598	4702.1	4757.0
	3	-2233.72689	4495.5	4580.8
	4	-2186.88956	4411.8	4527.6
	5	-2168.02966	4384.1	4530.4
	6	-2157.85377	4373.7	4550.5
	7	-2152.86299	4373.7	4581.0
	8	-2131.83972	4341.7	4579.4
	9	-2130.30804	4348.6	4616.8
	10	-2120.76759	4339.5	4638.2
Age group 70	1	-2466.93212	4941.9	4966.0
	2	-2224.51394	4467.0	4521.3
	3	-2124.14160	4276.3	4360.7
	4	-2080.06128	4198.1	4321.7
	5	-2052.04820	4152.1	4296.8
	6	-	-	-
	7	-2017.28796	4102.6	4307.6
	8	-1993.56204	4065.1	4300.3
	9	-1994.84877	4077.7	4343.0
	10	-2002.77934	4103.6	4399.0

5

Equity-dependent social willingness to pay for a QALY

Based on:
Bobinac A, van de Wetering E. J., Van Exel N.J.A. & Brouwer W.B.F. (2015). Equity-dependent social willingness to pay for a QALY. *Submitted*

ABSTRACT

Information about the relative cost-effectiveness of healthcare interventions is generally considered important for reimbursement decisions, although other considerations, such as equity, may also play a role. Numerous studies explored if and when it may be justifiable to sacrifice a part of the maximum attainable health to achieve more distributional fairness. Studies that explored which characteristics of health recipients warrant differential treatment suggest that age and severity of illness are the most important differentiators. The current study estimates the monetary value of QALY gains in patients with different illness severity and age profiles (i.e., age- and severity-dependent monetary value of a QALY), using the willingness to pay method (WTP) and directly test the social support for the concept of varying cost-effectiveness thresholds as a tool for evaluating the results of economic evaluations. Severity was operationalized as proportional shortfall and end-of-life concerns.

Results reveal that the size of health gain, proximity to death and age of respondents were most influential determinants of WTP. Severity arguments, although often discussed in the context of an equitable distribution of health care, were found to be less influential than the size of that gain on offer, indicating that respondents may generally lean towards the principle of health maximization in priority-setting. The concerns for patient's age appeared to dominate concerns for severity, with some preference reversals between the age groups. Finally, severity-related preferences were not independent of the age of beneficiaries, which illustrates the complexity of the relationship between different equity concerns.

5.1. INTRODUCTION

In several healthcare systems, information about the relative cost-effectiveness of healthcare interventions is considered important for reimbursement decisions [109]. However, cost-effectiveness is usually not the only criterion used to evaluate the desirability of healthcare interventions. Other considerations, such as fairness in the distribution of health and healthcare may also play a role in deciding which interventions to reimburse [110]. For instance, policy makers may have reasons to favour interventions targeted at children over those targeted at elderly, even when the latter interventions are the most cost-effective options. In health economic evaluations such preferential treatment can be operationalized by assigning more weight to health gains (typically expressed as Quality-Adjusted Life-Years, or QALYs) achieved in such groups [58]. Then, the QALY gains are *directly* weighted according to some scheme of relative weights. Alternatively, the conventionally derived cost-effectiveness ratios can be judged against divergent thresholds, where this threshold is higher for interventions targeting patients (or diseases) that are deemed a political or social priority [67]. In principle, both approaches should lead to identical outcomes.

So far, a multitude of normative and empirical studies explored if and when it may be justifiable to sacrifice a part of the maximum attainable health to achieve more distributional fairness. Studies have explored which characteristics of QALY recipients warrant differential treatment and what relative weight they should receive [56, 105, 111-114]. The results of these studies suggest that age and severity of illness are the most important differentiators, revealing general social support for prioritizing younger patients and patients who are more severely ill [18, 56, 95, 115-121]. The current study aims to contribute to this stream of literature by estimating the monetary value of QALY gains in patients with different illness severity and age profiles, using the willingness to pay method (WTP). The main question of this study is whether WTP per QALY depends on severity- and age-gradients of QALY recipients, and what the nature of this relationship is. To our knowledge, this is the first Contingent valuation study to systematically estimate equity-dependent monetary values of QALY gains.

We estimate the age- and severity-dependent monetary value of a QALY using data obtained from a relatively large representative sample of the general public in the Netherlands, and hence directly test the social support for the concept of varying cost-effectiveness thresholds[4] as a tool for evaluating the results of economic evalu-

[4] Note that the term threshold is used here to denote the social value of a QALY (v), rather than the opportunity costs of spending within a fixed health care budget (k). See e.g. Claxton et al.2011 for more background on both decision frameworks.

ations. The ongoing debates in the policy and academic arena's about the height of the threshold for judging the results of economic evaluations [122] warrants this line of research. While the discussion is ongoing, some countries are already attempting to put a specific operationalization of an equity concept in practice. Two well-known examples are the Dutch concept of proportional shortfall and the special treatment of 'end-of-life interventions' of NICE. In the Netherlands, decision makers have suggested that the cost-effectiveness threshold should explicitly incorporate concerns for the more severely ill, based on the concept of proportional shortfall [20, 66]. NICE recommends to funding of life extending "end-of-life" (EoL) treatments, under specific circumstances, even when these would not normally be considered cost-effective [47,123]. In this study, we will explore the societal support for both operationalization of severity: proportional shortfall and "end-of-life" concerns.

The design of this study is novel in several respects. First, to align with the decision makers' perspective and increase the relevance of our findings, we estimate WTP per QALY values from the social perspective. Past WTP per QALY studies have almost exclusively focused on individual WTP values,[5] although there are concerns about the usefulness of such individual values in healthcare decision making [23]. We define the social value of a QALY as the average WTP per QALY estimate provided by the payers of healthcare services acting as social decision makers, and expressed as the amount of own consumption they are willing to forego in order to fund a QALY gain achieved in an unknown member of society, through an increase in the health insurance premium.

Second, most previous studies exploring trade-offs between different distributional concerns have varied the attributes of interest (e.g., age or severity) while holding the size of the QALY gain constant [13, 126]. Some studies have allowed health gains to vary across choices and found that respondents preferred larger gains above all other attributes of interest [70, 127], apparently taking the amount of health produced as their maximand. In these studies, the variation in size of the gain was typically obtained by varying either length of life (LoL) or quality of life (QoL) [89, 128, 129]. In the current study, the size of QALY gains is allowed to vary both in terms of LoL and QoL, with values ranging from only marginal to very substantial lifetime gains, in different age groups and at various levels of severity. Hence, our design allows a more thorough exploration of equity-related preferences than previous studies.

In the next section we present the analytic framework of our study, discuss the two operationalization of severity in more detail, and formulate our hypotheses. In the methods section we present the study design. Finally, we will present and discuss our results.

[5] With some exceptions [23, 124, 125] which, however, addressed different research questions.

5.2. ANALYTIC FRAMEWORK AND HYPOTHESIS

The web-based survey included diagrams to better illustrate the decision problems to respondents [70, 130, 131]. The diagram in Figure 5.1 illustrates the presentation of a QALY profile in each scenario, with "quality of life" on the vertical axis and "life expectancy" on the horizontal axis. The point of origin is determined by the age of the hypothetical patient group considered in this study at the time the WTP and intervention decision is made (i.e., 10, 40 or 70 years), assuming that the years before the age at origin were spent in full health. Figure 5.1 is an example of a diagram for the group of 10 year olds (10Y). The QALY profile in the absence of illness is the sum of the different shaded areas: the target group would then live till 80 years old in full QoL (100, or U(h)=1). For the 10Y group depicted in Figure 5.1, that would be 70 years in full QoL, i.e. 70 QALYs. The light grey shaded area - demarcated by Age_U and QoL_U - represents the number of remaining QALYs in case of illness, without treatment. The QALY loss if untreated is represented by the sum of the middle grey and dark grey shaded areas (or, in this case, 70 QALYs minus the light grey area). The middle grey shaded area - demarcated by Age_T and

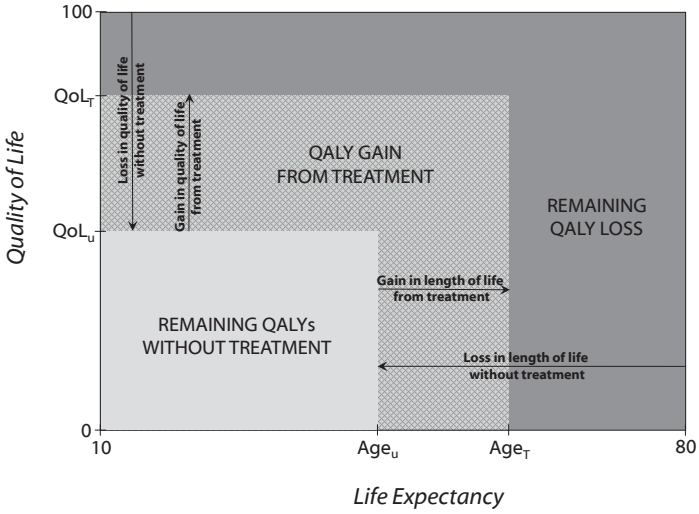


Figure 5.1: Graphical presentation of analytical framework
Note: Suppose that as a result of illness the life expectancy of a 10 year old in the target group is reduced from 80 to 30 years (Age_U = 30) and the quality of life from 1 to 0.65 (QoL_U = 65), effective immediately. If the 10 year old receives treatment, life expectancy increases from 30 to 50 years (Age_T = 50) and quality of life from 0.65 to 0.85 (QoL_U = 85).The light grey shaded area then represents the remaining health after illness without treatment, being 13 QALYs (=20 years * 0.65 QoL). The middle grey shaded area represents the expected QALY gain from treatment, a combination of quality and length of life gains, being 21 QALYs (= (50-30) * 0.65 QoL + (50-10) * (0.85-0.65) QoL). The dark grey shaded area shows the QALY loss that remains after treatment, 36 QALYs (= (80-50) * 0.85 QoL + (80-10) * (1-0.85) QoL). In this example, proportional shortfall is 0.81 (= (70-13) / 70).

QoL_T - represents the QALY gain from treatment. The dark grey shaded area represents the remaining QALY loss after treatment. The diagrams were accompanied by textual explanations of the information contained in the diagrams (as will be discussed further in section 5.3). The note to Figure 5.1 provides a numerical example.

5.2.1. Severity as Proportional shortfall

The concept of proportional shortfall (PS), endorsed by policy-makers in the Netherlands [20], is based on the proportion of remaining lifetime health lost due to illness and, therefore, can also be seen as an intermediate principle between fair innings - prioritizing those with less lifetime health - and prospective health - prioritizing those with the shortest remaining life expectancy [18, 58, 66, 89]. PS is represented as a number on a scale from 0 (no health loss) to 1 (full health loss or immediate death). Because PS is a relative measure, both younger and older individuals can experience a low or a high PS. Under the framework presented in Figure 5.1, PS can be defined as the ratio between the illness-related QALY loss (i.e., the sum of the middle grey and dark grey shaded areas) and the average number of total QALYs enjoyed in the absence of the disease (i.e., the full area of the diagram); see note to Figure 5.1 for numerical example. When severity is defined as PS, we test the following hypotheses:

- **H1**: Within each age group, WTP per QALY gained at lower levels of PS will be significantly lower than the WTP per QALY gained at higher levels of PS.
- **H2**: At each level of PS, WTP per QALY gained in younger patients (e.g., 10Y) will be significantly higher than for QALYs gained in the older patients (40Y and 70Y).

5.2.2. Severity as "End-of-life" concern

NICE has recommend that under certain circumstances it may be appropriate to fund "end-of-life" (EoL) treatments, even when these would not normally be considered cost-effective [47]. These are treatments aimed at small patient groups with short life expectancies, normally of 24 months or less and provide a life extension of at least three months [47]. In terms of our analytical framework, this pertains to patients for whom Age_U is close to the origin of the diagram.

When severity is defined as an EoL concern, we examine whether there is public support for giving higher priority to end-of-life treatments than to other types of treatments and test:

- **H3**: WTP per QALY will be significantly higher for gains situated increasingly closer to the end of a patient's life, keeping the size of the gain constant and providing a life extension of at least three months.[6]

[6] Having in mind that all life extensions in our design concerned extensions of more than 3 months and small patient groups

Finally, we explore whether society attaches higher value to gains in length of life than those in quality of life, independent of the current health state of the recipient, by testing hypothesis 4:

H4: WTP per QALY will be significantly higher for gains in length of life than gains in quality of life.

5.3. METHODS

We conducted a survey in a sample of the adult general public in the Netherlands (n=1,320), representative in terms of age (18+), gender and education, recruited by a sampling company. The survey was pilot-tested in a small representative sample (n=100) to determine the clarity of the tasks, the feasibility of the questionnaire, the range of the payment scales, and the time required to solve the questionnaire in a thoughtful manner. The contingent valuation questions were part of a larger survey that also contained a discrete choice experiment [131], and questions about socio-demographic characteristics.

The survey

In the introduction of the survey, the diagram shown in Figure 5.1 was built up on screen in small consecutive steps, using an animated presentation, to further the understanding of the concepts at hand [70]. Respondents were requested to think of themselves as social decision makers in the area of healthcare and to imagine an illness that would strike the Dutch population the next day, affecting a group of people aged either 10, 40 or 70 years old. Respondents were told that the treatment was currently not covered in the basic benefits package. Hence, if to be reimbursed, it would need to be financed through an increase in the community-rated health insurance premium that all adults in the Netherlands have to pay [132]. The cost of treatment would be equal for all people affected by the illness. The respondents were also told that they themselves could not become ill (i.e., be a part of the affected group), but that other people – including their loved ones – could. This allowed for the standardization of the beneficiary's age across scenarios and age groups and avoided the issue of inter-personal age variation. Next, respondents were asked to decide on an increase in the monthly health insurance premium that, in their opinion as social decision makers, would adequately represent the social value of the treatment. The premium increase was coercive (i.e., mandatory for all adult payers in the Netherlands, including the respondent) for the period of 1 year. By using coercive payments respondents were (arguably) stimulated to think not only of sympathetic benefits achieved in unknown patients but also of the sympathetic

costs caused by the increase in the insurance premium to other payers, thus potentially reducing the problem of double-counting [133].

Contingent valuation questions

WTP was elicited in a linked two-step procedure: a payment scale followed by a bounded direct open ended (OE) question [23, 134]. Respondents were first presented with an ordered low-to-high payment scale of monthly instalments ranging from €0 to €24 (in steps of €1) and asked to inspect the scale from the left side to the right and indicate the maximum amount they would certainly agree to pay. Next, they were asked to inspect the scale from the right side to the left and to indicate the minimum amount they would certainly not agree to pay. This first step was then followed by the OE question, which was bounded by the minimum and maximum value range obtained from the payment scale. When respondents chose €0 as their maximum WTP, they were asked to indicate the main reason for this preference.[7] When respondents choose €24 as the minimum amount on the scale, they were automatically directed to the OE question which was not bounded to any maximum amount.

5.3.1. Scenario design

Each WTP question was based on a single scenario, a QALY profile comprising four main attributes (Table 5.1). A total of 48 scenarios, well-dispersed across severity levels, were designed for each age group using efficient design for DCE in Nlogit [131] (see Appendix 5.1). Two extra scenarios were added for each age group: one scenario in which the QALY gain was particularly marginal in size (scenario 49; see Appendix 5.1) and one scenario in which the health gain equalled 1 full QALY followed by immediate death, which was used to test **H3** (scenario 50; see Appendix 5.1). The scenarios were distributed randomly to respondents, respecting scenario balance.[8]

The at-risk groups of beneficiaries in the Dutch population varied in size (i.e., 8,000 people in 70Y and 2,000 people in 10Y and 40Y) in order to keep the implicit minimum and maximum size of the lifetime QALY gains equal between the three age groups, and thereby the implicit maximum and minimum WTP per QALY values. The implicit minimum and maximum WTP per QALY estimates defined by scenario and payment

[7] Options were: 1) I am unable to pay more than €0; 2) avoiding the worse health state and remaining in the better health state in not worth more than €0 to me; 3) I am not willing to pay out of ethical considerations; 4) something else [with open text field for explanation]. Answer options 1) and 2) were considered as a "true" €0 WTP while option 3) as a protest answer. Option 4) was coded depending on the explanation given.

[8] A respondent was offered the WTP questions for the three age groups in random order, and within each age group a randomly selected scenario from the design. The relative probability a scenario was offered to a respondent decreased with the number of times it had been offered to respondents before.

Table 5.1: Attribute levels

Group	Attributes	Main attribute levels				Additional attribute levels[a]	
10Y	Quality of life before treatment	45	65	85		95	0
	Life expectancy before treatment	30	50	70		10	10
	Quality of life gain	5	15	25	35	1	100
	Life expectancy gain	5	10	15	20	2	1
40Y	Quality of life before treatment	45	65	85		95	0
	Life expectancy before treatment	50	62	74		76	40
	Quality of life gain	5	15	25	35	1	100
	Life expectancy gain	2	6	10	14	1	1
70Y	Quality of life before treatment	45	65	85		95	0
	Life expectancy before treatment	73	76	79		70	70
	Quality of life gain	5	15	25	35	1	100
	Life expectancy gain	0.5	1.5	1	2	10	1

Note: [a] Immediate death (left column) and marginal health gain (right column), these scenarios were not a part of the attributes used to generate scenarios in Nlogit.

scale design ranged equally in the three age groups, from €4,600 to just over €1,600,000. Equal (implicit) maximum and minimum WTP per QALY values facilitated adequate comparisons between the results.[9] Every respondent solved one WTP question for each age group. Each of the 50 scenarios in each target age group was solved by approximately 27 respondents.

5.3.2. Analysis

Before we analysed the data and tested the hypothesis we discounted the value of each element of a QALY profile (e.g., a QALY gain or loss) to its present value. A fixed discount rate *r* was used for each year throughout the time period *t* such that, for instance, a discounted QALY gain *dGQ* for every scenario *i* in age group *y* was:

Eq 5.1 $$dGQ_{t,i} = GS_y \, x \sum_{t=0}^{n} \frac{Gq_{y,i,t}}{(1+r)^t}$$

[9] Because the gains in 70Y were considerably smaller than in the other two groups, the number of 70-year olds within the risk group was increased (from 2,000 to 8,000). An alternative would have been to vary the range of the payment scales between the three age groups (and different average QALY gains). In light of potential biases caused by varying payment scale ranges, we did not opt for this. We did not opt for smaller ex ante probabilities in the 10 and 40 year groups either, in order to avoid small sized gains and cognitive problems of interpreting differences in small probabilities. The equality of group sizes between 10Y and 40Y was a product of the design and partly a product of discounting which, over long stretches of time, reduced the size of the gains particularly in the 10-year olds.

where t=70 in 10Y, t=40 in 40Y and t=10 in 70Y and GS_y denotes the size of the at-risk group in y. A constant discount rate for health of 1.5% was used, as recommended in the Dutch pharmacoeconomic guidelines [135]. However, as a form of sensitivity analyses, we also tested our hypotheses using the undiscounted QALY values and a higher discount rate of 3.5% (as conventional in the UK setting) [136]. In the remainder of the text, values are discounted with a 1.5% discount rate, unless explicitly mentioned otherwise.

Furthermore, the sensitivity of scope of (OE)WTP was analysed as proposed in Bobinac et al. [94], by testing whether larger QALY gains receive statistically higher WTP values, and whether the relationship between the size of the gain and OE(WTP) was near-proportional. The theoretical validity of WTP estimates was tested using log-linear regressions, for two reasons: ease of interpretation in terms of constant elasticity and the correction for non-normality in the distribution. The covariates were tested for non-linearity (using quadratic terms) and relevant interactions (excluded if insignificant). Statistical differences were tested using the parametric t-test on log-transformed WTP estimates and the non-parametric Mann-Whitney u-test. The distributional properties of WTP estimates were analysed using Kurtosis and Shapiro-Wilk tests.

To test hypotheses **H1** and **H2**, discounted proportional shortfall (*dPS*) for every scenario *i* in age group *y* was calculated as:

Eq 5.2 $$dPS_{y,i} = 1 - \frac{dLE_y - dLq_{y,i}}{dLE_y}$$

where dLE_y represents the discounted number of QALYs enjoyed by people of the same age and gender in the absence of disease and $dLq_{y,i}$ represents the discounted QALY loss due to illness in scenario i.

To test hypothesis **H3** with respect to preferences for end-of-life treatments, we used two sets of scenarios: (1) the scenarios where respondents valued a QALY gain that prevents immediate death in all three age groups (i.e., scenario 50 in 10Y, 40Y and 70Y and scenario 6 in 40Y; see Appendix 5.1); and (2) the scenarios in 70Y where treatments offered life extension of <2 years versus >2 years (regardless of the HRQOL)[47].

To test hypothesis **H4**, gains in life expectancy and quality of life were included in the regression analysis separately, instead of the combined QALY gain.

The average social WTP per QALY in each age group *y* was calculated as:

Eq 5.3 $$WTP\ per\ QALY_y = \frac{OE(WTP)_{y,i,k}}{dGq_{y,i}} \times 12 \times 13{,}260{,}000$$

where $OE(WTP)_{y,i,k}$ represents the open-ended WTP estimate provided by respondent *k* in scenario *i* and age group *y*, and $dGQ_{y,i}$ represents the discounted QALY gains from treatment. This average of ratios is multiplied by the number of 1-year monthly instalments (i.e., 12) and the approximate number of health insurance payers in the Netherlands (i.e. 13,260,000).[10]

Analyses were performed in STATA13.

5.4. RESULTS

The pilot study estimated the minimal time required to carefully read and solve the WTP questions to be minimally 3.5 minutes. We therefore discarded the responses from subjects who solved the questions in less than 3.5 minutes, yielding final average response time of 6.6 minutes. In total, the answers of 1,320 respondents were retained in the sample. The average household monthly income (before tax) of €2,430 per average number of 2.5 household members adequately represented Dutch national figures (Table 5.2) [137].

Table 5.2: Summary statistics

Variable		Mean	SD	Min	Max
Age		46	15.5	18	95
Gender (% female)		50			
Partner (% yes)		68			
Children (% yes)		57			
Personal before-tax income		2,430	1,619	999	10,000
Income groups (% yes)	- < €1,000	25			
	- €1,000 - €1,999	31			
	- €2,000 - €3,499	29			
	- ≥ €3,500	15			
Number of people living on household income		2.49	1.08	1	20
Education status (% yes)	- Elementary school	12			
	- High school	54			
	- University	34			
Health status:	- EQ-5D (Dutch tariff)	0.85	0.20	-0.32	1
	- EQ-VAS	79.3	16.1	10	100
Completion time of the questionnaire (seconds)		394	150	210	1340

[10] Number of persons over >18y in the Netherlands, minus the number of defaulters (CBS, 2015, 2006).

5.4.1. WTP estimates

Most OE (WTP) fell within the range of €0 to €15. Only 0.73% of respondents indicated the highest level offered on the payment scale (€24) as their maximum WTP. Few respondents indicated €0 WTP (i.e., 15 in 10Y, 13 in 40Y and 35 in 70Y) and no respondents indicated €0 WTP in all three WTP questions. The reasons for stating zero WTP were uniformly distributed among the explanations (see footnote 4). No relevant relationship between the size of the health gain, household income and zero WTP was observed. Although the average size of dGQ and dPS were fairly similar between 10Y, 40Y and 70Y, as envisaged by study design (Appendix 5.2), the OE(WTP) values significantly and systematically differed between the age groups, with $OE(WTP)_{10y} > OE(WTP)_{40y} > OE(WTP)_{70y}$ ($p<0.05$; Table 5.3). The same pattern was observed for the discounted WTP per QALY values, with QALY gains achieved in 10Y valued highest at €216,600, followed by €173,900 in 40Y and €118,500 in 70Y ($p<0.05$).

Table 5.3: OE(WTP) and WTP per QALY values

Discount rate	Age group	Average OE(WTP)[a]	Average WTP per QALY	Average WTP per QALY in scenario 50[b]
r=1,5%	10Y	€13,1	€216,600	€ 819,000
	40Y	€11,6	€173,900	€ 750,300
	70Y	€ 9,0	€118,500	€ 175,700
r=3,5%	10Y	€13,1	€628,000	
	40Y	€11,6	€256,300	
	70Y	€ 9,0	€143,700	
r=0%	10Y	€13,1	€ 95,700	
	40Y	€11,6	€151,000	
	70Y	€ 9,0	€115,900	

Note: WTP values are rounded to hundreds. [a] OE(WTP) presented here are reported as monthly averages. [b] One full QALY followed by immediate death.

Sensitivity of scale

Figure 5.2 depicts the relationship between the size of the QALY gains (dGQ) and OE(WTP). It reveals a concave relation between the value of a QALY and the size of dGQ, in all three age groups. If sensitivity to scale as a higher-order condition for validity was defined in terms of a near-proportional relationship between the size of the gain and WTP [94], then the sensitivity to scale was higher at smaller QALY gains. All other things equal, Figure 5.2 also reveals a tipping point after which the value of additional QALYs gained in a single patient became negative, especially in the 70Y group where the inverse U-shaped function exhibited a sharp decline in value. The tipping point in 10Y and 40Y was around 13 dGQ. Hence, although from a social perspective we found that

the gains in younger patients are deemed more valuable throughout, as they received higher (OE)WTP at all levels of QALY gains, we also found that the extra value is diminishing with the size of QALY gains. This suggests that restoring patients to high levels of post-treatment health, after a point, was not valued differently than restoring patients to lower levels of post-treatment health. The inverse U-shaped relationship between OE(WTP) and QALY gains was not the consequence of study design (nor discounting), since smaller and larger QALY gains were fairly normally distributed in all age groups. In fact, using different methods and designs, Lancsar et al. [70] , Norman et al. [72] and Brazier et al. [107] observed similar patterns.

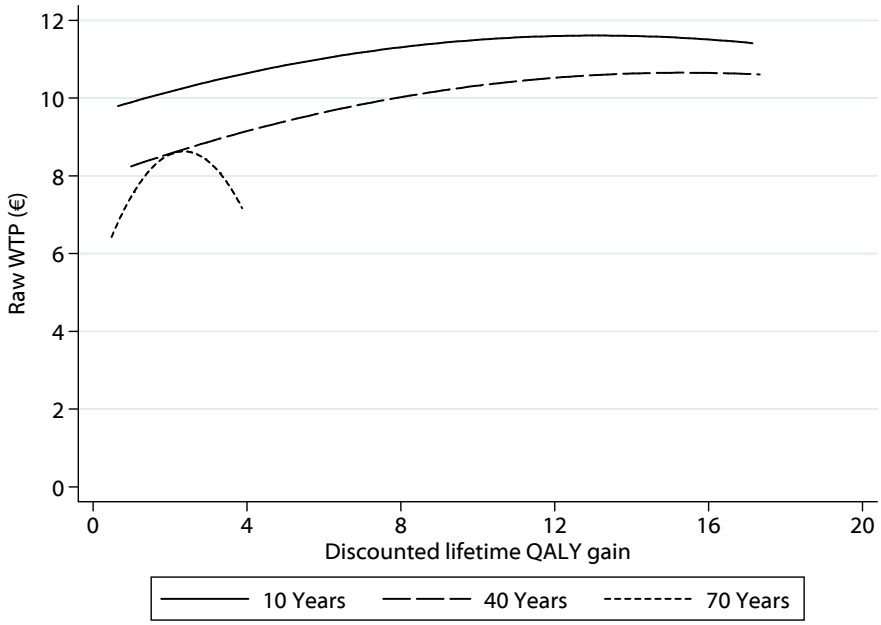


Figure 5.2: OE(WTO) by dGQ for 10Y, 40Y and 70Y

5.4.2. Severity as proportional shortfall – testing H1 and H2

When accounting for the level of severity at which QALYs were gained, generally we found no important relationship between PS and WTP per QALY (rejecting **H1**). First, in regression analyses (Model 1; Table 5.4), PS was found to be a significant predictor of WTP per QALY only in 70Y ($p<0.05$), also when the non-linearity between the OE(WTP) and QALYs was modelled using a quadratic term. We tested the hypothesis that the level of PS and the size of the gain were evaluated in combination and quantified this effect through an interaction term (Model 2). While the level of PS and the interaction term were both statistically significant in 10Y, they were insignificant in 40Y and 70Y.

Importantly, income was found to be a non-significant explanatory variable, likely due to the small monthly OE(WTP) amounts (not limited by respondents' ability to pay) and the social decision maker perspective framing.

The graphical presentation of the relationship between WTP per QALY and dPS (Figure 5.3) reveals a weak U-shaped relationship in 10Y, and almost no relationship in 40Y and 70Y.[11] Combined data from all age groups showed a very similar pattern. However, Figure 5.3 again reveals evidence of ageism similar to that presented in Table 5.3 and Figure 5.3: at each level of PS, WTP per QALY gained in younger patients is valued more, hence confirming **H2**.

Table 5.4: The impact of the most relevant predictors on raw log(WTP) estimates

Age group	Variable	Model 1[a]				Model 2[b]			
		Beta	s.e.	P	St Beta	Beta	s.e.	P	St Beta
10Y	dPS	0.03	0.12	0.83	0.06	0.33	0.18	0.07**	0.08
	dGQ	0.02	0.06	0.00*	0.08	0.06	0.02	0.00*	0.28
	dPS*dGQ	-	-	-	-	-0.07	0.03	0.03*	-0.24
	Higher education	-0.00	0.05	0.97	-0.00	-0.00	0.05	0.91	-0.00
	Log income	0.03	0.04	0.42	0.023	0.03	0.04	0.44	0.02
	Constant term	1.77	0.34	0.00*	-	1.6	0.34	0.00*	-
40Y	dPS	0.14	0.11	0.34	0.18	0.12	0.22	0.59	0.03
	dGQ	0.03	0.00	0.00*	0.13	0.03	0.03	0.33	0.11
	dPS*dGQ	-	-	-	-	0.01	0.04	0.89	0.02
	Higher education	-0.03	0.05	-0.56	-0.02	-0.03	0.05	0.59	-0.02
	Log income	0.07	0.04	1.52	0.04	0.06	0.04	0.13	0.04
	Constant term	1.24	0.33	0.00	-	1.27	0.36	0.00*	-
70Y	dPS	0.29	0.12	0.02*	0.06	-0.05	0.31	0.88	-0.01
	dGQ	0.05	0.03	0.08**	0.05	-0.08	0.11	0.46	-0.8
	dPS*dGQ	-	-	-	-	0.22	0.18	0.22	0.15
	Higher education	-0.04	0.05	0.52	-0.02	-0.03	0.05	0.55	-0.02
	Log income	0.04	0.05	0.43	0.02	0.04	0.05	0.43	0.02
	Constant term	1.25	0.35	0.00*	-	1.45	0.39	0.00*	-

Note: [a] R^2 was 0.013 for 10Y, 0.019 for 40Y, and 0.005 for 70Y. [b] R^2 was 0.012 for 10Y and 40Y, and 0.008 for 70Y. dPS = discounted proportional shortfall, dGQ = discounted QALY gain. *=p<0.05, *=p<0.1.

[11] When the same relationship is tested using regression analyses, we find a significant coefficient of PS (negative sign) and PS squared (positive sign) in 10Y, and insignificant PS coefficients in 40Y and 70Y.

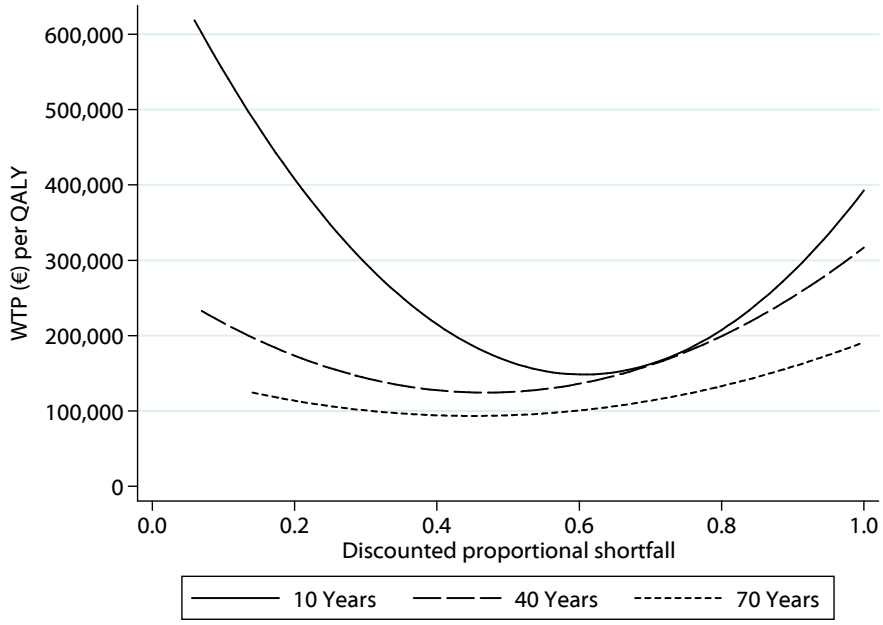


Figure 5.3: WTP per QALY estimates at different levels of proportional shortfall and age

5.4.3 Severity as the end-of life concern - testing H3

The end-of-life argument received reasonable support in our data. The last column of Table 5.3 shows that the value of 1 QALY increased four-fold in 10Y and 40Y in the context of immediate death compared to other type of QALY gains, but only about 50% in 70Y. Table 5.5 shows the results of the effect of the end-of-life argument on the WTP per QALY in age group 70. The further the patient was from her death the smaller the WTP per QALY, thus confirming **H3**. However, judging from the linear regression (Table 5.5), there seemed to be a threshold since the value of a QALY did not differ significantly for gains that were achieved within three years before death if untreated (not 2 years, as proposed in the UK). That is, if the patient will die from the illness if untreated within three years' time, it makes little difference as to *when* within these three years death will occur. If the illness strikes after 3 years or later, the value of a QALY decreases substantially (Table 5.5).

5.4.4. Composition of the QALY gain - testing H4

The variation in WTP per QALY was also investigated with respect to the composition of QALY gains. When controlling for the position of the gains in terms of the remaining QALY profile without treatment, we found that the gain in LE determines the value of a QALY strongest, while the gain in terms of QOL generally was not a significant predictor of

Table 5.5 WTP variance per QALY accounting for LE without treatment in age group 70

Variable	WTP per QALY in 70Y[a]	Model 1 [b]			
		Beta	s.e.	P	St Beta
LE without treatment					
70y	€ 176,000	Omitted			
71y	€ 147,000	0.14	0.28	0.63	0.02
73y	€ 141,000	-0.2	0.2	0.3	-0.09
76y	€ 110,000	-0.4	0.2	0.04*	-0.19
79y	€ 90,000	-0.58	0.05	0.00*	-0.24
Higher education		0.00	0.6	0.99	0.00
Log income		0.06	0.05	0.21	0.04
Constant term		11.1	0.43	0.00*	-

Note: [a] in absolute values, rounded to hundreds. [b] R^2=0.02, * =p<0.05.

OE(WTP) values (models not presented here). This finding confirms **H4** and is at variance with the conventional QALY model (based on individual rather than social preferences), in which the composition of the QALY gain should be irrelevant. This result suggests that, in social decision making, QALYs could be weighted based on their composition. In fact, previous studies also showed that life-extending and quality-of-life-enhancing QALYs are regarded differently [107, 138, 139].

5.5. DISCUSSION

The main focus of this study was to estimate the social willingness to pay for QALY gains at different levels of disease severity, in beneficiaries of different ages. We explored two well-known examples of operationalizations of severity that are already put in practice, the Dutch concept of proportional shortfall [20] and the special treatment of 'end of life interventions' of NICE [47].

The evidence in support of PS as a predictor of WTP per QALY can be characterized as sporadic. Most support for PS was found when gains were relatively small in size. As the size of QALY gains increased, the level of PS ceased explaining the variance in WTP per QALY. Based on the results of this study, the proposition of a strictly increasing cost-effectiveness threshold for increasing levels of PS (CVZ, 2009) is refuted. Support for PS when gains are small in size is in line with Rowen et al. (2014), who found some support for burden of illness (absolute instead of relative severity) and equity-dependent social value of a QALY. Their findings (like ours) were not consistent across all situations, suggesting that preferences over different equity considerations are likely tied to particular contexts and thus not uniform. We tested several potential explanations for the sporadic

relationship between PS and WTP. It is conceivable that a very low QoL after treatment may relate *negatively* to WTP per QALY estimates, potentially confounding the results. In other words, interventions that offer extensions in life expectancy but leave patients in very low QoL after treatment may receive lower priority. However, even when only the subgroup who returned to health above 40%, 60% and 80% was analysed (hence the groups with the highest post-treatment QoL), the support for the differentiation based on the PS argument remained scarce. We found that different "types" of QALY gains (either life-extending or quality-of-life-enhancing QALYs) were valued differently, a finding that is also consistent with earlier studies [107, 138, 139].

We did find support for the end-of-life premium. QALY gains situated close to death (<3 years) were valued the highest and gaining a QALY in a situation of immediate death was valued highly. Similar findings were reported by Pennington et al. [140], Rowen et al. [130] and Baker et al. [126]. However, many other studies report no evidence to support this argument [106, 141].

Overall, our results may imply that severity-related issues were *less* important than the size of that gain on offer and that respondents generally leaned towards the principle of health maximization in priority-setting and were reluctant to introduce severity-related considerations when prioritizing among beneficiaries. However, concerns for age appeared to dominate concerns for severity, with some preference reversals between the age groups.

The findings with respect to age warrant some further discussion. It seems that the differentiation of the social value of a QALY largely depends on the recipient's age, justifying our initial decision to separate the exploration of the preferences for the severity arguments by the age of the beneficiaries. Gains in younger patients were systematically valued higher than gains in the older population. Our finding that the value of a QALY decreases as the age of beneficiaries increases is also consistent with earlier studies [66, 78]. Next to the higher monetary values assigned to younger patients' QALY gains, ceteris paribus, our results show that the impact of patients' age on preferences is very profound and appears in various forms. That is, the conclusions regarding the preferences for the severity of patient's condition seem to depend on patient's age, just like the conclusions regarding the sensitivity of OE(WTP) estimates to the scale of the QALY gains. For instance, adding a life-saving QALY to a 70-year old patient is only about 50% more valuable than adding other type of QALY gains in other circumstances, while it is 400% more valuable in 10Y and 40Y. These results indicate that the severity-related preferences are not independent from the age of beneficiaries and are very complex.

Finally, the (OE)WTP values reported here were sensitive to scope, since the results showed that respondents assigned higher monetary value to larger QALY gains, al-

though at a declining rate. The diminishing relationship between the OE(WTP) and QALYs is not likely to be the result of the study design, since smaller and larger discounted QALY gains were fairly normally distributed in all age groups. In addition, using different methods and study designs, Lancsar et al. [70], Norman et al. [72], Brazier et al. [107] and Rowen et al. [130] observed similar patterns.

Although it is difficult to compare the results of stated preference methods due to differences in design, the estimates of social WTP per QALY (r=1,5%) reported here average at 170,000 Euro per QALY, which is about 3,7 times the Dutch GDP per capita (World Bank, 2015), in the ballpark of the values suggested by WHO CHOICE (2015). The 170,000 Euro per QALY is higher than values obtained in previous studies directly exploring the value of a QALY [67, 142]. However, a recent review of Ryen and Svensson [143] reported a wide spread range of WTP estimates of 1,000 to 4,800,000 Euro per QALY, depending on the methodology, country of study, perspective taken and sample population. Because valuation of QALYs and preferences for a fair distribution of healthcare are complex, it need not surprise that there is a lot of variation in the reported results across studies. According to our results, people seem to hold different preferences over the trade-off between severity-related arguments and the size of QALY gains, depending on the age group these gains are realized in. This further highlights the complexity of estimating equity-weights for QALYs and the complexity of the task faced by policy-makers and researchers. This may partially explain the non-explicit-weighting position currently adopted by HTA agencies and governments around the world [70], along with the unwillingness to define exact thresholds.

Several limitations need to be mentioned. First, it could be argued that our study design may have stimulated strong age-related preferences. This may be based on the fact that age was easier to differentiate than severity, since that is how WTP questions were posed (i.e., three separate questions for three age groups). However, the size of gain varied considerably in each age group. Hence, simply stating a lower OE(WTP) for older people irrespective of the size of the gain presented in a scenario, would not have led to such a consistent relationship between the size of the gain, raw WTP and age. Secondly, this was a hypothetical WTP exercise, and hence our results are only indicative of actual consumption behaviour and underlying preferences that may be observed in, for instance, real-life choices or a field experiment. An online format of the survey likely did not contribute to creating a real-life contingent market and payment situation. These are typical problems related to preference elicitation, and not unique to our study. Thirdly, it could be argued that our results were partly driven by the substantial amount of information that respondents were required to account for in their valuations, which some respondents may have found difficult to process, hence focusing only on a subset of the information when making their choices – mainly the size of the gain. The finding

that life expectancy after treatment was the single most relevant point on the diagram may support this claim, although diagrams and animations were designed to explain and enhance the understanding of the task. On the other hand, respondents *were* able to distinguish between the size of the gains on offer in the theoretically predictable way (increasing yet diminishing relationship between OE(WTP) and QALYs), indicating that they apparently did understand the task and were able to capture the relevance of the information presented. Moreover, although the size of the gain was allowed to vary in our study, this did not hamper respondents to express their concern for the younger patients over the older, again indicating that respondents understood the task. This was not the case for level of PS or BoI, indicating that these preferences may indeed be a reflection of true preferences and not an artefact of the design or misunderstanding of the task by respondents.

As proposed by Bobinac et al. [144], future health profiles and gains were discounted in order to reduce the value of an element of a QALY to its present value, reflecting the time preferences of society. Discounting is a much debated topic in the health economics literature, since it may have a considerable impact on the outcome of economic evaluations [68]. Given this ongoing discussion in the literature, we also analysed undiscounted data (Appendix 5.3). Except for the absolute WTP per QALY values, the undiscounted results showed similar patterns with respect to the relative QALY values for different operationalizations of severity and age and thus do not change our main conclusions. Note that the undiscounted values show relatively smaller WTP per QALY values in 10Y than in 40Y and 70Y. This may be an artefact of the design, since the design was based on the equality of population-wide QALY gains between the three age groups, where QALYs were discounted at 1,5% rate. This resulted in a very large size of undiscounted life-time QALY gains in 10Y while the payment scale was held constant.

The way in which the societal perspective was applied in this study deserves some discussion. From the regression models, it seems that applying a societal perspective in this particular form (i.e., valuing a gain in an unknown member of society, using a coercive increase in health insurance premium as the payment vehicle) renders own income and personal traits insignificant. On the one hand this is reasonable, given that respondents were required to think as societal decision makers and make decisions on behalf of society. Moreover, the societal perspective as defined in this study separates own monetary contribution from the good being valued in the sense that this good (QALY gain) cannot be consumed by the payer directly. We chose to define the societal perspective as not inclusive (which would have seen the payer as a part of the risk group). The intention of this study was to stay close to the perspective of social decision makers, who make decisions that usually do not directly concern themselves, at least not at the same point in time. Neither the theoretical nor the empirical literature has seen a lively debate on

the influence of different perspectives on WTP estimates, nor about the specification of the "social perspective" within CV studies (e.g., which payment vehicle can be used in which contexts, or how best to define the health insurance premium vehicle, how to calculate the social WTP, etc.). We defined the social value of a QALY as the amount of own consumption individuals are willing to forego in order to contribute to a health gain achieved in society, through an increase in their health insurance premium. However, it is possible to think of other definitions of social value and further research in this area seems warranted.

Concluding, this paper has further investigated the social valuation of QALY gains in different patient groups. Size of health gain, proximity to death and age of respondents were most influential in this study. Severity arguments, although often discussed in the context of an equitable distribution of health care, were found to be less influential. Moreover, the underlying preferences for different health gains appear to be complex and context dependent. This warrants further investigation of this important topic.

Acknowledgement

The authors wish to thank the Netherlands Organization for Health Research and Development for financing this study (ZonMW grant number 152002038) and Steef Baeten for his invaluable assistance with the design of the graphs for the experiment.

APPENDIX 5.1: ATTRIBUTES AND LEVELS PER AGE GROUP

Table A.5.1: Attributes and levels per age group

id	10Y					40Y					70Y				
	QoL	LE	QoL^G	LE^G	PS	QoL	LE	QoL^G	LE^G	PS	QoL	LE	QoL^G	LE^G	PS
1	85	50	15	15	51%	85	50	5	6	79%	45	73	15	0,5	86%
2	45	50	5	10	74%	65	74	25	2	45%	85	76	15	2	49%
3	45	70	25	5	61%	65	62	35	14	64%	85	76	5	0,5	49%
4	65	70	5	5	44%	65	62	35	2	64%	65	76	15	2	61%
5	65	50	25	5	63%	45	62	15	10	75%	45	73	25	2	86%
6	45	30	25	10	87%	0	40	80	2	100%	65	79	15	1	41%
7	65	30	15	20	81%	65	62	5	2	64%	45	76	5	1	73%
8	65	30	25	20	81%	95	76	5	4	14%	45	73	35	1	86%
9	65	50	5	15	63%	45	50	15	6	89%	85	79	15	1	23%
10	45	30	35	15	87%	85	50	5	2	79%	65	73	5	0,5	80%
11	85	70	5	5	27%	65	50	15	14	84%	45	79	25	0,5	59%
12	85	70	15	10	27%	65	62	5	6	64%	65	73	35	1,5	80%
13	65	50	35	20	63%	85	50	15	14	79%	85	79	5	0,5	23%
14	85	30	5	15	76%	65	62	15	14	64%	65	79	35	0,5	41%
15	45	30	15	10	87%	45	62	5	14	75%	45	76	35	1,5	73%
16	45	70	15	10	61%	85	50	5	10	79%	65	73	25	2	80%
17	45	30	35	20	87%	85	74	15	6	28%	45	71	40	1	96%
18	65	50	35	10	63%	45	74	35	6	62%	65	76	25	0,5	61%
19	45	50	5	15	74%	65	50	25	10	84%	45	76	25	1	73%
20	85	30	5	5	76%	85	74	15	2	28%	65	76	5	1,5	61%
21	45	50	25	20	74%	45	62	25	10	75%	65	73	35	1	80%
22	65	50	15	5	63%	45	50	35	2	89%	45	76	15	1,5	73%
23	45	70	35	10	61%	45	74	25	2	62%	85	73	5	2	74%
24	85	30	15	5	76%	45	50	5	10	89%	45	73	5	1,5	86%
25	65	30	5	5	81%	65	62	15	10	64%	45	79	35	1	59%
26	85	70	15	5	27%	85	62	15	6	53%	45	79	5	0,5	59%
27	85	50	5	10	51%	45	74	15	6	62%	45	79	25	1	59%
28	45	50	25	15	74%	85	74	5	6	28%	45	73	35	0,5	86%
29	45	30	5	20	87%	65	50	25	6	84%	85	73	15	1,5	74%
30	65	50	35	15	63%	45	50	5	2	89%	85	73	5	1,5	74%
31	85	70	5	10	27%	95	76	5	4	14%	65	73	25	1,5	80%
32	45	50	35	5	74%	65	74	35	2	45%	65	76	25	1,5	61%
33	45	70	35	5	61%	85	62	5	14	53%	65	76	5	0,5	61%
34	85	50	15	20	51%	45	50	25	14	89%	45	76	35	2	73%
35	85	30	15	10	76%	85	74	5	2	28%	65	76	5	1	61%

Table A.5.1: Attributes and levels per age group (continued)

id	10Y					40Y					70Y				
	QoL	LE	QoL^G	LE^G	PS	QoL	LE	QoL^G	LE^G	PS	QoL	LE	QoL^G	LE^G	PS
36	65	30	35	20	81%	85	50	15	10	79%	85	79	5	1	23%
37	45	30	15	15	87%	65	62	5	10	64%	65	76	15	1	61%
38	45	50	15	20	74%	45	50	35	10	89%	85	73	5	1	74%
39	65	50	25	15	63%	65	50	35	10	84%	65	73	15	2	80%
40	45	30	25	5	87%	45	50	15	2	89%	95	76	5	0,5	43%
41	65	50	25	10	63%	45	74	5	2	62%	65	76	35	2	61%
42	45	70	5	5	61%	85	62	15	2	53%	45	73	15	2	86%
43	65	30	35	10	81%	45	62	35	14	75%	65	79	35	0,5	41%
44	65	70	25	10	44%	45	62	25	6	75%	85	73	15	0,5	74%
45	65	30	15	15	81%	65	50	5	14	84%	45	76	25	2	73%
46	85	30	5	20	76%	65	74	15	6	45%	85	79	5	1	23%
47	95	78	5	2	8%	65	50	35	6	84%	45	73	25	1	86%
48	65	70	5	10	44%	65	62	25	2	64%	85	76	15	0,5	49%
49	95	79	5	1	6%	95	79	5	1	7%	95	79	5	1	14%
50	0	10	100	1	100%	0	40	100	1	100%	0	70	100	1	100%

Note: QoL = quality of life before treatment; LE = life expectancy before treatment; QoL^G = quality of life gain from treatment; LE^G = life expectancy gain from treatment; PS = proportional shortfall.

APPENDIX 5.2: THE EQUALITY OF THE MAIN PARAMETERS TO BE ESTIMATED, FOR EACH AGE GROUP

Table A.5.2: The equality of the main parameters to be estimated, for each age group

Age group	Variable (average values)	Mean	SD	Min	Max
10Y	OE(WTP)	13,1	39,5	0	1000
	dGQ	7,5	4,2	1	17
	Risk group	2000			
	GQ	15,6	7,5	1	34
	dPS (in %)	0,6	0,2	0	1
40Y	OE(WTP)	11,6	39,3	0	999
	dGQ	7,3	3,8	1	17
	Risk group	2000			
	GQ	8,9	4,6	1	22
	dPS (in %)	0,6	0,2	0	1
70Y	OE(WTP)	9,0	28,2	0	999
	dGQ	1,9	0,9	0	4
	Risk group	8000			
	GQ	2,0	1,0	1	4
	dPS (in %)	0,6	0,2	0	1

APPENDIX 5.3: UNDISCOUNTED RESULT

Table A.5.3.1: The impact of the most relevant predictors on raw log(WTP) estimates

Age group	Variable	Model 1				Model 2			
		Beta	s.e.	P	St Beta	Beta	s.e.	P	St Beta
10Y		R^2=0.0100				R^2=0.0116			
	PS	0.14	0.11	0.20	0.04	0.41	0.22	0.06	0.10
	GQ	0.01	0.00	0.00	0.09	0.03	0.01	0.03	0.23
	PS*GQ	-	-	-	-	-0.02	0.02	0.15	-0.17
	Higher education	-0.01	0.05	0.88	-0.00	-0.01	0.05	0.84	-0.01
	Log income	0.04	0.04	0.42	0.02	0.03	0.04	0.44	0.02
	Constant term	1.67	0.34	0.00	-	1.52	0.35	0.00	-
40Y		R^2=0.0190				R^2=0.0193			
	PS	0.17	0.10	0.10	0.05	0.06	0.22	0.80	0.01
	GQ	0.02	0.01	0.00	0.12	0.01	0.02	0.63	0.06
	PS*GQ	-	-	-	-	0.02	0.03	0.54	0.08
	Higher education	-0.03	0.05	0.58	-0.02	-0.03	0.05	0.61	-0.01
	Log income	0.07	0.04	0.13	0.04	0.07	0.04	0.13	0.04
	Constant term	1.23	0.34	0.00	-	1.31	0.36	0.00	-
70Y		R^2=0.0072				R^2=0.0085			
	PS	0.03	0.13	0.02	0.07	-0.06	0.31	0.85	-0.01
	GQ	0.05	0.03	0.07	0.05	-0.09	0.11	0.42	-0.09
	PS*GQ	-	-	-	-	0.22	0.17	0.20	0.16
	Higher education	-0.04	0.05	0.52	-0.02	-0.03	0.05	0.54	-0.02
	Log income	0.04	0.05	0.43	0.02	0.04	0.05	0.43	0.02
	Constant term	1.24	0.36	0.00	-	1.47	0.40	0.00	-

Table A.5.3.2: Explaining the variance in WTP per QALY values by accounting for the level of LE without treatment (i.e., end-of-life argument) in age group 70Y

Variable	Beta	s.e.	P	St.Beta	Descriptive analysis
Log(WTP per QALY)	R^2=0.03				WTP per QALY in 70Y, in absolute values, by LE without treatment[a]
70y	Omitted				€200,500
71y	-0.09	0.27	0.73	-0.01	€181,500
73y	-0.29	0.20	0.14	-0.14	€220,500
76y	-0.48	0.20	0.02*	-0.19	€162,000
79y	-0.69	0.20	0.00*	-0.23	€124,500
Higher education	0.04	0.06	0.54	0.02	
Log income	0.06	0.05	0.27	0.03	
Constant term	11.66	0.42	0.00*	-	

Note: * p<0.05. [a] values rounded to hundreds.

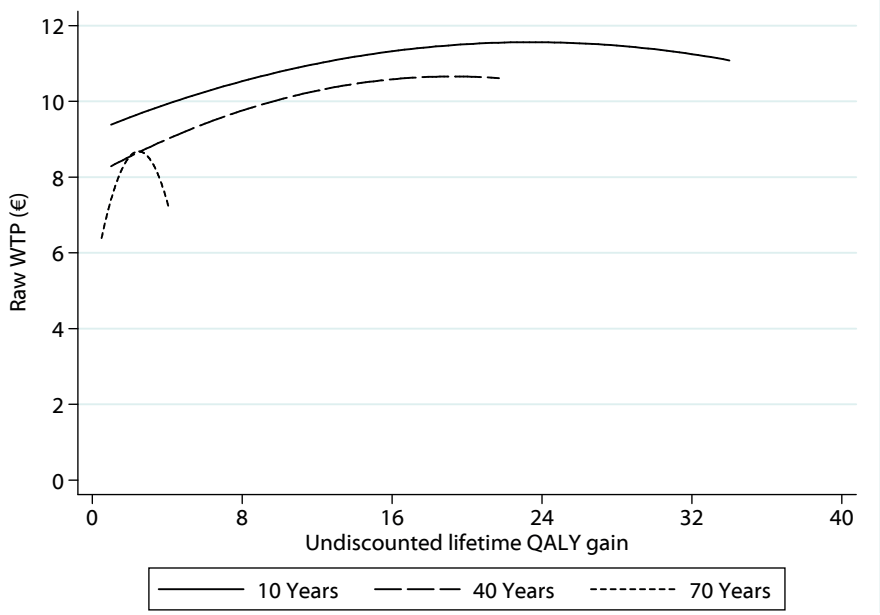


Figure A.5.3.1: OE(WTO) by dGQ for 10Y, 40Y and 70Y

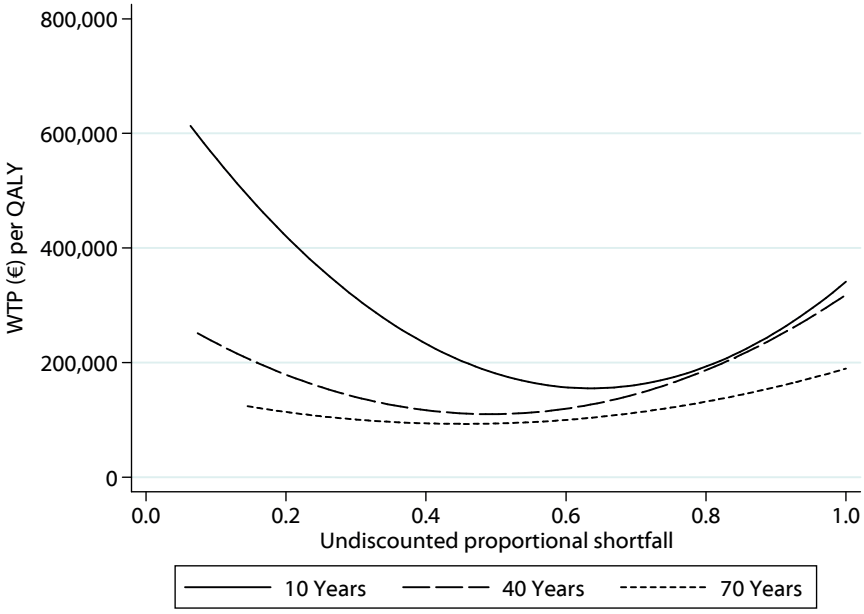


Figure A.5.3.2: WTP per QALY estimates at different levels of proportional shortfall and age

6

The risk of conditional reimbursement: stopping reimbursement can be more difficult than not starting in the first place!

Based on:
Van de Wetering, L., van Exel, J. & Brouwer, W. B. (2015). The risk of conditional reimbursement: Stopping reimbursement can be more difficult than not starting in the first place! *Submitted*

ABSTRACT

Conditional reimbursement of new health technologies is increasingly considered as a useful policy instrument. It allows gathering more robust evidence regarding effectiveness and cost-effectiveness without delaying market access. However, literature suggests that ending reimbursement and provision of a technology when it proves not to be (cost-) effective in practice, may be difficult.

The aim of this study was to investigate how policy makers and the general public in the Netherlands value removing a previously reimbursed treatment from the basic benefits package relative to not including a new treatment. To investigate this issue, we used discrete choice experiments. MMNL models were used to analyse the data. Compensating variations values and changes in probability of acceptance were calculated for withdrawal of reimbursement.

The results show that, ceteris paribus, both the general public (n=1,169) and policy makers (n=90) prefer a treatment that is currently reimbursed over one that is currently not yet reimbursed.

Apparently, ending reimbursement is more difficult than not starting reimbursement in the first place, both for policy makers and the public. Loss aversion is one of the possible explanations for this result. Policy makers in health care need to be aware of this effect when engaging in conditional reimbursement schemes.

6.1. INTRODUCTION

The growing pressure on health care budgets increases the need for a more explicit decision making process regarding the allocation of scarce resources. In this context, health care policy makers seek scientific evidence to inform these complex allocation decisions. Economic evaluation of new health technologies have become an important source of evidence for policy makers to guide these allocation decisions [145, 25]. However, the information on (cost-) effectiveness of new technologies is often imperfect at the time that the reimbursement decision needs to be made. For instance, first evidence about cost-effectiveness is typically generated in controlled trials and primarily aimed at meeting the regulators' requirements about efficacy and safety of the new technology, but it is usually inconclusive about effectiveness in real life use and the cost-effectiveness compared to existing technologies. As a result, many allocation decisions have to be made under considerable uncertainty, which makes it difficult for health care policy makers to make a well-considered long term reimbursement decision [146, 147, 24].

This leads to the question how to deal with this uncertainty in the decision making process. One option could be to postpone reimbursement in order to generate more evidence, but this may delay patients' access to new promising technologies [146]. Therefore, conditional reimbursement has been proposed as an alternative option to deal with uncertainty regarding health care allocation decisions, without delaying market access [24, 25, 147]. With conditional reimbursement, a new technology is included in the basic benefits package (in a health insurance system) or paid by government (in a National Health Services system) for a given period of time, under specified conditions. One of the common conditions is the collection of real-world data on costs and effectiveness of the technology. Hence, the policy instrument has also been labelled as coverage with evidence development (CED), funding with evidence development (FED) or access with evidence development (AED). Based on the additional evidence a better informed long term reimbursement decision can be made at the end of the conditional reimbursement period. In other words, using this policy instrument, policy makers can make new promising technologies available to patients at an early stage, while the long term reimbursement decision can be postponed until more robust evidence of the performance of the technology in daily practice has become available [146, 147].

Some countries have already implemented some form of conditional reimbursement as a policy instrument for health care allocation decisions, as for example US, Canada, UK, Australia, France, Sweden and Belgium [24, 148, 149, 150, 149, 150]. In the Netherlands, conditional reimbursement was implemented in 2006 to ensure early access to new expensive inpatient drugs, with a budget impact of at least €2.5 million. In 2013, this policy was extended to a selected groups of outpatient drugs that met the criteria for

temporary reimbursement. If a drug is conditionally reimbursed, hospitals receive an additional ear-marked budget to cover the expenses. This is combined with the obligation to gather data on appropriate use and cost-effectiveness in real-world clinical practice. After four years an evaluation is carried out to inform the final reimbursement decision [148, 151].

First experiences of countries using conditional reimbursement in practice showed that the re-assessment process appears to be a complex and politically sensitive procedure [152]. Therefore, researches and policy makers primarily aim to determine the conditions under which conditional reimbursement can be considered a feasible or optimal strategy and the type of evidence that needs to be gathered during the period of conditional coverage [153, 154]. Nevertheless, withdrawing funding for a technology that has been implemented and reimbursed –even if this was labelled as 'conditional'– appears to be far from straightforward.

A traditional reimbursement decision is typically related to allowing a new technology to either enter or not enter the health (insurance) system. Given that the technology was not yet funded this implies the status quo (not entering) or a gain (entering). Under conditional reimbursement, the second ('final') decision is either to continue funding (status quo) or to end –temporary- funding, which may be considered a loss. This difference is by no means trivial. Once a technology like a pharmaceutical is used in practice, ending reimbursement may be less feasible than deciding not to reimburse in the first place, in particular when the technology was proven to be effective in practice, but not cost-effective. This relates to the general tendency to value equally sized gains and losses differently, with losses looming larger than gains. This phenomenon of loss aversion is a well-known aspect of prospect theory [26, 155].

In the context of allocation decisions loss aversion may imply that policy makers may be willing to accept higher cost-per-QALY ratios for technologies already reimbursed (under conditional reimbursement) than they would accept for technologies not yet reimbursed (in the conventional decision making context). So far, this asymmetry in removing something from the package versus allowing something in the package has remained largely unexplored. The aim of this study was to investigate how policy makers and the general public in the Netherlands value removing an existing treatment from the basic benefits package relative to not including a new treatment in the first place, in the context of health care allocation decisions. That is, we investigate the value of removing an existing treatment from the basic benefits package relative to not including a new treatment in the first place. In other words, is stopping indeed more difficult than not starting? A discrete choice experiment (DCE) was designed to investigate preferences for different technologies, with a set of relevant criteria for health care allocation

decisions obtained from the literature and information on the current reimbursement status of the treatment as choice attributes. Data was collected both from health care policy makers and the general public.

6.2. METHODS

6.2.1. Discrete choice experiments

DCEs have proven to be a useful method to elicit individuals' preferences in health care decision making [70, 97]. DCEs are based on random utility theory, which assumes that a respondent, confronted with a choice between different scenarios, always chooses the alternative that gives the highest utility. The utility of alternative *j* in a choice situation for respondent *n* is given by:

Eq 6.1 $U_{nsj} = V_{nsj} + \varepsilon_{nsj}$

U_{nsj} can be separated into two components, V_{nsj}, the observed component of utility, and ε_{nsj}, the residual unobserved component. The observed component of utility is assumed to be a linear relationship of attribute levels, *x*, of each alternative *j* and their corresponding parameter weights, β, such that:

Eq 6.2 $U_{nsj} = \lambda_j \sum_{k=1}^{k} + \beta_k x_{njk} + \varepsilon_{nsj}$

where β_{nk} denotes the marginal utility associated with attribute *k* for respondent *n*. In the basic multinomial logit model the unobserved component, ε_{nsj}, is assumed to be independently and identically (IID) extreme value type 1 (EV1) distributed [73, 97].

6.2.2. Identification and presentation of attributes and levels

The focus of this study was to explore how policy makers and the general public value removing an existing treatment from the basic benefits package relative to not including a new treatment in the first place. Therefore, the main attribute of this study was the current reimbursement status of a treatment. To emphasize the fact that a certain treatment was not only reimbursed but also used by patients, the levels of the attribute were defined as 'existing treatment, currently reimbursed and used in practice', and 'new treatment, currently not reimbursed and not used in practice'.

Besides this main attribute of the study, additional criteria potentially relevant in health care allocations decisions were identified from the literature. Recent related studies by Koopmanschap et al. [156] and Van de Wetering et al. [97] were used as primary

sources of information for potentially relevant choice attributes in the Dutch policy context. Based on the literature, the following additional attributes were selected: age of patients, quality of life before treatment, health gain from treatment, cost per QALY, budget impact and probability that the cost per QALY would double. The attributes and corresponding levels were identical for policy makers and the general public. An overview of the attributes and the corresponding levels can be found in Table 6.1.

Since the general public is less familiar with the terminology and the common interpretation of absolute levels of cost effectiveness and budget impact in the policy context, we gave them an indication whether a certain level could be considered favourable or not (Table 6.1).

Table 6.1 Attributes and levels

Attributes	Levels*
Age of patients (AGE) *in years*	15, 40, 65, 90
Quality of life before treatment (QOL) *scale 0-100*	5, 30, 55, 80
Health gain from treatment for one year (GAIN) *scale 0-100*	5, 10, 15, 20
Reimbursement status of treatment (STATUS)	Existing treatment, currently reimbursed and used in practice New treatment, currently not reimbursed and not used in practice
Cost per QALY (ICER)	€ 10.000 (very favourable) € 50.000 (favourable) € 90.000 (unfavourable) € 130.000 (very unfavourable)
Budget impact (BUDGET)	€ 5 million (very low) € 30 million (low) € 55 million (high) € 80 million (very high)
The probability that the cost per QALY will double (UNCERTAINTY)	10% 15% 20% 25%

Note: * Labels between brackets were only shown to the general public.

The attributes, levels and presentation of choice sets were pilot-tested in a sample of 156 respondents of the general public. To be able to evaluate the attributes and construction of the design, a number of questions concerning the complexity, plausibility and comprehensibility of the choice options were added to the pilot study (see Appendix 6.1). The pilot study revealed that 39.7% of the respondents thought it was difficult to opt for one of the groups in the choice sets. However, from respondents' explanations to this question it appeared that this was mainly due to the fact that

people preferred not to choose between groups of patients at all. The results showed that 65.4% of the respondents took all attributes in consideration while making their decision. The attribute which was most often considered important was health gain as a result of treatment, followed by quality of life before treatment and age. Only 44% of the respondents considered the probability that the cost per QALY would double to be an important argument for their choice between groups of patients. Finally, people were asked whether they had sufficient information to make a well-considered decision; 17.9% of the respondents answered that they needed additional information, predominantly on the personal circumstances of the patients, the success rate of treatment and the life expectancy after treatment. Given that more than 80% of respondents indicated that the scenarios provided sufficient information and the large variety in proposed additional attributes for the experiment, with limited support for each, the design of the study was left unchanged.

6.2.3. Questionnaire

The questionnaire started with a short introduction and a detailed description of all attributes included in the study. Respondents were asked to imagine themselves being a policy maker facing allocation decisions in health care. They were then asked to imagine that there were two comparable treatments in consideration for reimbursement. The two treatments only differed in terms of the attributes in the choice sets. Due to a limited budget, only one of the treatments could be included in the basic benefits package. The respondents were asked which of the two treatments they, as policy makers in the health care sector, would prefer to include in the basic benefits package. They could also opt for neither (see Figure 6.1 for an example of a choice scenario).

The program Ngene 1.1 was used to generate an efficient multinomial logit design for the main study. An efficient design attempts to optimize the information obtained from each choice set by minimizing the predicted standard errors of the parameters. The D-error, which is the most widely used measure of efficiency, was used to determine the efficiency of the design [74]. The estimates of the pilot study were used as Bayesian priors for the main study. Bayesian priors optimize on prior distributions instead of on fixed parameters which makes them more robust to misspecification [74]. Furthermore, we used 1,000 Halton sequence draws and included all interaction effects between attributes. Halton draws are more effective than random draws because they are based on a set of values which are well spaced over the distribution [73].

A design with 24 choice sets was generated. For the general public the choice sets were divided over 3 versions with 8 choice tasks using a blocking variable. The questionnaire for policy makers consisted of 2 versions with 12 choice tasks each. Respondents were randomly assigned to one of the versions of the questionnaire. The alternatives were

You can only include one treatment in the basic benefits package. Which of the two treatments below would you prefer?

Treatment A	Treatment B	Neither
The patients are 70 years old, have a quality of life of 55 and gain for one year a quality of life of 15	The patients are 40 years old, have a quality of life of 80 and gain for one year a quality of life of 10	
It involves an existing treatment currently reimbursed and used in practice, the cost per QALY are €50.000 the budget impact is €55 million and there is a 10% probability that the cost per QALY will double	It involves an new treatment currently not reimbursed and not used in practice, the cost per QALY are €90.000 the budget impact is €5 million and there is a 20% probability that the cost per QALY will double	

I would prefer…

Treatment A	Treatment B	Neither
❑	❑	❑

Figure 6.1: Example of choice scenario

unlabelled and the choice sets were randomized within blocks to avoid order biases in the results. The questionnaire for the general public also included two control questions to detect inconsistent respondents: one dominant choice set was presented as first choice set in all versions, also to introduce and familiarise respondents with the type of questions. In addition, the fourth choice set was repeated as the last choice set, but with left and right placed scenarios reversed. If a respondent chose the dominated scenario in the first choice and reversed preferences in the last choice, the respondent was removed from the dataset. Furthermore, based on the pilot study we set a minimum meaningful completion time for the ten choice sets in the general public sample of 150 seconds. Finally, policy makers were presented three follow-up questions to verify and better understand their responses to the DCE (see Appendix 6.1), which were also included in the pilot study.

6.2.4. Data collection

The questionnaire was distributed by a professional Internet survey company to a representative sample of the Dutch adult population in terms of gender, age and level of education. In addition, a convenience sample of (future) policy makers in health care in the Netherlands was invited to participate in this study to get insight into their prefer-

ences. Policy makers were invited personally by email, using the professional networks of the researchers conducting this study. Participants included policy makers from the National Health Care Institute and other policy makers and (academic) HTA experts involved directly or indirectly in the process of health care allocation decisions in the Netherlands. In addition, a sample of PhD and master students with a background in health economics and health policy were invited to participate, as potential future policy makers or advisors. The student sample received an incentive of for participation (i.e., a €10 gift card).

Prior to the start of the questionnaire, respondents were informed about the purpose of the study, told that participation was voluntary and anonymous, and requested to give informed consent.

6.2.5. Analyses

The deterministic components of the elemental alternatives for each age group were represented by:

Eq 6.3 $V_A/\lambda = \beta_1 AGE + \beta_2 QOL + \beta_3 GAIN + \beta_4 STATUS + \beta_5 ICER + \beta_6 BUDGET + \beta_7 UNCERTAINTY$

$V_B/\lambda = \beta_1 AGE + \beta_2 QOL + \beta_3 GAIN + \beta_4 STATUS + \beta_5 ICER + \beta_6 BUDGET + \beta_7 UNCERTAINTY$

$V_C/\lambda = \beta_0$

where V is the deterministic component of the random utility function, λ is the scale parameter and β_i are the parameters to be estimated. The coding of the corresponding variable is reported in Table 6.1. The constant terms represent the expected utility for no treatment over treatment.

To analyse the data we estimated MNL models as well as panel Mixed MNL models (MMNL), which allow for preference heterogeneity among the population by using random parameters. Likelihood ratio tests were used to test different specifications of the utility functions (e.g. categorical or numerical attribute levels and interaction effects) on the dataset of the general public. Dummy variables of quality of life before treatment and age were included in the model since they significantly improved the model fit. All variables were included as random variables with a normal distribution except for the ICER variable. The ICER variable was held fixed to avoid implausible welfare values [101, 157]. The MMNL models were estimated with 1,000 Halton draws.

To assist interpretation of the model results we defined three scenarios and calculated for each scenario the compensation variation value (CV) and the difference in the prob-

ability of acceptance between new and existing treatments. We defined an 'average scenario' where all attributes were set at the second level of the four level attributes, a worst case scenario, where all attributes were set at the least preferable level, and a best case scenario, where all attributes were set at the most preferable level, all from the perspective of the general public. The CV value reflects the welfare effect of a change in the attributes while taking into account the uncertainty regarding which alternative will be chosen in a DCE. The CV was calculated according to the Small and Rosen (1981) formula:

Eq 6.4 $$CV = \frac{1}{\lambda} \left[ln \sum_{j=1}^{j} e^{v_j^0} - ln \sum_{j=1}^{j} e^{v_j^1} \right]$$

where λ is the marginal utility of income, V^0 and V^1 are the utilities of each alternative before and after the change, respectively, and *j* is the number of alternatives in each choice set [158]. Instead of income, the marginal utility of the ICER attribute was used in this study to express the effect of conditional reimbursement. We took 10,000 Halton draws from the estimated distributions of the random parameters to simulate the CV using the above formula. The reported CV values are the means over these 10,000 replications [159, 160]. The probabilities of acceptances were simulated over the data with the simulation feature in Nlogit 5.0 [161, 162].

The calculations of CV and probabilities of acceptance eliminate the scale between different datasets which makes is possible to directly compare the results between the general public and policy makers.

6.3. RESULTS

Demographic characteristics of the general public and policy makers are provided in Table 6.2. The sample of the general public (n=1,169) was representative for the Dutch population in terms of age, gender and level of education. A total of 90 of the 156 invited (future) policy makers completed the questionnaire (57.7% completion rate). Among these respondents, 22.2% identified as policy makers, employed at the National Health Care Institute of the Netherlands and 36.6% identified as policy makers or HTA experts indirectly involved in process of health care allocation decisions. 41.1 % of the sample consisted of PhD and master students in health economics and health policy.

The results of the MNL and the MMNL models for the general public and the policy makers are presented in Table 6.3. The MMNL models performed significantly better than the corresponding MNL models. The standard deviations of almost all parameters were

Table 6.2: Demographic variables

Variable	Category	General public (n=1,169)	Policy makers (n=90)
Gender	Female	49.0%	55.6%
	Male	51.0%	42.2%
	Prefer not to say	-	2.2%
Age	Mean (SD; min-max)	48.8 (15.5; 18-75)	37.0 (11.5; 22-62)
Education status	Low	31.6%	-
	Middle	43.5%	-
	High	24.9%	100%

statistically significant, which reflects heterogeneity in preferences and thus supports the use of panel MMNL models.

6.3.1. General public

In the general public sample, all random parameters were statistically significant. This indicates that all attributes included in the design had a significant effect on respondents' choices. The coefficients of the attributes represent their relative importance for respondents, while the statistically significant positive constant term shows that respondents preferred choosing one of two treatments over neither of both. All attributes in the model performed as expected with respect to the signs of the coefficients.

The results show that the general public preferred reimbursement of an existing treatment that is currently reimbursed and used in practice over reimbursement of a new treatment, currently not reimbursed or used in practice. Furthermore, respondents preferred treatments for relatively younger patients over treatments for older patients and treatments with relatively larger health gains. A higher cost per QALY, a higher budget impact or a higher probability that the cost per QALY will double, all significantly reduced the probability of selection of a treatment for reimbursement. Furthermore, respondents preferred treatments targeting patients with relatively lower initial health states except for a health state of 5 units of quality of life before treatment. The negative coefficient for the initial health state of 5 units of quality of life indicates that respondents have a negative preference to treat this particular groups of patients compared to patients with an initial quality of life of 80 units. All standard deviations were significant, which shows considerable preference heterogeneity in the sample over all attributes.

6.3.2. Policy makers

The results in the policy makers sample were very similar to those in the general public, with the lowest level of the quality of life before treatment attribute as only exception. All random parameters were statistically significant, indicating that all attributes included

Table 6.3: Results MNL and Mixed MNL

Variable	General Public		Policy makers	
	MNL	MMNL	MNL	MMNL
Age of patients[a] 40	-0.357***	-0.611***	-0.478***	-1.122***
65	-0.455***	-0.903***	-0.915***	-1.829***
90	-1.661***	-3.108***	-2.488***	-6.685***
Quality of life before treatment[b] 55	0.190***	0.369***	1.115***	2.446***
30	0.122***	0.193***	1.548***	3.561***
5	-0.186***	-0.392***	1.472***	3.567***
Health gain from treatment	0.025***	0.051***	0.075***	0.179***
New treatment	-0.085***	-0.108**	-0.340***	-0.489*
Cost per QALY (euro/QALY)	-0.004***	-0.007***	-0.016***	-0.034***
Budget impact (euro)	-0.005***	-0.010***	-0.013***	-0.035***
Probability that the cost per QALY will double (%)	-0.008***	-0.008*	-0.006	-0.045*
Constant	2.464***	4.850***	2.511***	5.368***
SD random parameters				
Age of patients[a] 40		1.132***		1.119***
65		1.717***		1.514***
90		2.948***		3.678***
Quality of life before treatment[b] 55		0.881***		1.519***
30		1.408***		2.711***
5		1.826***		4.226***
Health gain from treatment		0.054***		0.160***
Include new treatment		0.536***		1.637***
Budget impact (euro)		0.011***		0.026***
Probability that the cost per QALY will double (%)		0.031**		0.163***
Constant		3.557***		4.621***
Log-likelihood at zero		-10274.222		-1186.501
Log-likelihood at convergence	-8166.054	-7100.122	-780.272	-677.164
McFadden's Pseudo R2		0.309		0.429

Note: [a] reference level: 15. [b] reference level: 80. ***, **, *: significance at 1%, 5%, 10% level; MNL: multinomial logit model. MMNL: mixed multinomial logit model. SD: standard deviations.

in the design significantly influenced the reimbursement decisions of policy makers. Also for policy makers, removing an existing treatment from the basics benefits package significantly reduced utility as compared to not including a new treatment in the first place. There was significant preference heterogeneity in the sample with respect to all attributes included in the model.

The results of the questions presented to policy makers after the DCE (see appendix 6.1) are shown in Table 6.4. Only 38.9% of the policy makers considered all attributes in their

Table 6.4: Statements policy makers (n=90)

Variable	(very) important	neutral	(very) unimportant
Age of patients	86.7	5.5	7.8
Quality of life	92.3	3.3	4.5
Health gain	92.2	4.4	3.3
Reimbursement status of treatment	18.9	21.1	60.0
Cost per QALY	76.7	8.9	14.4
Budget impact	52.2	28.9	18.9
Probability that the cost per QALY will double	17.8	35.5	46.7

decision. Policy makers rated quality of life before treatment and health gain as a result of treatment as the most important characteristics for their decision to include a treatment in de basic benefits package. Remarkably, the current status of treatment in the basic benefits package was most frequently valued as unimportant; 60% considered this attribute unimportant and only 18.9% thought it was important for their decisions. This seems to contradict the results of the DCE, since it showed a significant coefficient in the MMNL model. The probability that the cost will double was insignificant in the MNL model, while the answers on the statements show that about half of the respondents considered this important.

Half of the policy makers indicated that they needed more information than the attributes included in the design to be able to make a well-considered decision about reimbursement of treatments. The additional information they desired included: more detailed information about the disease (e.g. symptoms and life expectancy) and the treatment (e.g. prevalence), duration and side effects of the treatment, and the available scientific evidence.

6.3.3. Compensating variation and probability of acceptance

Table 6.5 shows the results of the CV and the differences in probability of acceptance between new and existing treatments, both for the general public and policy makers. The CV values reflect the acceptable increase in ICER for an existing treatment compared to a new treatment. This means that for the average scenario with an ICER of €50.000 the general public would accept an €7.360 higher ICER for an existing treatment compared to a new treatment. For both the general public and policy makers, CV is the lowest for the least preferred scenarios which seems reasonable since these scenarios will not be chosen by most respondents. The CV of the most preferred scenarios are the highest. Policy makers are willing to accept higher ICER ratios for existing treatments compared to new treatments than the general public, except for the least preferred scenario.

Table 6.5: Changes in predicted probability of acceptance and the compensation variation value if the intervention already exist (compared to new intervention)

	Average scenario	**Least preferred scenario (GP)**	**Most preferred scenario (GP)**
Attributes			
Age of patients	*40*	90	15
Quality of life before treatment	*30*	5	55
Health gain from treatment	*10*	5	20
Include new treatment	*New -> existing*	*New -> existing*	*New -> existing*
ICER	*€ 50.000*	€ 130.000	€ 10.000
Budget impact	*€ 30 million*	€ 80 million	€ 5 million
Probability that the cost per QALY will double	*15%*	25%	10%
Differences between an existing treatment compared to a new treatment within each scenario			
Change in predicted probability (general public)	*+ 1.1%*	+0.86%	+ 1.3%
Change in predicted probability (policy makers)	*+ 2.8 %*	+ 1.7%	+ 4.2 %
CV general public ICER (euro)	*€ 7.360*	€ 4.146	*€7.414*
CV policy makers	*€ 7.959*	€ 60	€ 7.971

As shown by the positive values of the change in probability of acceptance between new and existing treatments, existing treatments have a higher chance of acceptance compared to new treatments. For the average scenario, an existing treatment had a 1.1% higher probability of acceptance compared to a new treatment in the sample of the general public, a 2.8% higher probability in the sample of policy makers. The magnitude of change in probability of acceptance was not large. Both for the general public and policy makers the difference was lowest in the least preferred scenario and highest in the most preferred scenario.

Overall, policy makers appear to be more sensitive to the current status of treatment than the general public.

6.4. DISCUSSION

Conditional reimbursement is increasingly recognized as a useful policy instrument for health care allocation decisions. However, conditional reimbursement changes the decision making process, since the decision to end reimbursement of a technology after a period of time, in which additional evidence was gathered on effectiveness and cost-effectiveness in daily practice, differs from the decision not to reimburse a technology in the first place. The main objective of this study was to explore how the Dutch general public and policy makers in health care value existing treatments relative to

new treatments. The results of the discrete choice experiment showed that both the general public and policy makers preferred to reimburse an existing technology, which is already reimbursed and used in practice, over reimbursing a new technology, which is currently not reimbursed or used. This finding may have important implications for the consideration to use conditional reimbursement as policy instrument for allocation decisions.

The result of the MMNL models showed the relative importance of the different attributes in the choice sets. The magnitude of the coefficients are not directly comparable between the samples of the general public and the policy makers due to differences in scale. However, the signs and the significance of the attributes were very comparable between the two data sets, with the exception of patients in a very poor health state before treatment. The results of the general public showed a negative preference to treat patients in a very poor health state. Although this may seem counterintuitive, previous studies showed similar results and hypothesized that people may consider the health state after treatment rather than before treatment in their decisions which treatment to prioritize. If patients remain in a poor health state after treatment, this reduces the probability for a treatment to be reimbursed [14, 70, 163]. Since the maximum health gain in this study was 20 QoL units, the reluctance to allocate resources to patients with very low health state before treatment (i.e., QoL 5), who would remain in a poor health state after treatment (i.e., in the best case improve to QoL 25), may be a reasonable explanation for this finding. In the policy makers sample the sign of the coefficient of quality of life before treatment was in the anticipated direction, indicating that policy makers give lower priority to treatments for relatively healthy patients (i.e. QoL 80).

The compensating variation values and the changes of probability of acceptance provided the opportunity to directly compare the effect of conditional reimbursement between the general public and policy makers. Both preferred to reimburse an existing treatment over a new treatment, ceteris paribus. Although differences were small, policy makers appeared to be more sensitive to the current status of the treatment than the general public. For instance, in the average scenario, the general public was willing to accept a €7.360 higher ICER for an existing treatment as compared to a new treatment. While it might be expected that policy makers are not sensitive to the process of the allocation decision, they were willing to accept a €7.959 higher ICER for an existing treatment as compared to a new treatment. Apparently, policy makers are more aware of the political consequences of ending reimbursement than expected.

After completing the DCE, policy makers were presented with a number of statements related to the attributes of the choice experiment. The responses to these statements were not completely consistent with the choices made in the DCE. For example, the

attribute 'status of treatment' was significant in the DCE but the response to the corresponding statement showed that only 18% of the policy makers thought this characteristic was important for their decisions, and 60% even thought it was unimportant. These differences may of course stem from the different methods used to elicit preferences. The DCE elicits *relative* preferences of the different attributes while the statements focus on one attribute at the time. Previous studies have already highlighted that different methods and different study contexts tend to produce different results [163]. Therefore, our results should be interpreted with caution and more research is required to confirm these findings.

Apart from the general limitations of online surveys and discrete choice experiments, some limitations deserve explicit mentioning here. The significant standard deviations of the coefficients indicated that there was considerable heterogeneity in preferences in both samples. This should be kept in mind when interpreting the results. Some standard deviations were larger than the relevant coefficient which implies that preferences switch sign within the sample. For example, in both samples the status of the treatment attribute had a negative coefficient, indicating a negative preference for removing a treatment from the basic benefits package. However, the size of the standard deviation showed that a considerable proportion of the respondents had a positive preference for this attribute, indicating a preference for new treatments over existing ones. In addition, although the sample of the general public was representative for the Dutch population and therefore has some external validity, the policy makers concerned a convenience sample and therefore the external validity of these results is questionable. Obtaining a sizeable and representative sample of policy makers proved very challenging, as many different people in different roles and organizations are involved in the process; it is therefore difficult to define clear sampling criteria, let alone get access to this sample. We have attempted to collect a relevant dataset for the purposes of this study, by purposively recruiting a considerable number of representatives from important stakeholder settings in the decision making process in the Netherlands, complemented with PhD and master students in training as health economists and policy makers. Although this sample was small and possibly selective, we think that the sample provided interesting information for the aim of this study. Finally, the pilot study showed that about 40% of respondents found it was difficult to choose between groups of patients in the scenarios. This was not because the choice options were not clear or too difficult, but because they preferred not to choose at all. Some recent studies into societal preferences for resource allocation [164, 165] observed similar preferences, related to an egalitarian view on resource allocation –giving everyone equal access to care - and an apparent denial of the scarcity of health care budgets – giving everyone access to all the care they need.

To circumvent that respondents with such a preference were forced choosing between groups of patients an opt-out option was offered.

Overall, this study showed that the impact of excluding a treatment from the basic benefits package on the utility both the general public and policy makers in the Netherlands is higher than including that same treatment. In other words, stopping reimbursement seems more difficult than not starting reimbursement in the first place. We may conclude from these results that conditional reimbursement evokes loss aversion, suggesting that higher cost-effectiveness ratios are considered acceptable for existing treatments than for new treatments. This has important implications for allocation decisions in health care, in particular because this suggests that the *decision making process* directly influences the outcome of a reimbursement decision. Policy makers should be aware of this potential risk when they consider conditional reimbursement as a policy instrument.

More research is required to further investigate the possible consequences of conditional reimbursement for the allocation of scarce resources in health care. It would be interesting to replicate this study in other countries considering or already using conditional reimbursement as a policy instrument for health care allocation decisions.

APPENDIX 6.1: ADDITIONAL QUESTIONS FOR PILOT STUDY AND POLICY MAKERS

1. Have you considered all attributes in your decision making process?

❑ No

❑ Yes

2. How important were the following characteristics for your decision which treatment to prioritize?

	Very unimportant			Neutral		Very important	
Reimbursement status of treatment	☐	☐	☐	☐	☐	☐	☐
Age of the patients	☐	☐	☐	☐	☐	☐	☐
Quality of life before treatment	☐	☐	☐	☐	☐	☐	☐
Health gain from treatment	☐	☐	☐	☐	☐	☐	☐
Cost per QALY	☐	☐	☐	☐	☐	☐	☐
Budget impact	☐	☐	☐	☐	☐	☐	☐
Probability that the cost per QALY will double	☐	☐	☐	☐	☐	☐	☐

3. Have you considered any other characteristics in your decision which treatment to include in the basic benefits package?

❑ No, as a policy maker, I had enough information to make a well-considered choice between the two treatments.

❑ Yes, namely….

Part B

7

Health or Happiness? A note on trading off health and happiness in rationing decisions

Based on:
Van de Wetering, E., van Exel, N., & Brouwer, W. (2015). Health or Happiness? A note on trading off health and happiness in rationing decisions. Value in Health. *In press.*

ABSTRACT

Economic evaluations typically value the effects of an intervention in terms of Quality-Adjusted Life-Years, which combine length and health-related quality of life. It has been suggested that economic evaluations should incorporate broader outcomes than health-related quality of life. Broader well-being, for instance measured as happiness, could be a better measure of the overall welfare effects in patients due to treatment.

An underexplored question is whether and how people trade-off information on health and broader outcomes from treatment in rationing decisions. This paper presents the results of a first experiment aimed to explore such trade-offs between health and happiness.

We used a web-based questionnaire in a representative sample of the public from the Netherlands (n=1,015). People made choices between two groups of patients differing in terms of their health and happiness levels before treatment and gains from treatment.

The results show that about half of the respondents were willing to discriminate between patient groups based on their health and happiness levels before and after treatment. In this trader group, health gains were considered somewhat more important than happiness gains. However, our findings suggest that *both* health and happiness levels of patients may play a role in priority-setting.

7.1. INTRODUCTION

Economic evaluations can support informed decision making regarding the optimal allocation of scarce resources. An important issue is to decide *what* it is that should be optimized through these decisions. In health care the answer to that question has often been 'health', measured and valued in some acceptable way [5]. In the field of health care this focus on the value of health as outcome measure may be considered a logical one, as many interventions will aim to improve the health of beneficiaries. Health benefits of interventions can be measured using clinical outcome measures, but in the context of economic evaluations, Quality-Adjusted Life-Years (QALYs), capturing length and health-related quality of life, are often recommended as measure of outcome [1, 6, 28].

Nonetheless, there is increasing debate on whether other, broader outcome measures may be more appropriate for use in economic evaluations [10]. An important reason for this is that the narrow focus on the value of health (or health-related quality of life in QALYs) may not always align well with the goals of interventions in the health care sector. For instance, in long-term care, the aim of interventions may not be to improve 'health' (alone), but to contribute to overall wellbeing, including elements such as autonomy and security. In such cases, outcome measures capable of capturing wellbeing may be better suited to capture the full benefits of interventions [10]. Alternative measures that aim to capture wellbeing outcomes have been proposed, such as the multi-attribute instruments ICEpop CAPability measure for Older people (ICECAP-O) [166, 167], and Adult Social Care Outcomes Toolkit (ASCOT) [168], but also subjective wellbeing measured as happiness [10, 28, 169]. In the rest of this paper, we will use the terms (subjective) wellbeing and happiness interchangeably, as happiness could be seen as a representation of overall wellbeing [32], of which health is one (important) element.

Subjective wellbeing has also been advocated as an alternative outcome measure because it would better reflect *experienced utility* from health changes, as appreciated in real life, while QALY calculations are commonly based on *expected utility*, as anticipated when making decisions [170, 171]. Due to mechanisms such as adaptation and coping, experienced utility may be higher than expected utility [12, 172]. One may therefore claim that using QALYs in allocation decisions protects patients against the consequences of adaptation, since adaptation reduces the utility gain from treatment. However, at the same time, QALY scores do not represent the full wellbeing gains of patients, and therefore may discriminate against those diseases in which adaptation is more difficult (e.g. mental illness) [173, 174].

This debate highlights important differences of opinion about the content of what should be optimized (health or wellbeing) and how these concepts should be opera-

tionalized and valued. The discussions on appropriate outcome measures in (economic evaluations of) health care will likely continue, including whether such alternative outcome measures should replace or complement health (or health-related quality of life) outcomes. While more theoretical debates about this issue remain crucial, empirical studies can highlight how trade-offs between different outcome measures are made.

An important question for instance is how allocation decisions in health care would be affected when information is provided on changes in both health and subjective wellbeing of patients. This could indicate whether people focus on either of the two outcomes in allocating scarce health care resources, or trade them off. In the latter case, their relative importance in allocation decisions obviously becomes interesting. This note presents a first, explorative study to investigate this. We asked members of the public from the Netherlands to choose between treatment of two groups of patients, differing in terms of health and happiness levels before treatment and gains from treatment.

In the experiment, we focus on the question whether respondents in prioritising patients in health care focus on health or happiness levels and gains, or use both types of information. We chose happiness as outcome measure next to health, since it is intuitively understandable for respondents. In the experiment, we focus purely on the issue of outcome measure (health or happiness), and, by using health and happiness ratings, not on experienced or expected utility. Our results therefore provide first insight in the intriguing and largely unexplored question of whether and how people trade-off information on health and happiness gains from treatments in the context of health care.

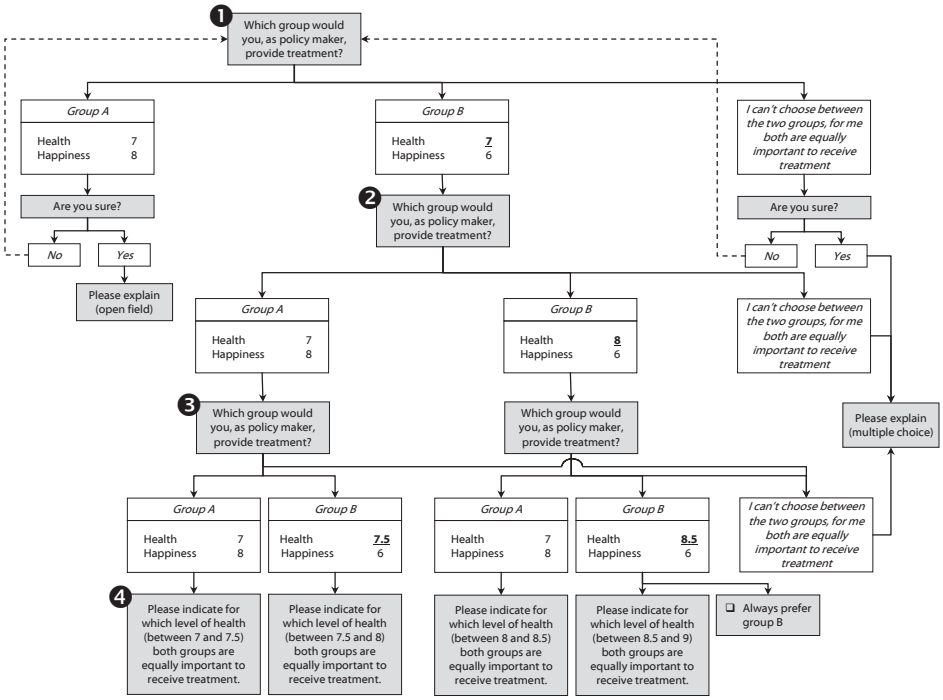
7.2. STUDY DESIGN AND HYPOTHESES

A web-based questionnaire was designed in which respondents were asked to imagine being a healthcare decision maker facing priority-setting decisions between two groups of patients that only differed in the amount of health and happiness they had and could gain from treatment.

First, respondents were asked to rate their own health and happiness on a VAS scale, to familiarize them with the outcome measures and measurement scale used in the experiment. Then, they were asked to make treatment choices between groups of patients, using three distinct scenarios. In all three scenarios, both patient groups were described as being healthy and happy, with levels of 9 on scales from 'worst conceivable health' [0] to 'best conceivable health' [10] and 'completely unhappy' [0] to 'completely happy' [10] before the onset of illness. Next, respondents were informed that each group would be

affected by a different (unspecified) illness, with diverging effects on their health and happiness levels. It was specified that happiness could be affected differently due to health changes, for instance because of the impairing effects of illness on daily activities and the possibility of adaptation to illness. Finally, respondents were informed that the group receiving treatment would return to their original levels of health and happiness (i.e., 9 on a scale from 0 to 10), whereas the group not receiving treatment would remain at the lower levels of health and happiness indefinitely. Groups were said to be of the same size, and costs of treatment identical (see Appendix 7.1 for an example question). The three scenarios only differed in terms of the size of the effect of illness on health and happiness.

In scenario 1 respondents were asked to choose between two groups of patients experiencing the same effect of illness on their health, but different effects on their happiness, with Group A suffering less in terms of happiness than Group B. Hence, if respondents



Note: Example concerns scenario 1, version 1: Group A and happiness Group B were fixed, health Group B varied to look for indifference. Multiple choice options in case indifferent between Groups A and B: (1) Everyone is equally important to treat, irrespective of differences in happiness; (2) The effect of treatment on happiness should not be taken into account; (3) The difference in happiness is not large enough to prioritize the one group over the other; (4) Other, namely … [open field].

Figure 7.1: Example of the iteration process

considered only health to be relevant, they would be indifferent between treating A or B. If they considered only happiness or both health and happiness to be relevant, they would prefer to treat Group B. In choice 1, Group A is a dominated choice and therefore considered as an inconsistent response; respondents selecting this group were asked to explain (see Figure 7.1) and were further disregarded for this scenario. In order to determine the point of indifference for treating either group, the status of Group A was held constant in following treatment choices and the health effect in Group B was adjusted up- or downwards based on the previous treatment choice of the respondent (see Figure 7.1 for an example of the iteration process). After a maximum of three choices (i.e., ❶ to ❸ in Figure 7.1) respondents were asked to indicate the health effect in Group B that, given the difference in happiness effect, would make them indifferent between providing treatment to group A or group B (i.e., ❹ in Figure 7.1). In case respondents chose Group A in choice 1 they were asked to explain their choice (open field). When respondents indicated to be indifferent between the groups at any stage in the scenario, they were asked to clarify their choice (multiple choice; see note to Figure 7.1).

Scenario 2 was very similar to scenario 1, except that now the two groups of patients experienced the same effect on their happiness levels but different effects on their health levels. Subsequently, the effect of illness on the happiness level of group B was varied to determine the respondent's indifference point. In scenario 3 the starting levels of both health and happiness were varied in order to offer respondents a wider range to trade-off outcomes. The results of Scenario 3 are not presented in any detail here. From the many inconsistent choices (and explanations of them in relation to follow-up statements), we concluded that this last scenario had been too complex (at least in the presented online format) for respondents and was therefore excluded from the analysis.

The questionnaire consisted of four versions with different starting levels for health and happiness (see Table 7.2 for scenario 1 & Table 7.3 for scenario 2). The units of change in health or happiness were constant over the versions (Figure 7.1) except for version 4. Given the low starting levels before treatment in version 4, the units of change in health/happiness were doubled in order to cover the entire evaluation space within three questions. Respondents were randomly assigned to one of the four versions and completed the scenarios in a fixed order. Next, respondents were presented with six statements related to the choice experiments (Table 7.1) which were included to verify and better understand the choices respondents made. Finally, some socio-demographic data were collected.

Two hypotheses were formulated for this experiment:

(A) People consider both the effects on health and happiness in health care decision making.

(B) People consider health to be more important than happiness for health care priority-setting.

For the analysis, respondents were divided into 'traders' and 'non-traders' subgroups. Respondents expressing indifference between the groups of patients throughout the experiment, apparently unwilling to trade-off health and happiness in priority setting, were assigned to the 'non-traders' subgroup. Differences between the subgroups in background characteristics and response to the six follow-up statements were explored. For respondents in the 'traders' subgroup, relative weights of health and happiness gains were calculated for each scenario by dividing the level differences at the indifference point. In case respondents indicated to 'always prefer group B' (see Figure 7.1, bottom right), the point of indifference for health or happiness was set to the original level before illness (i.e., 9), potentially leading to a ceiling effect in the results.

Table 7.1: Summary statistics (n=1,015)

Variable	Traders (n=495)		Non-traders (n=520)		P value
	%	Mean (SD)	%	Mean (SD)	
Gender (female)	43.8		55.0		0.000
Age		39.1 (13.1)		42.3 (12.4)	0.000
Level of education*					
- Low	11.7		14.2		
- Middle	50.9		53.5		
- High	37.4		32.3		0.005
Health Status		78.6 (17.7)		78.4 (17.5)	0.900
Happiness		71.1 (18.6)		74.0 (19.1)	0.014
Statements					
- All patients are equally entitled to health care, irrespective of the effect of the illness on their happiness†		5.5 (1.6)		6.2 (1.5)	0.000
- The ultimate goal of the health care system is to promote health†		5.6 (1.6)		5.8 (1.5)	0.004
- The ultimate goal of the health care system is to promote happiness†		4.3 (1.6)		4.2 (1.7)	0.133
- Happiness is more important than health†		3.6 (1.7)		3.3 (1.5)	0.016
- Health is more important than happiness†		4.7 (1.7)		4.6 (1.6)	0.626
- A person that is unhealthy cannot be happy†		3.0 (1.8)		2.7 (1.7)	0.003

Note: * Low = lower vocational or primary school. Middle = middle vocational or secondary school. High = higher vocational or academic. †scale from 1 (totally disagree) to 7 (totally agree).

7.3. RESULTS

7.3.1. Sample

The questionnaire was administered online to an internet panel and we obtained a sample of 1,015 respondents, representative of the population of the Netherlands in terms of age, gender and education level. About half of the sample (n= 520; 51.2%) was identified as non-trader. The other 495 respondents did trade off health and happiness in at least one of the scenarios and were assigned to the 'traders' subgroup. Non-traders were significantly older and more often female (Table 7.1). As explanation for not trading-off, 63.5% consistently answered that everyone is equally important to treat, irrespective of differences in happiness (scenario 1) or health (scenario 2).

Non-traders agreed more often with the statements "All patients are equally entitled to health care, irrespective of the effect of the illness on their happiness" and "The ultimate goal of the health care system is to promote health", but less often with "Happiness is more important than health" and "A person that is unhealthy cannot be happy" (Table 7.1).

Given their explanation to the choices in the scenario's and their response to the statements it seems that this subgroup largely consists of genuine non-traders, unwilling to differentiate between patient groups according to health or happiness levels in priority-setting.

7.3.2. Health versus happiness trade-offs

Table 7.2 shows the choices of the traders in scenario 1. Respondents considered the effect on happiness (i.e., choose Group B) more often when the start level of health was higher. Overall health effects received more weight than happiness effects, but relative weights came closer to one when starting levels were lower. This may relate to a ceiling effect in the experiment, which is more likely when starting levels are higher.

Table 7.2: Choices in scenario 1; underlined number varied with choices made (n=495)

Ver.	Baseline health / happiness	N	Initial choice			Always prefer B[a]	Weights[b]	SD	Median	Health > happiness[c]	Happiness > health[d]
			A	B	Dif.						
1	7/8–7/6	135	17.8%	68.1%	14.1%	25.0%	0.61H = 1WB	0.32	0.5	0.285**	-0.185
2	6/7–6/5	115	21.7%	48.7%	29.6%	23.0%	0.68H = 1WB	0.51	0.5	0.461**	-0.317*
3	5/6–5/4	108	19.4%	47.2%	33.3%	15.7%	0.78H = 1WB	0.63	0.5	0.143	-0.206
4	4/5–4/3	137	28.5%	46.0%	25.5%	15.9%	0.88H = 1WB	0.80	0.5	0.339**	-0.208

Note: *p<.05;**p<.01. [a] Percentage of those with initial choice 'Group B'. [b] H=health, WB=wellbeing (happiness). [c] Correlation coefficient between relative weight and agreement with statement 'Health is more important than happiness'. [d] Correlation coefficient between relative weight and agreement with statement 'Happiness is more important than health'.

The sign of the correlation coefficients between relative weights and response to the statements 'health is more important than happiness' and 'happiness is more important than health' (see last column of Table 7.2) indicates a consistent preference for health over happiness.

Table 7.3 shows the choices in scenario 2. The distribution of initial choices was comparable to scenario 1, although respondents more often 'always preferred group B', the group with the lowest health level, irrespective of happiness level. This effect was stronger when the start level of health was lower. The relative weights of health and happiness were also comparable to those found in scenario 1, except in version 1. This is probably the result of a ceiling effect, as the relative weight in this version by design was lower than one. The correlation coefficients between weights and statements show that respondents still found health more important than happiness (see last column of Table 7.3).

Table 7.3: Choices in scenario 2; underlined number varied with choices made (n=495)

Ver.	Baseline health / happiness	N	Initial choice			Always prefer B[a]	Weights[b]	SD	Median	Health > happiness[c]	Happiness > health[d]
			A	B	Dif.						
1	8/7–6/7	108	14.8%	68.5%	16.7%	29.0%	1H = 0.78 WB	0.27	1.0	-0.183	0 .271*
2	7/6–5/6	137	21.9%	65.5%	13.1%	32.0%	1H = 1.00 WB	0.46	1.0	-0.217*	0.323**
3	6/5–4/5	135	14.8%	57.8%	27.4%	42.0%	1H = 1.12 WB	0.70	0.95	-0.276*	0.203
4	5/4–3/4	115	14.8%	61.7%	23.5%	36.0%	1H = 1.35 WB	0.88	1.0	-0.242*	-0.066

Note: *p<.05;**p<.01. [a] Percentage of those with initial choice 'Group B'. [b] H=health, WB=wellbeing (happiness). [c] Correlation coefficient between relative weight and agreement with statement 'Health is more important than happiness'. [d] Correlation coefficient between relative weight and agreement with statement 'Happiness is more important than health'.

7.4. DISCUSSION AND CONCLUSIONS

The aim of this explorative study was to gain more insight in the relative importance of health and happiness in health care priority setting by the members of the public. This experiment was performed against the background of the debate on broader outcome measures and experienced wellbeing in (economic evaluations of) health care.

Before discussing the main implications of our findings, we stress that our experiment had some noteworthy limitations. First, the range in which respondents were able to trade-off happiness and health effects was relatively limited. Therefore, especially in the scenarios with higher starting levels respondents could not always indicate their indifference point precisely. Secondly, the health attribute was always presented first (i.e. above) the happiness attribute in the choice sets and we varied only health in scenario 1 and only happiness in scenario 2. This may have drawn respondents' attention more

to one attribute than to the order, possibly resulting in order or focus bias. In order to overcome these first two limitations, we designed scenario 3, in which health and happiness levels were varied simultaneously. However, this proved to be too complex for respondents resulting in many inconsistent answers. Future studies may use simpler designs or a different study setting (e.g. face-to-face). Thirdly, in order to keep the exercise as simple as possible, we asked respondents to trade-off health scores with happiness scores, without further definition. However, as warm-up question to familiarize respondents with the outcome measures and measurement scale used in the experiment, they were asked about their own health and happiness levels. Nonetheless, respondents may have given own interpretations to the content of health and happiness ratings. Given our main aim, this may not be problematic. Still, future research could make the concepts more precise when presenting them to the respondents. Fourthly, we chose not to focus on the issue of experienced versus expected utility as we felt this would make the exercise even more complex. Yet, even *within* the health and happiness domains, it remains interesting to see how people would trade off these outcomes. Fifthly, in the design of the experiment we presented two groups of patients affected by two distinct diseases, with different effects of the diseases and respective treatment on health and happiness. This allowed us to vary the levels of health and happiness before and after treatment. However, this may also have caused respondents to base their choices on other elements than the benefit of improvements in health and happiness, such as equity considerations or expectations regarding the type of disease involved. Future studies may ask respondents to make a choice between treatments for the same group of patients, to increase their focus on the trade-off between health and happiness. Finally, the choice task presented to respondents may have been difficult. While these limitations warrant caution in interpreting the relative value of health and happiness for decision making in health care, some interesting general findings can be highlighted.

First, about half of the respondents were unwilling to discriminate between patient groups on the basis of their health and happiness levels. This may reflect concerns for equity. Recent studies have shown that a considerable portion of the public may have fundamental difficulties with making choices between patient groups based on patient, disease or treatment characteristics as well as on health and broader effects of treatment (e.g., [91]).

Second, hypothesis A was that individuals would consider both health and happiness levels in priority-setting decisions. Our results indicate that, within the trader group, this was largely confirmed. Hence, when provided information on subjective wellbeing (here expressed as happiness), health care decisions are not based solely on health maximization considerations, but also on levels of and gains in happiness. Still, traders in general

attached more weight to health (gains) than to happiness (gains) in making trade-offs, thus confirming hypothesis B.

It is also important to stress that our results indicate that people appear more willing to discriminate between groups of patients when health and happiness levels before treatment are higher. Especially when people were asked to discriminate based on different happiness levels respondents were more willing to do so when health and happiness levels before treatment were higher. This may reflect a general interest in the distribution of health and happiness, next to the size of gains.

This experiment was a first, explorative step in investigating the relevance and relative weight of the outcome measures health and happiness for allocation decisions in health care. While we stress the limitations of this study, in light of current debates on the appropriate (mix of) outcome measures to guide allocation decisions in health care, it nonetheless revealed some interesting results. Importantly, they suggest that from the perspective of the public both health and happiness could play a role in priority setting. While the debates so far have focused on a 'either or' choice between these outcomes, in part perhaps because of the interdependency between them, our results indicate that both receive weight in choices between interventions. A full welfare economic assessment of interventions therefore may require a broader set of outcome measures than currently the case [29]. "Health or happiness?" may thus not be the right question, as both seem to matter!

APPENDIX 7.1: EXAMPLE QUESTION (VERSION 1, SCENARIO 1, CHOICE ❶)

Introduction

When a person becomes ill, this may affect this person's happiness as well as their health. The influence of illness on happiness may vary between diseases. This may depend on the type of disease, the impact of the disease on daily life, and how well people can adapt to the limitations from the disease. For example, it may matter to people whether they have problems walking because of the disease, or have pain, or psychological problems. In addition, it can make a difference whether the disease has an effect on work or social life, and the extent to which people can learn to cope with the limitations.

Question

Please imagine that two comparable groups of patients are affected by different diseases. Before they developed the disease, they were in good health and happy. They gave their health and happiness levels both a 9 on a scale from 0 to 10, with 0 representing 'worst conceivable health' / 'completely unhappy' and 10 'best conceivable health' / 'completely happy'.

As a result of disease, health levels dropped to 7 in both groups. Furthermore, in group A happiness dropped to 8 and in group B happiness dropped to 6.

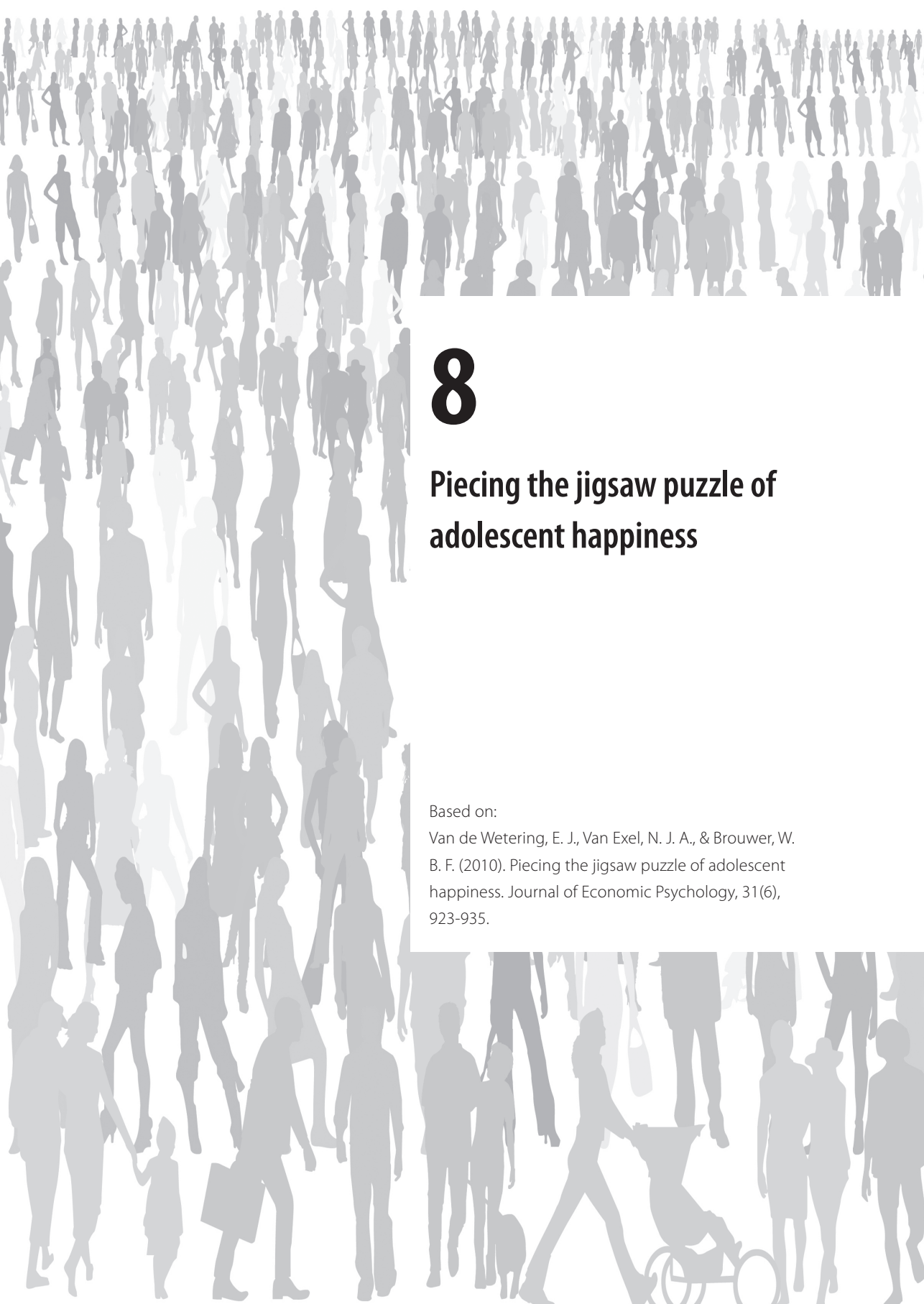
Imagine that you are a decision maker and you have to decide which one of the two groups of patients will receive treatment. Treatment will bring patients in that group back to their original levels of health and happiness of 9 on a scale from 0 to 10, whereas the untreated group will remain at the lower levels indefinitely.

Both groups have the same size and treatment costs are the same.

	Group A			Group B		
	Before treatment	After treatment	Gain from treatment	Before treatment	After treatment	Gain from treatment
Health	7	9	+2	7	9	+2
Happiness	8	9	+1	6	9	+3

As a decision maker, which group of patients would you give treatment?

- ☐ I would give treatment to Group A
- ☐ I would give treatment to Group B
- ☐ I can't choose between the two groups, for me both are equally important to receive treatment



8

Piecing the jigsaw puzzle of adolescent happiness

Based on:
Van de Wetering, E. J., Van Exel, N. J. A., & Brouwer, W. B. F. (2010). Piecing the jigsaw puzzle of adolescent happiness. Journal of Economic Psychology, 31(6), 923-935.

ABSTRACT

Happiness is increasingly recognized as a proxy for utility and therefore a valuable maximand for policy decisions. As a result many studies have investigated happiness and the associated determinants in both overall and specific life domains. Adolescent happiness, however, remains largely unexplored. The aim of this study is to explore the relative importance of happiness of young Dutch adolescents at home, school, and leisure, and their associations with a broad array of personal and context characteristics within each of these domains.

We used data from a study which investigated adolescents' health behaviour in relation to their attitudes regarding health and lifestyle as well as their considerations and expectations regarding the future consequences of their behaviour (n=1,436). Variables were selected on the basis of findings in the literature or significant univariate Pearson correlations between the variable and domain-specific or overall happiness. The data was analysed using multiple hierarchical stepwise regressions.

In line with international findings, most adolescents reported high levels of overall happiness with a mean score of 7.69 (SD= 1.23) on a scale from 0 to 10. Personal and context characteristics were associated with adolescent overall happiness either directly or indirectly, via domain specific happiness. Happiness at home, at school, and during leisure hours contributed approximately equally to overall happiness but were associated with different characteristics. Finally, the results demonstrate that adolescents differentiate their happiness levels between life domains, which support the relevance of a multidimensional approach in happiness studies.

This study provides additional insight over single dimensional studies of happiness and a more comprehensive explanation of previously published findings.

8.1. INTRODUCTION

Happiness is typically conceived as a subjective measure of the overall enjoyment of life, generally defined as 'the degree to which an individual judges the overall quality of his life favourably' [31] and is frequently designated as an important life goal [175]. In disciplines like psychology, biology, and sociology, happiness has been widely accepted as a significant concept [176, 177].

The concept of happiness has a long standing, yet controversial, tradition in economics. It can be traced back to the work of the utilitarian philosopher Jeremy Bentham (1789), who used the pursuit of 'the greatest happiness for the greatest number' as a central argument in his reasoning [178]. Whether 'happiness' can be equated with 'utility' or 'welfare', however, has been long-disputed. The use of happiness became controversial when economists started to question whether subjective well-being as a measure of welfare could be a good and stable indicator – or even synonym – for 'utility', especially since it cannot be measured objectively and interpersonal comparisons may thus be considered 'unscientific', also because of differences in reference point, coping etcetera. At the time, economics therefore moved away from using (self-reported) well-being [179-183]. More recently, however, the study of happiness has regained popularity in economics and studies are increasingly using it as a proxy for utility [184-186]. In this genre, happiness is considered to be a (partial) index of welfare, or as Ferrer-i-Carbonell and Frijters (2004) describe it, 'a positive monotonic transformation of an underlying metaphysical concept called welfare' [5].

Many studies have been conducted to better understand what constitutes happiness, showing that determinants of happiness may differ across individuals, life stages, and life domains [33]. The happiness of adolescents (ages 12 to 18), however, remains largely unexplored. At present, our knowledge of what constitutes adolescent happiness can best be seen as a large, unsolved jigsaw puzzle. Some studies have explored the determinants of overall adolescent happiness or happiness within a specific life domain [35, 187]; others have focused on the relations *between* domain-specific and overall adolescent happiness [188]. What is lacking, however, are studies investigating the relations between overall and domain-specific happiness alongside a broad array of possible determinants. In other words, many, but certainly not all, pieces of the adolescent happiness jigsaw puzzle have been identified and we are only at the very early stages of piecing it together.

This article aims to contribute to the completion of this puzzle by exploring the relative importance of happiness of young Dutch adolescents at home, at school, and during

leisure hours to their overall happiness, in association with a broad array of personal and context characteristics.

The structure of the article is as follows. First, we present an overview of the relevant recent literature on adolescent happiness. We then describe the dataset and statistical analyses, followed by the results of the study. Last, we discuss our findings.

8.2. BACKGROUND

A number of studies have focused on overall happiness of adolescents. Although different measurement instruments have been used, it appears that most adolescents rate their overall happiness levels positively [189]. UNICEF (2007) initiated a study on the overall well-being of adolescents in 21 industrialized countries. Well-being was assessed in six different dimensions: material well-being, health and safety, education, peer and family relationship, behaviours and risks, and peoples' own subjective sense of well-being. Within this study young adolescents were asked to rate their overall life satisfaction on a 'life satisfaction ladder' (0 =worst possible life, 10 = best possible life; scores above the midpoint, i.e., scores of 6 or more were treated as positive). A great majority of young adolescents scored their life satisfaction positively, ranging from 77.4% (15-year-old girls) to 88.1% (11-year-old boys). The Netherlands reported the highest overall well-being of young adolescents: more than 90% of adolescents between the ages of 11 and 15 rated life satisfaction positively.

The studies investigating determinants of adolescent happiness have largely focused on the influence of personal characteristics. Some studies have reported that adolescents differentiate their happiness levels across life domains (e.g. UNICEF 2007), indicating that differentiation of the analysis to specific life domains may provide incremental information in relation to overall happiness [190]. We will first discuss the influence of personal characteristics on adolescent happiness, followed by domain-specific variables that may influence (domain-specific) happiness.

Beginning with personal characteristics, age was shown to be an important determinant of adolescent happiness. Happiness scores tend to decrease through the teenage years with the lowest level at age 16, followed by a small recovery up to age 18 [191]. Contradictory results have been reported regarding the association between gender and adolescent happiness. A number of studies have reported significant differences in happiness scores between boys and girls, with girls generally being happier [192, 193], while a few others have found no significant association between gender and adolescent happiness [187, 191].

In 2005 Mahon, Yarcheski et al. studied the relation between adolescent happiness and health. The results show significant and positive correlations between adolescent happiness and three aspects of health: perceived health status, wellness, and clinical health, with correlation coefficients of 0.61, 0.55, and 0.14, respectively. Given the strength of the correlations it seems that subjective perceptions of health are more strongly related to happiness than objective measures. This was consistent with earlier findings [187].

No clear evidence has been found regarding the relation between social class and adolescent happiness. Some studies have reported no significant differences between social classes, while others have noted small differences in favour of higher social classes [190]. One study reported large differences in happiness between social classes, but the relation was arbitrary: Young people living in working class communities on average reported the highest level of happiness, followed by middle class, poor, and upper class, and the lowest level of adolescent happiness was found in the upper middle class [191].

Whereas Robbins and Francis (1996) reported a positive association between religious beliefs and happiness, Lewis and colleagues (1997) found no clear relationship in their study. These differences could however be explained by the use of different measurement techniques or definitions of happiness and religiosity [194].

Many studies have reported the importance of personality traits. The majority of these studies indicated that extraversion is positively associated with adolescent happiness whereas neuroticism shows a negative association to adolescent happiness [35]. Lounsbury, Saudargas et al. (2005) moreover observed that the importance of personality traits differed between general and domain-specific life satisfaction in a sample of college students.

Adolescent happiness has been shown to be related to domain-specific variables as well. In this study we address three central life domains: home, school, and leisure time (for reasons explained later). Evidence from studies that have focused on the relationship between home-related variables and adolescent happiness demonstrated that satisfaction with family relationships is strongly associated with adolescent happiness [195]. Suldo and Huebner (2004) concluded that an authoritative parenting style was positively associated with happiness among youth. Social support appeared to be most strongly associated, but 'strictness' and 'psychological autonomy granting' were also significantly related to adolescent happiness. In addition, the more parents' and youths' beliefs, views, and attitudes were in agreement, the higher the happiness level reported by the youth. Therefore, communication between parents and their children seems to be very important [195, 196]. No significant differences in happiness levels were found between children with and without siblings [197].

Several studies have explored the associations between school-related variables and adolescent happiness. Self-evaluated school performance, rather than actual grades, was found to be positively associated with adolescent happiness [198]. In addition, Csikszentmihalyi and Hunter (2003) observed lower than average happiness levels during time spent on certain school-related activities such as listening to lectures. Finally, Park (2004) reported that perceived safety at school was significantly related to overall happiness. Youths who felt unsafe or had been victims of psychologically violent behaviour tended to report lower happiness levels.

Csikszentmihalyi and Hunter (2003) included several leisure time variables in their study of adolescent happiness. Their results showed that doing homework or reading a book, even for pleasure, lowers happiness. Several studies have shown an association between the number of friends and happiness. Young people that socialize more with peers were found to be happier, and young people who spent more time alone were generally less happy [191]. Cheng and Furnham (2002) have also found that peer friendship is a direct predictor of adolescent happiness. Demir, Ozdemir, and Weitekamp (2007), however, have concluded that while many studies suggest that an individual benefits from having a large number of friends, it is actually only 'having a best friend' that contributes to happiness. Finally, involvement in sports appeared to be positively associated with leisure time satisfaction as well as with overall happiness [35].

While a growing number of studies addressed the importance of domain variables and domain-specific happiness, the majority of studies have nonetheless focused on the association of determinants with overall happiness and have not considered the relevance of happiness in different life domains [190]. Other studies have investigated domain specific happiness and overall happiness, but have focused on happiness in only one specific life domain or included only very few underlying determinants [34, 35, 193]. Besides, a variety of methods and happiness measures have been used, making it difficult to compare results between studies.

To summarize, past studies have identified a range of determinants of adolescent happiness, but the findings have been ambiguous and have varied across studies. In particular the personal characteristics 'health' and 'personality traits' have been reported as important determinants of adolescent happiness.

The current study aims to clarify how a variety of determinants used in past studies fit in the overall jigsaw puzzle of adolescent domain specific and overall happiness and to identify additional determinants which to date have remained undetected.

8.3. METHODS

8.3.1. Participants

We used secondary data from a study which investigated adolescents' health behaviour in relation to their attitudes regarding health lifestyle and their expectations and consideration of the future consequences of their health behaviour [165, 199]. The original dataset was collected in May 2005 by means of an extensive questionnaire (the "Health & Future" questionnaire; see Van Exel, Koolman et al. (2006)) and included 2,006 adolescents from 12- to 14-year-old (53% girls, M=13.2, SD=0.68) in the first and second grade of pre-vocational and general secondary education, from ten variously-sized schools in variously-urbanized areas throughout the Netherlands. Female gender, general secondary education level and Dutch ethnicity were slightly overrepresented, so that the sample was not entirely representative for the reference population. For more details see Van Exel, Koolman, et al. (2006).

8.3.2. Materials

The participants completed the questionnaire in class and under teacher supervision. It contained multiple-choice items and closed-end scales addressing a wide range of variables, many of them potentially associated with happiness. Of the eight sections of the questionnaire, happiness was measured in the following: 'about you' (overall happiness), 'about home', 'about school', and 'about your leisure time' (domain-specific happiness measures). Our study uses the data that resulted from these four sections.

Since happiness is a subjective measure, there has been much discussion in the literature about its measurability. Research has shown that the validity of happiness measured by self-report is good [32]. In this study, participants assessed their happiness by using a rating scale that ranged from 0 (very unhappy) to 10 (very happy). The exact happiness question for overall happiness was posed as follows:

Please grade your happiness. Place a mark on the scale below on the place that best fits with how happy you feel, in general. The '0' means "very unhappy" and the '10' means "very happy".

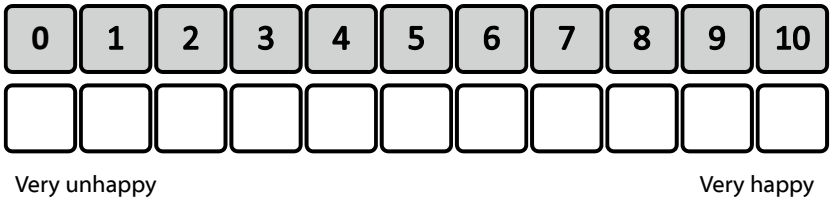


Figure 8.1: Happiness rating scale

The domain specific happiness questions were posed identically, with the following additions: (i) the first sentence was phrased 'Please grade your happiness at home' (or 'at school', 'in leisure time'); (ii) the final part of the second sentence included the same reference, i.e. '…how happy you feel at home, in general' (or, again, 'at school', 'in leisure time').

To assess personality, the questionnaire contained a 30-item short version of the Goldberg's adjective 100 list for the Big-Five personality dimensions: openness to experience, conscientiousness, extraversion, agreeableness and neuroticism [200]. This short version was used and validated in similar Dutch populations [201, 202].

8.3.3. Procedure

Selection of the variables and the sample

A database containing the full list of variables from the four sections of the questionnaire mentioned above was scanned for potentially relevant variables for our analysis. Variables were selected on the basis of the literature discussed in the previous section (henceforth called the 'core variables') and on univariate Pearson correlations between the variable and happiness (henceforth called the 'additional variables'). The latter variables were retained in case of a statistically significant ($p<0.05$) association with either overall or domain-specific happiness. The final selection of variables for analysis is pre-

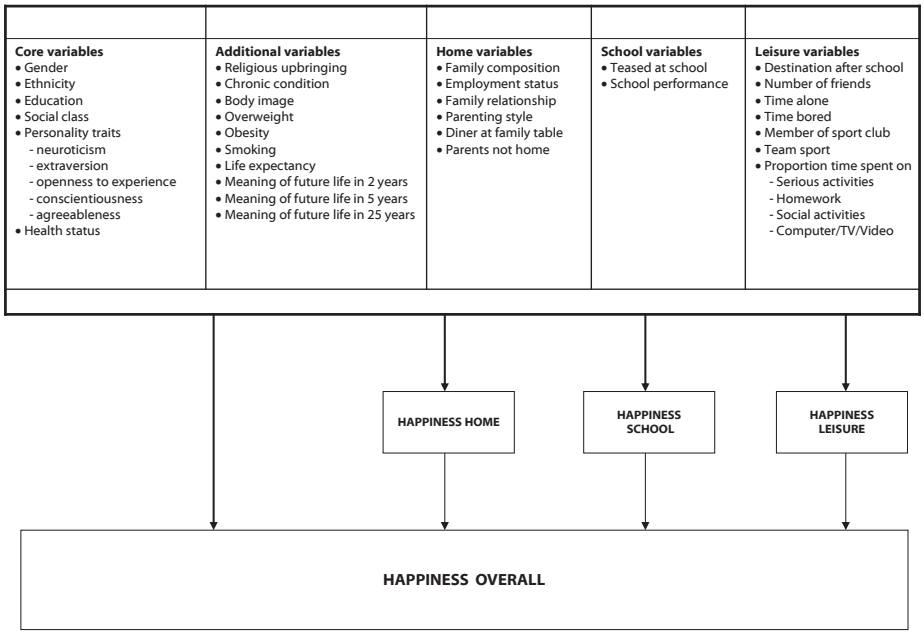


Figure 8.2: Variables selected for analysis

sented in Figure 8.2, which also provides a general conceptual model for our analyses. Within this model a distinction was made between personal characteristics based on the literature and those selected as a result of their univariate associations with happiness.

Observations with a missing value on any of the selected variables were excluded from further analysis. An independent-samples t test showed there were no systematic differences in age, gender, health, or happiness levels between included and excluded respondents. Educational level, however, significantly differed: the included observations showed a relatively higher percentage of general secondary education ($p<0.001$).

Statistical Analysis

Previous happiness studies have demonstrated that assumptions about a cardinal or ordinal interpretation of happiness are relatively unimportant to the study results [32, 186, 203, 204]. Therefore, we primarily applied a cardinal interpretation of the happiness data and used ordinary least square (OLS) regression techniques in the analyses. In order to check the implications of the cardinality assumption for the results, ordinal logit analyses were conducted as well (using the same explanatory variables). The results of this additional analysis are presented in appendix 8.1.

First, the percentage of respondents with identical levels of general and domain-specific happiness was calculated to verify whether respondents differentiated their happiness scores between general and domains. Next, the associations between overall and domain-specific happiness were explored. Paired-sample t tests were performed to test whether the average happiness levels statistically differed between overall happiness, happiness at school, happiness at home, and happiness during leisure time. Additionally, Pearson's correlations and multivariate regression analysis were conducted to explore associations between overall and domain-specific happiness and the relative impact and explanatory power of domain-specific happiness on overall happiness.

Subsequently, a series of multiple hierarchical stepwise regressions were conducted to explore the associations of the explanatory variables with adolescents' domain-specific happiness. The ordering of variables was based on Figure 8.2. The first step included the core individual variables, which originated from the overview of the literature, and were entered all at once. In step two, the additional individual variables and the domain-specific variables were added using the stepwise regression method to make a statistical selection from these variables [205]. Variables were entered (and removed again) depending on the statistical significance of their coefficients ($\alpha=0.10$ for entry and $\alpha=0.15$ for removal). This procedure was repeated for the three life domains and overall happiness. Robustness checks, varying the threshold values of the stepwise regressions, did not notable influence the results.

To determine possible differences in explanatory power of the personal characteristics (core individual variables) and domain variables in relation to domain-specific happiness, we assessed the adjusted R square. Furthermore, in order to explore direct and indirect effects of the variables on overall happiness, we conducted additional regression on overall happiness following the same procedure but including the domain-specific happiness measures as control variables (forced into the model, between steps 1 and 2).

Statistical analyses were conducted using the statistical package SPSS version 17.0.

8.4. RESULTS

8.4.1. Characteristics of the Study Population

The final study population included 1,436 adolescents, of which 54.7% were female and 91.2% were Dutch. The mean age of the study sample was 13.2 years. Most adolescents reported high levels of overall happiness, with a mean score of 7.96 (SD=1.23) (Figure 8.3). The descriptive statistics are reported in Table 8.1.

8.4.2. Happiness Scores

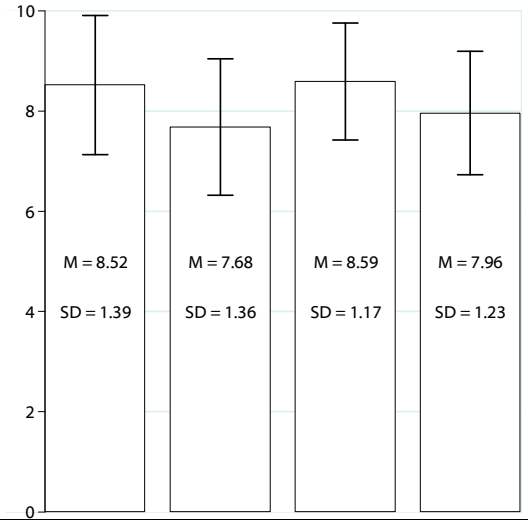
The domain-specific and overall happiness levels are presented in Figure 8.3. Mean happiness levels ranged from 7.68 to 8.59 on a scale of 0 to 10. On average, young adolescents were happiest during their leisure time, less happy at home, and least happy at school. Most respondents had variations in domain-specific happiness; only 7.2% of the sample rated their domain-specific and overall happiness levels all the same. Paired sample t tests showed that mean happiness scores differed statistically significantly between the specific life domains and overall happiness ($p<0.001$). Furthermore, although the domain-specific and overall happiness scores were positively and significantly correlated, each life domain represented substantial unique variance not accounted for by the other life domains, as indicated by moderate correlation coefficients, ranging between 0.275 and 0.430 (see Figure 8.3).

The standardized multiple regression (Beta) coefficients indicate the relative importance of happiness between the three domains. Although the differences were small, happiness at school and at home received the highest weights. Together, the three domain-specific happiness measures explained 31.8% of the variance in overall happiness (Figure 8.3).

8.4.3. Domain-specific Happiness

Happiness at Home

The core individual variables accounted for 13.1% of the variance in happiness at home (Table 8.2). The second step of the regression added 10 extra explanatory variables to



	Happiness Home	Happiness School	Happiness Leisure time	Happiness Overall	Standardized Beta
Happiness Home		0.316**	0.334**	0.430**	0.268**
Happiness School			0.275**	0.419**	0.271**
Happiness Leisure time				0.394**	0.229**
Happiness Overall					
Adjusted R^2					0.318

Note: ** $p < 0.001$

Figure 8.3: Average happiness scores, correlations and regression results (n=1,436)

the model and raised explained variance to 26.0%. The results indicate that girls were less happy than boys. Adolescents who rated their family as less wealthy than others reported lower scores for happiness at home. Better health had a significant positive effect on happiness. A higher life expectancy was associated with marginally higher happiness scores.

Several factors within the home category were significantly associated with happiness. Adolescents who receive compliments from their parents and those experiencing supportive behaviour from parents reported higher happiness scores, while adolescents from a single-parent family or with strict parents reported lower levels of happiness. Furthermore, from the leisure time variables, adolescents going directly home from school reported higher happiness levels. Adolescents that often felt alone or bored showed significantly lower happiness levels. Having three or more friends significantly increased reported happiness levels. Finally, better self-reported performance at school was associated with being happier at home.

Table 8.1: Descriptive statistics (n = 1,436)

Variable	Category	Share*	Mean	SD
Core variables				
Gender	Female	54.7		
Ethnicity	Dutch†	91.2		
Education level	General Secondary	62.5		
Social class	Less wealthy than other Dutch families	7.6		
Health Status	0-10 Scale (0/10 best/worst conceivable)		7.72	1.24
Additional variables				
Religious upbringing	Yes	74.4		
Chronic condition	Yes, at least one chronic condition	14.8		
Body image~	Too thick	32.9		
Overweight^	Yes	7.8		
Obesity^	Yes	0.6		
Smoking	Yes	6.5		
Total monthly allowance (€)			45.6	64.5
Life expectancy	Years of age		85.13	10.79
Meaning of future life in 2 years °	(very) important	89.8		
Meaning of future life in 5 years °	(very) important	93		
Meaning of future life in 25 years °	(very) important	89.1		
Home variables				
Family composition				
Single parent	Yes	10		
Only child	Yes	6		
Employment status of parents	Both unemployed	1.3		
Satisfied with relationship mother	1. (Totally) Disagree	7.2		
	2. Agree	57.1		
	3. Totally agree	35.7		
Satisfied with relationship father	1. (Totally) Disagree	9.6		
	2. Agree	55.7		
	3. Totally agree	34.7		
Parenting style				
Parents are strict	Yes	26.7		
Parents are interested	Yes	95.1		
Parents give compliments	Yes	93.7		
Having dinner at family table	Daily	87.5		
At least one parent at home after school	Yes	89.1		

Table 8.1: Descriptive statistics (n = 1,436) (continued)

Variable	Category	Share*	Mean	SD
School variables				
Teased at school	Never	84		
School performance	1. Very good	14.2		
	2. Good	48.6		
	3. Satisfactory	27.8		
	4. (Very) Unsatisfactory	9.4		
Leisure time variables				
Destination after school	Going home after school	90.5		
Number of good friends	Three or more	87.5		
Time Alone	1. (Very) Often	11		
	2. Sometimes	39.5		
	3. Rarely	39		
	4. Never	10.5		
Time Bored	1. (Very) Often	4.9		
	2. Sometimes	32		
	3. Rarely	42.5		
	4. Never	20.6		
Member of sports club	Yes	57.7		
Team sport	Yes	27.7		
Proportion time spent on…[#]				
Homework			0.13	0.09
Serious/ hobbies			0.26	0.14
Social activities			0.26	0.12
Computer/video			0.34	0.17

*Note: Total equals 100%; [†] Both parent are born in the Netherlands; ~ Subjective evaluation of their own body; ^ Based on international cut-off points for overweight and obesity in adolescents by gender and age [203]; º Respondents were asked how important it is to them what their life will be like in 2,5,25 years; [#] Proportions of weekly time spent on hobbies.

Happiness at School

The core individual variables were responsible for 17.2 % of the variance, while the total regression accounted for 23.9% of the variances in happiness at school. The regression results show that girls were significantly happier at school than boys. Both being less wealthy and being healthier were positively associated with happiness at school. The personality trait agreeableness was also positively associated with happiness at school, whereas extraversion and neuroticism were negatively associated with it. Adolescents who placed relatively more importance on their situation five years in the future (i.e., about the end of secondary school) and those with a religious upbringing were hap-

Table 8.2: Results multiple hierarchical regressions (n=1,436)

Variables[†]		Happiness Domains			Happiness Overall			
		Home	School	Leisure Time	Excluding domains		Including domains	
		B[˜]	B[˜]	B[˜]	B[˜]	Beta[‡]	B[˜]	Beta[‡]
Core individual variables								
Gender	Female = 1, Male = 0	-0.140*	0.136*	-0.127*	-0.061	-0.025	-0.024	-0.010
Ethnicity	Dutch = 1, Other = 0	0.023	0.172	0.245*	0.234**	0.054*	0.188*	0.043*
Education	Higher = 1, Lower = 0	-0.125	-0.063	0.052	0.042	0.017	0.072	0.028
Less wealthy	Yes = 1, No = 0	-0.278*	0.293*	-0.125	-0.032	-0.007	-0.007	-0.002
Health	0-10 Scale (worst conceivable = 0, best conceivable = 10)	0.162**	0.213**	0.166**	0.303**	0.306**	0.216**	0.217**
Agreeableness	Continuous variable (higher score: more agreeable)	0.141**	0.259**	0.149**	0.230**	0.185**	0.140**	0.112**
Conscientiousness	Continuous variable (higher score: more conscientious)	0.042	-0.012	-0.053	0.032	0.026	0.046	0.038
Openness to experience	Continuous variable (higher score: more open to experience)	0.066*	0.058	0.014	0.044	0.036	0.015	0.012
Extraversion	Continuous variable (higher score: more introverted)	-0.040	-0.092*	-0.078*	-0.179**	-0.147**	-0.160**	-0.132**
Neuroticism	Continuous variable (higher score: more neurotic)	-0.051	-0.082*	0.036	-0.107**	-0.087**	-0.101**	-0.082**
Constant^		5.275**	3.830**	8.011**	5.126**		1.969	
Adjusted R^2		*0.131*	*0.172*	*0.118*	*0.290*	*0.290*	*0.290*	*0.290*
Happiness domains								
Happiness Home	0-10 Scale (very unhappy = 0, very happy = 10)						0.179**	0.204**
Happiness School	0-10 Scale (very unhappy = 0, very happy = 10)						0.159**	0.176**
Happiness Leisure time	0-10 Scale (very unhappy = 0, very happy = 10)						0.164**	0.156**

Table 8.2: Results multiple hierarchical regressions (n=1,436) (continued)

Variables†		Happiness Domains			Happiness Overall			
		Home	School	Leisure Time	Excluding domains		Including domains	
		B˜	B˜	B˜	B˜	Beta‡	B˜	Beta‡
Additional individual variables								
Smoking	Yes = 1, No = 0		-0.249		-0.276*	-0.056*	-0.247*	-0.050*
Overweight	Yes = 1, No = 0		0.200					
Too thick	Yes = 1, No = 0				-0.106	-0.041	-0.105	-0.040
Religious upbringing	Yes = 1, No = 0		0.173*	-0.113				
Life expectancy	In years	0.007*	0.008*					
Meaning of future life in 5 years	(very) important = 1, (very) unimportant = 0		0.240	-0.236*				
Monthly budget	€				0.001*	0.045*	0.001*	0.045*
School variables								
School performance	(Very) unsatisfactory = 0, Satisfactory=1, Good=2, Very good = 3	0.181**	0.241**		0.156**	0.106**	0.082*	0.056*
Teased at school	Never = 0, Sometimes/ Often = 1		-0.418**		-0.177*	-0.053*		
Home variables								
Satisfied with relationship mother	Totally disagree = 0, Disagree = 1, Agree = 2, Totally agree = 3			0.084	0.102*	-0.052*	0.092*	-0.047*
Only Child	Yes = 1, No = 0				0.274*	0.053*	0.280*	0.054*
Parents are strict	Yes = 1, No = 0	-0.519**			-0.114	-0.041		
Parents are interested	Yes = 1, No = 0	0.603**	0.521**					
Parents give compliments	Yes = 1, No = 0	0.898**		0.280*	0.353*	0.070*		
Single Parent	Yes = 1, No = 0	-0.378**						

Table 8.2: Results multiple hierarchical regressions (n=1,436) (continued)

Variables†		Happiness Domains			Happiness Overall			
		Home	School	Leisure Time	Excluding domains		Including domains	
		B˜	B˜	B˜	B˜	Beta‡	B˜	Beta‡
Leisure time variables								
Going home after school	Yes = 1, No = 0	0.343*					-0.204*	-0.049*
Friends	Less than three= 0, Three or more = 1	0.230*	0.419**		0.165*	0.045*		
Proportion time homework				-1.209**				
Proportion time computer/ video					0.505*	0.069*	0.461*	0.063*
Member of sports club	Yes = 1, No = 0		-0.153*				0.113*	0.046*
Alone	Never= 0, Rarely=1, Sometimes = 2, (Very) Often =3	-0.219**	-0.104*	-0.229**	-0.131**	-0.088**		
Bored	Never= 0, Rarely=1, Sometimes = 2, (Very) Often =3	-0.112*		-0.392**	-0.110*	-0.074*		
Adjusted R^2		*0.260*	*0.239*	*0.247*	*0.335*	*0.335*	*0.426*	*0.426*

Note: † the variables: chronic condition, employment status, obesity, meaning of future life in 2 and 25 years, diner at family table, parents home after school, team sport, proportion time spent on serious activities and proportion time spent on social activities, of Figure 8.2 were excluded as a result of the stepwise hierarchical regression analyses; ˜ unstandardized coefficients, ‡ standardized coefficients; * p<0.05, ** p <0.001; ^ constant varies with each step of the regression.

pier at school. The same was true for those who were overweight. Smokers, however, reported lower happiness at school.

Both school performance and being teased at school showed a significant effect on happiness at school, with plausible signs. Of the home-specific variables, parental involvement with the child was significantly associated with higher levels of happiness at school. Adolescents with more friends (a leisure time variable) showed higher school-related happiness scores, whereas being alone more often and, surprisingly, being a member of a sports club were associated with lower levels of happiness at school.

Happiness in Leisure Time

The core variables explained 11.8% of the variance, the lowest of the life domains, while the overall set of explanatory variables accounted for 24.7% of the variance in happiness during leisure time. Boys and adolescents with Dutch ethnicity tended to be happier during their leisure time in comparison to girls and non-Dutch adolescents. The same holds for those who were healthier and those who were more agreeable. Adolescents with higher scores on the personality trait extraversion, those with a religious upbringing or those who attached relatively low weight to their situation in five years reported lower happiness levels during leisure time.

With respect to the leisure time variables, adolescents who spent a higher proportion of their time on homework and those who were more lonesome or bored reported lower happiness scores. School variables did not contribute to explaining leisure time happiness. The home variable 'parents give compliments' was significantly and positively associated with happiness.

8.4.4. Overall Happiness

The core individual variables accounted for 29.0% of the variance in overall happiness, which was considerably higher than in the domain-specific models, whereas the domain-specific explanatory variables contributed far less. Many of the same core and additional individual characteristics as in the previous models were associated with overall happiness, with the exception of body image (i.e., adolescents with a bulky body image reported lower overall happiness) and monthly budget (i.e., having more to spend was associated with higher overall happiness). The variables 'only child' and 'proportion time spent on computer/video' entered the model for the first time and showed substantial positive associations with overall happiness.

In order to explore the direct and indirect effects of the explanatory variables on overall happiness, we estimated a model that included the domain-specific happiness scores as control variables. The last two columns of Table 8.2 show the results. They generally confirm the equivalent relations between overall and domain-specific happiness pre-

sented in Figure 8.3, after controlling for personal characteristics and domain variables. A number of variables dropped from the model and most of the coefficients declined in magnitude, suggesting that variables clearly differ in terms of their direct and indirect relation to overall happiness, but the overall model remained fairly robust and the explained variance rose considerably to 42.5%. In order to better understand the indirect effect of the variables we performed three additional regressions, each controlling for happiness in a single life domain, and in each case inspected which variables dropped from the model presented in Table 8.2.

Considering variables related to the home situation, both 'parents are strict' and 'parents give compliments' dropped from the model after controlling for the happiness scores in all three domains. The regressions controlling for happiness in one life domain at a time, showed that these two variables also drop from the model when controlling only for happiness at home. Controlling for happiness at school had no statistical effect on the two variables 'parents are strict' and 'parents give compliments'. In the regression which controlled for happiness during leisure time only the variable 'parents are strict' dropped from the model.

Of the two school variables, 'teased at school' dropped from the model in both the regression which controlled for happiness in all three life domains and the one which only controlled for happiness at school. Controlling for happiness at home and happiness during leisure time had no effect on the school variable 'teased at school'.

Furthermore, after controlling for happiness in all three life domains simultaneously, the leisure time variables having friend, being alone and being bored dropped from the model. This was also true when controlling only for happiness during leisure time. Having friends dropped from the models when controlling for the other life domains separately as well, while being alone and being bored did not. The variables 'going home after school' and being 'member of a sports club' entered the model when controlling for the happiness scores in all life domains simultaneously and when controlling for happiness at home and at school separately (results not shown).

The standardized coefficients of the regression of overall happiness after controlling for all life domains at once indicate that health is a major variable associated with the overall happiness of adolescents, followed by three of the 'Big-Five' personality traits. A range of other variables have comparable but lower standardized coefficients, among them ethnicity, smoking, body image, monthly budget, school performance, strict upbringing, and proportion of time spent on computer/video.

8.5. DISCUSSION

The main purpose of this study was to explore the associations between domain-specific happiness and overall happiness within the context of a broad array of potential underlying determinants. The results show that adolescents were able to differentiate their happiness levels among the specific life domains from overall happiness. This finding is in line with previous studies. Gilman and Huebner (2003) and UNICEF (2007) already demonstrated that adolescents were able to distinguish their happiness levels to specific life domains. By including multiple life domains in our model we found that the adolescents in our study weigh their happiness in the three life domains approximately equally with respect to overall happiness. At this point, we can only speculate about possible explanations. In part, this may relate to the fact that home, school and leisure time are not isolated domains within space, time, and social relations and that we may thus expect some overflow of happiness between domains, blunting explicable differences in happiness between domains as well as their relative importance to overall happiness.

The average happiness levels in this study ranged from 7.68 to 8.59 on a scale of 0 to 10. Such high levels of happiness are consistent with other studies, both involving adolescents and adults [190, 206, 207].

Consistent with many previous studies, gender did not show a significant effect on overall happiness [191]. However, here we found that girls were happier at school while boys tended to be happier at home and during their leisure time. On the level of overall happiness, this effect appears to cancel out. The same seems to hold for the variable 'less wealthy'.

The importance of health in predicting happiness has regularly been reported in the literature as well [34, 187]. In accordance with these findings 'health status' showed a positive significant effect on happiness in all life domains and on overall happiness. This indicates that health is indeed strongly associated with overall happiness, directly and indirectly (i.e., through domain-specific happiness). The domain-specific effects of health, however, differed substantially. Differences in health appear to have a stronger effect on happiness at school than on happiness at home.

The domain variables were, as expected, most strongly associated with happiness in their respective life domains. However, some domain variables, like 'school performance', 'parents are interested', 'friends', and 'alone' showed significant relations with more than one domain. The importance of the relationship with parents and loneliness for overall happiness has been demonstrated in previous studies [195, 196], as has the positive association between the number of friends and happiness [208]. Interestingly, the num-

ber of friends does not relate to overall happiness (controlled for domain happiness) or happiness during leisure time. One explanation may be that it is important to have at least one friend, which is possibly captured in the variables 'alone' and 'bored', and that the marginal utility of additional friends diminishes sharply.

Some limitations of this study should be acknowledged. First of all, female gender, general secondary education and Dutch ethnicity were slightly over-represented in the sample. As a result, the average happiness levels of this study are not entirely representative for adolescents in the Netherlands, and we should be cautious with generalizing our results, although we found only low correlations of gender and ethnicity with happiness. Female gender was associated with slightly lower happiness at home and in leisure time, slightly higher happiness at school, but no difference in terms of overall happiness. Adolescents with Dutch ethnicity showed slightly higher happiness overall, but no difference between domains. All in all, this indicates that overall happiness possibly is slightly lower in the reference population than reported here. However, given the explorative nature of our analysis, we expect these minor deviations pose no significant limitations. The purpose of this study was to explore determinants of happiness and the relative importance of domain specific happiness to overall happiness, not to draw generalized conclusions about happiness of Dutch adolescents.

Furthermore, our analysis relies on secondary data, using a database with many interesting background variables for the purpose of our study. Given that only a few studies have focused on adolescent and/or multidimensional happiness, a completely theoretically driven selection of variables would have limited the explorative character of our study considerably. Therefore, we have extended the selection of variables based on earlier studies with a stepwise procedure based on a statistical selection of variables. However, the scope of this study was limited to the available life domains and explanatory variables included in the original database. Although the study included an extensive amount of potentially interesting and relevant determinants, other than the available variables will undoubtedly contribute to the jigsaw puzzle pieces of adolescent happiness.

It needs noting that we cannot fully rule out problems with endogeneity. Subjective measures of happiness and health or factors like body image may have all been influenced by unobserved factors. The results from the regression analyses could be biased by such influences to some extent. Moreover, given the cross sectional nature of the current dataset we cannot investigate causality of the reported relationships. E.g., while increased health may result in higher happiness, increased happiness may also result in more health. The reported relationships should therefore be interpreted as associations.

Finally, it has been suggested that happiness questions should be placed at the beginning of surveys to minimize order bias [209]. In the Health & Future questionnaire, how-

ever, the domain-specific happiness questions were placed at the end of each domain section so that adolescents would include all their responses to domain variables in their assessment of domain happiness. Arguably, the happiness scores may be 'biased' by the prior questions in the specific section.

8.6. CONCLUSION

Happiness is increasingly measured and analysed in economics. Especially when happiness is seen as a maximand for policy decisions, the determinants of happiness need to be studied.

This study has made some important steps forward in studying adolescent happiness. The results demonstrate that adolescents' happiness levels vary with life-specific domains, which supports the relevance of a multidimensional approach in happiness studies. Consequently, this study provides additional insight over single-dimensional studies of happiness and a more comprehensive explanation of previously published findings. Variables associated with adolescent happiness differ between life domains, but the life domains appear to contribute approximately equally to overall happiness. Furthermore, broad arrays of determinants are either directly associated with adolescent overall happiness or indirectly by passing through a specific domain happiness. Finally, this study once again shows that results are robust to both cardinal and ordinal interpretations of the happiness measure.

Future studies can contribute further to understanding adolescent happiness by exploring additional determinants, the interaction of happiness between the different life domains, and potential alternative domain definitions. Another interesting research objective may be to determine the stability of adolescents' domain-specific and overall happiness over time. In any case, it will be a happy puzzle ☺.

APPENDIX 8.1: ANALYSIS BASED ON ORDINALITY ASSUMPTION

Table A. 8.1: Wilcoxon signed rank test (n=1,436)

Variables	Z	Sign
Happiness- Happiness home	-15.306[a]	0.000
Happiness- Happiness school	-7.558[b]	0.000
Happiness- Happiness leisure time	-16.956[a]	0.000

Note: [a] based on negative ranks. [b] based on positive ranks.

Table A. 8.2: Results ordered logit regression† (n=1,436)

Variables†		Happiness Domains			Happiness Overall	
		Home	School	Leisure Time	Excluding domains	Including domains
		B˜	B˜	B˜	B˜	B˜
Core individual variables						
Gender	Female = 1, Male = 0	-0.239*	0.223*	-0.200	-0.077	-0.034
Ethnicity	Dutch = 1, Other = 0	-0.064	0.247	0.385*	0.387*	0.309
Education	Higher = 1, Lower = 0	-0.253*	-0.122	0.071	0.094	0.172
Less wealthy	Yes = 1, No = 0	-0.226	0.480*	-0.153	0.059	0.057
Health	0-10 Scale (worst conceivable = 0, best conceivable = 10)	0.294**	0.381**	0.313**	0.607**	0.447**
Agreeableness	Continuous variable (higher score: more agreeable)	0.300**	0.450**	0.302**	0.446**	0.284**
Conscientiousness	Continuous variable (higher score: more conscientious)	0.054	-0.005	-0.065	0.039	0.057
Openness to experience	Continuous variable (higher score: more open to experience)	0.101*	0.106*	0.034	0.088	0.021
Extraversion	Continuous variable (higher score: more introverted)	-0.060	-0.152*	-0.153*	-0.339**	-0.333**
Neuroticism	Continuous variable (higher score: more neurotic)	-0.070	-0.150*	0.052	-0.176*	-0.170*
Happiness domains						
Happiness Home	0-10 Scale (very unhappy = 0, very happy = 10)					0.415**
Happiness School	0-10 Scale (very unhappy = 0, very happy = 10)					0.375**
Happiness Leisure time	0-10 Scale (very unhappy = 0, very happy = 10)					0.329**
Additional individual variables						
Smoking	Yes = 1, No = 0		-0.159		-0.538*	-0.487*
Overweight	Yes = 1, No = 0		0.381*			
Too thick	Yes = 1, No = 0				-0.178	-0.199
Religious upbringing	Yes = 1, No = 0		0.293*	-0.193		
Life expectancy	In years	0.005	0.009*			

Table A. 8.2: Results ordered logit regression[†] (n=1,436) (continued)

Variables[†]		Happiness Domains			Happiness Overall	
		Home	School	Leisure Time	Excluding domains	Including domains
		B˜	B˜	B˜	B˜	B˜
Meaning of future life in 5 years	(very) important = 1, (very) unimportant = 0		0.318	-0.487*		
Monthly budget	€				0.002*	0.002*
School variables						
School performance	(Very) unsatisfactory = 0, Satisfactory=1, Good=2, Very good = 3	0.343**	0.365**		0.315**	0.194*
Teased at school	Never = 0, Sometimes/ Often = 1		-0.659**		-0.390*	
Home variables						
Satisfied with relationship mother	Totally disagree = 0, Disagree = 1, Agree = 2, Totally agree = 3			-0.132	-0.174*	-0.206*
Only Child	Yes = 1, No = 0				0.425*	0.495*
Parents are strict	Yes = 1, No = 0	-0.781**			-0.211	
Parents are interested	Yes = 1, No = 0	0.500*	0.582*			
Parents give compliments	Yes = 1, No = 0	1.102**		0.416*	0.419*	
Single Parent	Yes = 1, No = 0	-0.464*				
Leisure time variables						
Going home after school	Yes = 1, No = 0	0.566*				-0.341
Friends	Less than three= 0, Three or more = 1	0.188	0.540**		0.309*	
Proportion time homework				-2.705**		
Proportion time computer/ video					0.991*	0.870*
Member of sports club	Yes = 1, No = 0		-0.238*			0.252*
Alone	Never= 0, Rarely=1, Sometimes = 2, (Very) Often =3	-0.383**	-0.241**	-0.423**	-0.259**	
Bored	Never= 0, Rarely=1, Sometimes = 2, (Very) Often =3	-0.208*		-0.763**	-0.200*	

Note: [†] the variables: chronic condition, employment status, obesity, meaning of future life in 2 and 25 years, diner at family table, parents home after school, team sport, proportion time spent on serious activities and proportion time spent on social activities, of Figure 8.2 were excluded as a result of the stepwise hierarchical regression analyses; * $p<0.05$, ** $p<0.001$.

9

How important is health for happiness?

Based on:
Van de Wetering, E. J.*, de Vries E.F*, Tsiachristas A., Van Exel, N. J. A., (2015). How important is health for happiness? *Submitted*

*Shared first authorship

ABSTRACT

Health care policy makers increasingly recognize the additional value of happiness, or wellbeing, as a broader outcome measure to inform health care allocation decisions, since health may not capture all effects of an intervention. However, the relationship between health and happiness is complex. Literature shows that health and happiness are strongly correlated, especially when they are both assessed by subjective measures. As a result, unobserved factors that are omitted from the model or reverse causality may bias the estimated effect of health on happiness. This study explores the causal effect of health on happiness in more detail using the SHARE database. OLS regressions and instrumental variable analyses were conducted to consider the possible impact of endogeneity. The results showed a causal effect of health on happiness and confirmed the additional value of instrumental variable analysis. Researches and policy makers should be aware of endogeneity bias in the relationship between health and happiness, since otherwise they might underestimate the effect of health on overall wellbeing.

9.1. INTRODUCTION

It is generally recognized that happiness, or wellbeing, can be viewed as an ultimate goal in life [203]. Therefore, happiness is emerging as an important societal aspiration. It is increasingly argued that happiness can be considered as proxy for utility or a (partial) index of wellbeing and thus that social policy making may (partly) be guided and evaluated by happiness measures [210]. Happiness is a subjective measure of the overall enjoyment of life, commonly defined as 'the degree to which an individual judges the overall quality of his life favourably' [31]. The concept of happiness extends back to the work of the utilitarian philosopher Jeremy Bentham. The essential argument is his reasoning was the pursuit of 'the greatest happiness for the greatest number' [178]. However, happiness has a controversial tradition in economics due to the subjective nature of its content and its measurability. For a long period of time, economists therefore avoided the use of happiness as an objective outcome measure and proxy of wellbeing. In recent years, following disciplines like sociology and psychology, economists appear to have (re)discovered happiness as a significant concept [12, 37, 173, 177, 211]. Happiness questions are now included in large national and international surveys and countries like Canada, France and the UK are seriously considering happiness as an outcome measure for public policy [37].

Also health care policy makers increasingly embrace the idea that happiness measures may provide relevant information to inform and evaluate health care policy decisions. While health care is still primary aimed to optimize health, it is increasingly recognized that the narrow focus on health (or *health-related* quality of life) as the main outcome measure may not always capture the full benefits of health care interventions. For example in elderly care, interventions may not be primary aimed to improve health, but (also) to contribute to other aspects of wellbeing, such as autonomy. In these cases, broader outcome measures such as happiness may be more in accordance with the goals of the intervention [10, 28]. One may even claim that from a wider societal perspective, using only health as an outcome measure to inform allocation decisions may result in suboptimal allocation decisions because it does not capture all benefits of an intervention [10]. In other words, an outcome measure for health care in terms of overall wellbeing improves the transparency about the contribution of health care to societal progress. Especially in times of ageing populations and growing health care expenditures, framing the benefits of health care in a more comprehensive way, such as happiness allegedly does, clearly shows to what extent health contributes to overall wellbeing and thus underlines the relevance of health care [27].

Given the increasing interest in happiness as an outcome measure in health care it is important to improve the understanding of the relationship between health and hap-

piness. It should be noted that any exploration of the relationship between health and happiness needs to consider the alternative approaches to measuring health. Health status can be measured in objective terms, as assessed by medical personnel or in terms of use of medical services, or in subjective terms, as an evaluation by the individuals themselves [212]. Most studies so far have used subjective measures of health to explore the association between health and happiness, but such self-reported measures of health have been shown to deviate from objective measures [37, 212]. Considering the subjective nature of both health and happiness, this complicates the identification of a causal effect of health on happiness. It is likely that endogeneity arises when happiness also has a causal impact on health (i.e. reverse causality) or when health and happiness are both correlated with important unobserved factors that are omitted from the data, such as personality [36, 37]. An individual with a relatively low overall wellbeing may, for instance, experience her health worse than her true health status. This type of under-reporting of the 'true' health status leads to biased results and biased estimations of the relationship between health status and happiness.

A substantial body of empirical research has examined the association between health and happiness. Since happiness is generally considered to be a broader concept than health, the majority of studies have focused on the direction from health to happiness and found that individuals with better subjective health also report higher levels of happiness [36, 37, 210]. While these studies clearly show a statistical association between health and happiness, they do no inform us about cause and effects. The association can be caused by the effect of health on happiness, but also by a simultaneous relationship between health and happiness or a spurious relationship with a third factor influencing both health and happiness [36].

A recent study of Garrido et al. [37] showed a simultaneous relationship between subjective health and happiness, in which unobserved variables that determine health and happiness were significantly related. They argued that a unidirectional approach using standard OLS measures may therefore underestimate the relation between health and happiness. However, their study was based on national data and limited to single measures for health (i.e., SF-6D) and happiness (i.e., Life satisfaction scale) [37]. Therefore, the aim of this study was to contribute to the existing literature by further exploring the causal effect of health on happiness and considering the impact of possible endogeneity. To achieve this aim, we used a large international data set including several health and happiness measures and a broad set of potential instrumental variables, performed thorough statistical analysis, and conducted sensitivity analysis to address the uncertainty and test the robustness of the instrumental variables.

9.2. METHODS

9.2.1. Data

This study uses data from the Survey of Health, Ageing and Retirement in Europe (SHARE), a large database of individuals over the age of 50 years from 19 European countries. For this cross-sectional study, we used data from the fourth wave, collected in 2010, which contained data of 58.000 respondents [213]. The dataset includes a large variety of variables potentially relevant for this study.

9.2.2. Specification of the models

Happiness data can be interpreted to be cardinal, which makes OLS an acceptable method for the analysis [214]. Based on an OLS regression, the relationship between health and happiness can be expressed by the following equation:

Eq 9.1 $$Y = \alpha + \beta_1 X_1 + \beta_i \pi_i + \varepsilon$$

where Y is the dependent variable happiness, α is the intercept, X_1 is the variable health, π_i represent other exogenous variables and ε is the error term. However, OLS regressions only estimate the correlation between health and happiness and may suffer from the endogeneity problem. Therefore, the relationship between health and happiness was also explored by instrumental variable analysis, estimated with two stage least square (2SLS). Instrumental variable analysis is aimed to find variables that can be used to correct for possible bias caused by the endogeneity of the health variable. In the first stage of 2SLS the endogenous variable health ($\hat{X}$) is predicted by its instrumental variable (Z_i):

Eq 9.2 $$\hat{X} = a + bZ_i + b\pi_i + u$$

where a represents the intercept and u the error term. In the second stage the initial Equation 9.1 is estimated with the predicted variable of health as a proxy for the original health variable:

Eq 9.3 $$Y = a + \beta_1 \hat{X}_1 + \beta_i \pi_i + \varepsilon$$

Finding a credible instrumental variable for 2SLS regressions, i.e. one fulfilling the requirements of validity and relevance, is considered the most important and difficult step in instrumental variable analysis [215, 216].

9.2.3. Selection of instrumental variables

Validity

The most important requirement for an instrumental variable is to be valid. This implies that the instrumental variable should only be related to the dependent variable through the endogenous variable, health in this case, and should not be correlated with the error-term [215]. A literature search was conducted to gather theoretical evidence from preceding studies to judge whether the candidate instruments fulfil this requirement; Garrido et al. [37] and Böckerman et al. [217] were important sources of information. In addition, definitions of the potential instruments in combination with keywords of health and happiness were used as a search strategy. Based on the theoretical evidence, the candidate instrumental variables were categorized as 'valid', 'questionable' or 'not valid'.

Relevance

Relevance is the second important requirement for an instrumental variable. This requirement implies that, after controlling for confounders, to be informative the instrumental variable should account for a significant variation of its endogenous variable [215, 216]. The stronger the association between the instrumental variable and the endogenous variable, the stronger the identification of the model [215]. Three statistical tests were performed to evaluate the relation between the candidate instruments and health. First, Stock and Yogo tests for 'weak' instruments were performed to test possible bias of the 2SLS estimates relative to the bias of the OLS estimates due to a weakly correlated instrumental variable. Since there is no consensus on which test statistic is most appropriate in this context, both the F-statistic and the minimum eigenvalue statistic were evaluated against the critical value of 16.38 [215]. This value was determined based on Stock and Yogo's guidance [218]. Secondly, Pearson's correlations between the potential instrumental variables and health were estimated. Thirdly, the partial R^2 indicating the marginal contribution of the instrument in explaining the endogenous variable after controlling for confounders, was examined. First stage regression results were reported to examine the statistical significance of the correlation. It should be noted that there is no consensus in the literature on the minimum value of Pearson's correlation and partial R^2. In this study, Pearson's correlations below 0.10 and partial R^2 of less than 0.01 were considered as too low. Pearson correlations below 0.10 may indicate efficiency loss when using 2SLS analysis opposed to OLS. A partial R^2 of less 0.01 indicates a relatively small contribution of the instrument in explaining health, after controlling for confounders [215]. Overall, a candidate instrument was considered 'strong' when the results of all statistical test exceeded their critical value. Preceding studies have shown that instruments that are categorized as 'questionable' in terms of their validity imply only small

biases when the instrument's relevance is strong [219]. Therefore, all instruments that were 'valid' or 'questionable' were tested on their relevance.

9.2.4. Additional tests

Wu-Hausman tests for endogeneity were performed to statistically confirm the additional value of 2SLS compared to OLS. A non-significant difference between the health coefficients of the IV analysis and OLS may indicate that the endogenous variable is exogenous in reality or that the instruments are problematic [215].

Sensitivity analyses were conducted to test the robustness of the results with respect to the selected outcome variable for happiness in the main analysis, i.e. CASP-100 (see below). The OLS and 2SLS regression analyses were remodelled with the LS-100 variable (see below) as outcome variable for happiness.

All analyses were conducted using Stata version SE 13.0.

9.2.5. Measures of interest

Happiness

The SHARE data includes two subjective measures of happiness, the life satisfaction scale and the CASP-12. The life satisfaction scale (LS) is widely used as a valid measure for happiness [37, 177, 180, 220]. This measure quantifies happiness by simply asking: 'on a scale from 0 to 10 how satisfied are you with your life?' where 0 means completely dissatisfied and 10 means completely satisfied.

The CASP-12 is a validated measure of happiness in older people, and a reduced form of the CASP-19 [221]. The CASP-12 captures four dimensions of well-being: control, autonomy, self-realization and pleasure. Each dimension is measured by three items on which respondents have to indicate the extent to which the item applies to them on a four point scale with response categories 'often', 'sometimes', 'rarely, 'never' [222].

It can be argued that the CASP is methodologically more objective than the LS scale. The CASP-12 has a clear structure of four dimensions and uses a four point scale to value the outcome, thus defining the concept of 'happiness' quite specifically. The LS scale uses a single overall score for satisfaction with life in general, without specifying what the respondent should take into consideration when answering the question. The CASP-12 thus give less room to different (subjective) interpretations of happiness than LS-100. For that reason, the CASP-12 was selected for the main analysis.

In this study the CASP-12 and the LS were linearly rescaled to a range of 0 to 100 (from worst to best) and labelled as CASP-100 and LS-100, respectively.

Health

SHARE contains two subjective health measures: the health thermometer and the EQ-5D. The health thermometer measures how the current health state is perceived on a Visual Analogue Scale (VAS) from 0 (the worst imaginable health state) to 100 (best imaginable health state) by asking: 'We would like you to indicate on this scale how good or bad your health is today, in your opinion'.

The EQ-5D is a widely accepted measure of health-related quality of life and captures the domains mobility, self-care, usual activities, pain/discomfort and anxiety/depression on three levels, corresponding with no, some or severe problems on that domain [223, 224]. A sum score or 'misery index' can be computed by assigning values of 0 to 2 to each level and summing these scores across domains, generating an overall sum score ranging from 0 (full health) to 10 (worst health). This sum score of the EQ-5D was included for the main analysis since it is a widely accepted measure of health and, for the same reasons as CASP-12 was preferred to LS for happiness, assumed to be more objective than the health thermometer [224]. The EQ-5D was linearly rescaled to a range of 0 to 100 (from worst to best) and labelled as EQ5D-MI-100.

Candidate instrumental variables for health

Besides these two health measures the SHARE dataset includes an extensive list of self-reported medical conditions that respondents can select in response to the question 'Has a doctor ever told you that you had any of the conditions on this card?' Previous studies have shown that self-reports of conditions that require laboratory or medical diagnostic procedures prove to be reliable observations [225]. Therefore, these conditions were used as candidate instruments for health. Medical conditions with a prevalence rate of at least 5% in the SHARE dataset were selected for this study. This minimum prevalence rate was set to ensure sufficient observations for the instrumental analysis. Further analyses should prove the validity and relevance of these selected variables as credible instruments for the main analysis.

Covariates

To control for personal characteristics the following variables were included in this study. First, age of the respondents, since the association between age and happiness is well documented. Most studies find measures of happiness to be U-shaped related to age, where both younger and older persons report being happier than middle-aged persons [203, 226]. Therefore, age squared divided by 100 was included to allow for non-linearity of age. Secondly, as proxy for socio-economic status we use both level of education and income. Education is found to be positively correlated with happiness [226]. SHARE contains the International Standard Classification of Educational Degrees (ISCED) variable which was recoded for this study into dummy variables indicating low

(ISCED 0 to 2), middle (ISCED 3 and 4) or high education (ISCED 5 and 6). For income, SHARE contains a variable that measures the extent to which a household is able to make ends meet. Various studies have argued that being able to provide in basic necessities is more important to health and happiness than the absolute level of income [29, 173, 227]. Finally, place of residence is considered to influence happiness [37]. In this study, a dummy variable was included indicating rural place of residence (rural area or village) or urban place of residence (big city, suburbs or town).

9.3. RESULTS

9.3.1. Descriptive statistics of the sample

Descriptive statistics of the study sample and the complete SHARE 2010 wave are reported in Table 9.1. Respondents with missing values for one of the outcome variables were excluded from the original dataset. This resulted in a study sample of 41,936 respondents with a mean age 68.6 years. Except for the level of education, all mean values of the descriptive variables differed significantly from the original SHARE dataset.

Table 9.1: Study Sample (n=41,936) and SHARE (n=58,000) characteristics

		Study sample Mean (SD)	**SHARE Mean (SD)**
Happiness	CASP100	70.6 (17.6)*	69.5 (17.9)
	LS100	77.2 (17.6)*	75.6 (18.7)
Health	EQ5D-MI-100	13.9 (17.1)*	14.1 (17.2)
	HT-100	71.5 (20.0)*	70.7 (20.7)
Age		68.6 (9.9)*	68.9 (10.1)
Educational level[a]	Low	38.6	38.4
	Middle	40.9	41.2
	High	20.5	20.4
Making ends meet[a]	Great problems	11.3*	12.0
	Some problems	28.2*	29.3
	Fairly easy	32.3	32.5
	Easy	28.3*	26.2
Place of residence[a]	Urban	64.4*	65.9
	Rural	35.6*	34.1

Note: To improve the comparability between the different instruments in this study the happiness and health variables are rescaled and linearly distributed in a range of 0 – 100, and labelled as LS-100, CASP-100, and HT-100, EQ5D-MI-100. [a] Dummy variable (%) * Statistical significance difference (P<0.05) between the study sample and SHARE.

9.3.2. Instrumental variable analysis

Candidate instruments

Table 9.2 provides an overview of the health conditions in the SHARE dataset that were selected as candidate instruments for health. Of the original 17 variables under consideration, Alzheimer's disease dementia or senility, asthma, benign tumour, hip or femoral fracture, osteoporosis, Parkinson, stomach or duodenal or peptic ulcer, stroke and other fractures were excluded because their prevalence rate was less than 5%.

Table 9.2: Validity of the candidate instrumental variables

Candidate instruments	Prevalence rate (%)	Validity
Cancer	5.1	Not valid
Cataracts	8.0	Not valid
Arthritis	23.8	Not valid
Chronic lung disease	6.2	Not valid
Heart attack	12.0	Not valid
Diabetes or high blood sugar	12.3	Valid
High blood cholesterol	23.5	Valid
High blood pressure or hypertension	38.2	Questionable

Validity

The results of the validity test for the candidate instruments are reported in the last column of Table 9.2. First, the definition of the variable cancer was considered as too non-specific to judge its validity and was therefore categorized as 'not valid'. Secondly, a number of medical conditions was categorized as 'not valid' because they lead to functional impairment, and therewith can be expected to affect wellbeing directly. Cataract directly influences the routines of daily life and places constraints on social contact and participation, which are important domains of overall wellbeing [210, 228, 229]. Previous studies reported significant effects of arthritis and chronic lung disease on social relationships, employment and independency [210, 230]. According to literature, an acute life threatening event like a heart attack has permanent negative consequences: physical limitations such as angina and shortness of breath not only place restrictions on daily activities but also have psychological consequences as anxiety and distress [231].

Finally, diabetes or high blood sugar, high blood pressure or hypertension and high blood cholesterol, without complications, are all silent medical conditions that are less likely to directly influence functional status. No additional significant effects of the variables diabetes or high blood sugar [217] and high blood cholesterol diabetes [37] on wellbeing were found when corrected for health. Therefore, these variables were categorized as 'valid'. The causal relationship between hypertension and wellbeing is

somewhat controversial since psychosocial factors may play a role in hypertension. Many mixed findings are reported in the literature but the exact association is still unclear [37, 217, 232, 233]. Therefore, we categorized the validity of this variable as 'questionable'.

Relevance

The results of the statistical tests for the relevance are presented in Table 9.3. For all candidates, the null hypothesis of 'weak instruments' was rejected since the results of the Stock & Yogo tests exceeded the critical value of 16.38. Diabetes or high blood sugar, high blood pressure or hypertension and high blood cholesterol were all sufficiently correlated with health, showing Pearson correlations above the critical value of 0.10. The partial R^2 scores indicate a considerable contribution of the instruments to explaining health, after controlling for confounders. Furthermore, the coefficients of all three candidate instruments were significant in the first stage regression analysis. Overall, the relevance of the three candidate instruments can thus be considered as 'strong'.

Table 9.3: Relevance of the valid candidate instruments

Instrument	Overall relevance	Stock & Yogo	F-statistic	Pearson Correlation	β first stage	Partial R^2
High blood pressure or hypertension	Strong	265.322	244.951	0.18	4.22***	0.0157
High blood cholesterol	Strong	178.797	162.943	0.11	3.89***	0.0106
Diabetes or high blood sugar	Strong	278.765	192.153	0.18	6.29***	0.0165

Note: ***statistically significant at the 1% level.

9.3.3. Regression results

Diabetes or high blood sugar, high blood pressure or hypertension and high blood cholesterol proved to be credible instruments for health. Therefore, these instruments were included in the OLS and 2SLS regressions with CASP-100 as measure for happiness. The results of these analyses are presented in the first four columns of Table 9.4. The Wu-Hausman tests showed statistical significant differences between OLS and 2SLS estimates for all instruments which confirms the additional value of 2SLS compared to OLS.

The OLS estimates show a significant positive association between health and happiness. This indicates that healthier people are also happier. For *age*, a reversed U-shaped relationship with happiness was found. Happiness increased with age from 50 years until the maximum at 63 years of age and then decreased. Furthermore, middle and higher educated respondents reported to be significantly happier than lower educated respondents. Living in a rural area and being able to make ends meet were also significantly and positively related to happiness.

Table 9.4: Results from OLS and 2SLS regressions

	CASP-100				LS-100			
		2SLS				2SLS		
	OLS	High blood pressure or hypertension	High blood cholesterol	Diabetes or high blood sugar	OLS	High blood pressure or hypertension	High blood cholesterol	Diabetes or high blood sugar
EQ5D-MI-100[†]	-0.42***	-0.56***	-0.83***	-0.68***	-0.32***	-0.57***	-0.60***	-0.48***
Age	0.45***	0.21	-0.26	0.01	-0.35**	-0.77***	-0.82***	-0.62***
$Age^2/100$	-0.36***	-0.14	0.28*	0.05	0.33***	0.72***	0.76***	0.58***
Education								
Middle	3.90***	3.48***	2.69***	3.14***	1.50***	0.78**	0.70*	1.04***
High	3.14***	2.61***	1.59***	2.17***	1.87***	0.93**	0.82*	1.27***
Place of residence								
Rural	3.24***	3.39***	3.69***	3.52***	0.95***	1.24***	1.27***	1.13***
Being able to make ends meet								
Some problems	5.20***	4.31***	2.67***	3.59***	9.15***	7.65***	7.48***	8.19***
Fairly easy	11.71***	10.17***	7.30***	8.91***	15.69***	13.05***	12.76***	14.01***
Easy	16.76***	14.82***	11.20***	13.24***	19.01***	15.67***	15.3***	16.87***
Wu-Hausman test		0.008	0.000	0.000		0.000	0.000	0.014

Note: [†] EQ-5D misery index ranges from 0 (best health state) to 100 (worst health state). ***statistically significant at the 1% level. **statistically significant at the 5% level. *statistically significant at the 10% level.

The magnitude of the health coefficients of all three instruments was larger in 2SLS than in OLS regressions. This suggests stronger positive associations between health and happiness than indicated by the OLS regression. The high blood pressure or hypertension model showed an increase of nearly 30%. The strongest association between health and happiness was found in the model based on the instrument high blood cholesterol. Age and happiness were not significantly related in the 2SLS regressions except for the positive effect of age square in the high blood cholesterol model. The changes in coefficients of the other confounders were small.

9.3.4. Sensitivity analysis

The last four columns of Table 9.4 show the results of the sensitivity analysis with the life satisfaction scale (LS-100) as measure for happiness. No noteworthy differences were found with respect to the relationship between health and happiness. Also in these models, the association between health and happiness appeared to be stronger using 2SLS instead of OLS and the strongest association was found in the high blood cholesterol model. However, the associations between age and happiness reversed: happiness initially decreased with age up to 53 or 54 years of age, depending on the model, and then increased.

9.4. DISCUSSION

The increasing interest among researchers and policy makers in the health care sector for the use of broader outcome measures, such as happiness, to inform allocation decisions, raise the importance to improve our understanding of the relationship between health and happiness. The aim of this study was to explore the causal effect of health on happiness in more detail. The instrumental variable (IV) approach was used to disentangle the effect of possible endogeneity of the subjective health variable. The results showed that conventional OLS regressions, which ignore endogeneity, may underestimate the effect of health on happiness.

In explaining happiness, high blood pressure or hypertension, high blood cholesterol and diabetes or high blood sugar turned out to be credible instruments for health. High blood cholesterol and diabetes or high blood sugar were both valid and relevant, whereas the validity of high blood pressure or hypertension was considered questionable given the discussion in the literature with respect to reverse causality. However, previous studies have shown that instruments that are relevant and of questionable validity imply only small biases for instrumental variables [219]. The same seems to

apply in this study, since the results of the OLS and 2SLS for high blood pressure or hypertension were comparable to the other instruments.

The 2SLS regressions of all three instruments independently showed significantly stronger positive associations between health and happiness than the OLS regressions. The strongest association was found in the high blood cholesterol model. These findings indicate that subjective health is endogenous and may bias the results of OLS regressions. The fact that all instruments showed consistent results, also in the sensitivity analyses, was encouraging. The causal effect of health on happiness was somewhat smaller when using the more subjective measure LS for happiness instead of the more objective measure CASP. It might be that LS encompasses a broader scope of aspects influencing happiness, therewith reducing the effect of health.

With respect to the covariates, the relationship between age and happiness warrants some further discussion. The OLS regression with CASP as measure for happiness showed a significant inverted U-shaped relationship between age and happiness. Given the age group of the sample, it seems that we have observed the increasing slope after the lowest point of the parabola [37]. However, this association was not significant in the corresponding 2SLS models. An explanation might be that the instruments were correlated with age and thus picked up a part of the effect of age. Remarkably, the OLS and 2SLS regressions with LS as measure for happiness showed the more frequently observed U-shaped association between age and happiness, where happiness decreased from 50 until 53 or 54 years of age, depending on the model, and then increased. Apparently, while the more objective measure of happiness (CASP), capturing the domains control, autonomy, self-realization and pleasure, identified lower levels of happiness after a certain age, the more subjective measure (LS) indicates that these people themselves experienced higher levels of overall happiness.

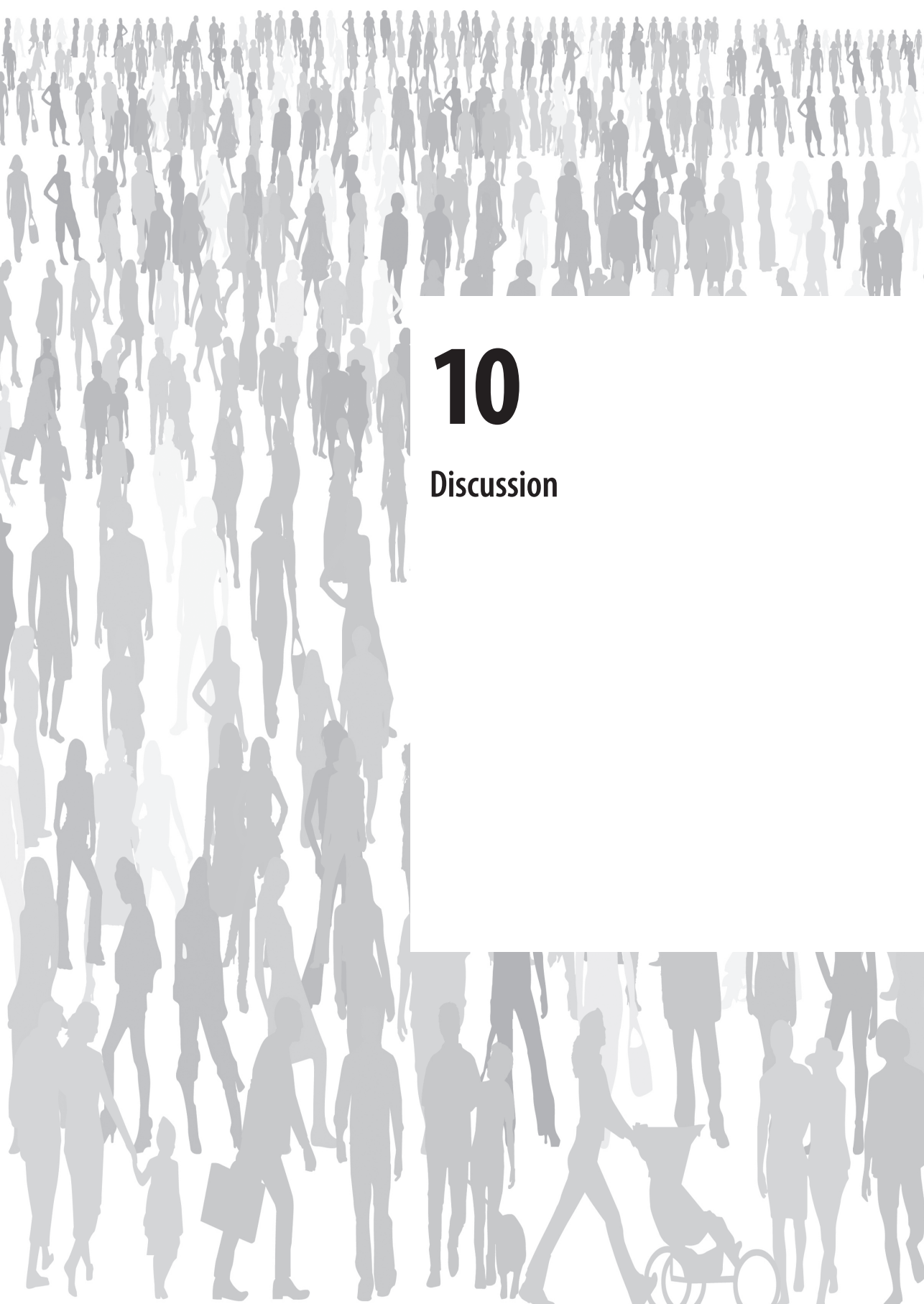
While empirical studies frequently demonstrate that the association between objective health and happiness is weaker than the association between subjective health and happiness [210, 234], this study seems to find the opposite. However, the results of our study are consistent with those of Garrido et al. [37], which also explored the simultaneous relationship between health and happiness. That study found that OLS - not accounting for endogeneity bias - may underestimate the effect of health on happiness. Further research is required to understand this effect, using different instruments for health and measures of happiness, in different respondent samples and contexts. Our study contributes to this literature by using a large, rich, and international dataset including two measures of happiness and a broad set of candidate instruments for the endogenous variable health. In this study different outcome measures for health and happiness were used for the main analysis, sensitivity analysis and robustness checks. It should be noted

that the identification of credible instrumental variables is not straightforward and has been topic of methodological debate [235]. Therefore, the robustness of the results was an important aspect of our methodological approach. We performed a literature search to assess the validity of the candidate instruments, and conducted a larger variety of statistical tests to test their relevance. Furthermore, to test the robustness of the selection process, the whole analysis was replicated with the health thermometer VAS as the endogenous variable for health. This resulted in a similar selection of instruments, with the exception that the partial R^2 of the high blood cholesterol model was just below the critical value of 0.01. This test increased the confidence in the credibility of the selected instruments for the regression analyses, and the robustness of the results. Still, development of more standardized guidance to test potentially relevant instruments on their validity and relevance is required.

This study has several limitations that need to be discussed. First, the SHARE data is representative for the European population aged over 50 years. However, due to many missing observations on important variables for this study, the final study sample used for analysis was no longer representative for the intended population, and the findings presented here thus lack external validity. However, given the purpose of this study, which was to explore the relationship between health and happiness, this is not considered problematic. Secondly, the study was based on secondary data, which potentially limits the availability of candidate instruments. Although we found strong correlations between the selected instruments and the endogenous variable health, it could be argued that a part of objective health was not captured by these instruments and, consequently, their effect on happiness is not represented in our analyses. Therefore, despite the robustness of our findings, we may not have fully eliminated the endogeneity bias. More objective measures of overall health are necessary in future research, and preferably health indicators that are not self-reported. Finally, this study is based on cross-sectional data, which was sufficient for the exploration of possible endogeneity bias in the relationship between health and happiness. Still, the interpretation of causality dictates extreme caution. For future research, a longitudinal panel approach is recommended to correct for unobserved time invariant factors and further investigate the causality between health and happiness [32, 37].

Notwithstanding these limitations, the results of this study showed a causal effect of health on happiness while considering the impact of endogeneity. These results are relevant for health care policy makers but also for policy making more generally. Policy makers are increasingly aware of the relevance of happiness as maximand for social policy. Taking well-being rather than health (or health-related quality of life) as measure of outcome brings the effects of interventions aimed at cure and those aimed at care under a common denominator, which may help to improve decisions about the allocation

of available resources within the health care sector between cure and care. But it can also improve a more efficient allocation of resources between sectors of the economy, based on their relative contributions to societal wellbeing. While most studies exploring the relationship between health and happiness use subjective measures of health, this study showed there is a causal relationship between health and happiness, and that this relation may be underestimated when not accounting for endogeneity bias.



10

Discussion

Economic evaluations are increasingly used to inform health care allocation decisions. However, there appears to be a discrepancy between the outcomes of economic evaluations and the socially desired allocation of health care resources. To reconcile this discrepancy and to make economic evaluations more useful for health care allocation decisions, there are still some important methodological challenges to address. This thesis investigated two of these methodological issues, related to measuring and valuing benefits of interventions. Part A investigated distributional concerns regarding health and health care. Applying equity weights to QALY gains might better align economic evaluations with societal preferences for a fair distribution of health and health care. Part B focused on happiness as a broader outcome measure for economic evaluations with the aim of capturing benefits of interventions beyond those captured in the QALY.

This final chapter discusses the main findings of Part A and Part B, including the limitations and research implications of this thesis, and addresses the overall conclusions and policy implications.

10.1. PART A: DISTRIBUTIONAL CONCERNS IN SOCIETY

To investigate the distributional concerns in society and their potential role in economic evaluations, the following research questions were formulated:

a) How is the equity efficiency trade-off operationalized in the decision making framework in the Netherlands?
b) What is the effect of providing information about severity of illness and fair innings on allocation decisions?
c) What is the social WTP for a QALY gain at different levels of proportional shortfall and in different age groups?
d) Does the social WTP per QALY depend on severity of illness and age characteristics of beneficiaries?
e) How does the current reimbursement status of an intervention affect the relative social value of a QALY?

Chapter 2 addressed research question *a* and evaluated how efficiency and equity considerations are combined in the Dutch decision making framework for health care allocation decisions. The Netherlands has made a first attempt to operationalize equity by using the concept of proportional shortfall. Although this provides important information to decision makers by quantifying the necessity of treatment, proportional shortfall does not always adequately capture societal distributional preferences. For example, the fact that avoiding a full loss of all remaining health would be equally important when it concerns either a very large or small *absolute* QALY loss seems hard to defend.

However, this and other shortcomings of proportional shortfall (highlighted in chapter 2) should not be misinterpreted as a plea to replace it with a different equity concept. It seems likely that whatever principle is chosen, similar problems will arise. The Dutch experiences with the operationalization of necessity provide important lessons for other countries and emphasize the need to further refine the operationalization of equity and its use in practice. In that context, more insight into the distributional preferences in society is important.

The empirical results of chapters 3 to 5 (answering research questions *b* to *d*) showed distinct preferences for different operationalizations of severity of illness as an equity concern for the allocation of health care resources. In chapter 3, support was found for both severity of illness (operationalized as proportional shortfall) and fair innings (operationalized as age). The results indicated that, while the severity of illness and fair innings approaches may have been contrasted in the literature [18], the general public showed some support for both principles. Providing more information to the choice scenarios affected the allocation decisions as well as the relative preferences for the two equity principles. Chapters 4 and 5 took the approach one step further by investigating the monetary valuation of a QALY in relation to these equity principles. Chapter 4 presented the estimated WTP per QALY in different age groups and found no support for the fair innings argument. We did find support for considering severity of illness among a substantial minority of the public. However, since the interaction terms between health state without treatment and QALY gains were not significant, we could not conclude that the value of a QALY significantly differed for different levels of severity of illness (operationalized as proportional shortfall). In chapter 5, a contingent valuation (CV) approach was used to estimate the WTP per QALY at different levels of severity, here operationalized as proportional shortfall and end-of-life concerns, in different age groups. Support was found for the end-of-life premium but no strong evidence was found for differentiating QALY gains according to proportional shortfall. Stronger preferences were found for the size of the QALY gain, indicating support for efficiency arguments, and for ageism, supporting the fair innings principle.

The diverse results between the chapters need not surprise since they are consistent with previous findings in the literature. Differences across studies seem to be the result of the framing of the decision, the operationalization of the equity concerns and the different methods used [22]. Notably, the inclusion of a cost component in chapters 4 and 5 might be problematic since it has been argued that respondents are not used to direct payment or monetary valuation of health care resources [236]. For that reason, in this thesis increases in monthly health insurance premiums were used to frame the cost component. We cannot conclude that the inclusion of a cost component caused the differences in preferences for severity of illness and fair innings between chapter 3 and

the subsequent two chapters, since there were more differences between these studies. For instance, in chapter 3, severity was defined in terms of losses in quality of life, while chapters 4 and 5 included QALY losses as a combination of length and quality of life.

It is interesting to note that the results of all empirical studies showed indications that the public considered the 'end point after treatment' in their decisions about which patient group to prioritize. This suggests that respondents attach weight to the desirability of the health state after treatment, with interventions which do not bring people back to desirable end states receiving less weight. These preferences are irrespective of whether this was a consequence of the health state before treatment or the health gain from treatment. The findings are consistent with previous studies [22, 237, 238].

In chapter 4, WTP per QALY values ranged from €206,408 for patients aged 10 years old to €296,756 in patients aged 40 years old. In the CV study presented in chapter 5, undiscounted WTP values were somewhat lower, ranging from €94,700 per QALY for 10 years old to €151,000 for 40 years old. A recent review by Ryen and Svensson [143] reported a wide range of WTP estimates of less than €1,000 to €4,800,000 euro per QALY, depending on the methodology, country, perspective taken and sample population of the study. The WTP per QALY values reported in this thesis should be interpreted within the explorative context of the presented studies; they were primarily designed to enable a better understanding of how the general public differentiates QALY values for different groups of patients.

With respect to research question *e*, the results of the DCE described in chapter 6 showed that the current reimbursement status of an intervention affects the allocation decision. Loss aversion may be a logical explanation of this finding: ending reimbursement is perceived differently from not starting reimbursement in the first place. Given an average scenario with an ICER of €50,000, the general public was willing to accept a €7,360 higher ICER for existing treatments. For policy makers this figure was €7,959. The latter finding could be explained in two ways: (i) as a genuine own preference for existing over new treatments or (ii) a strong awareness of the preferences among the public these policymakers need to represent. Withdrawing reimbursement of an intervention may lead to more protests than not reimbursing the same intervention in the first place, and policymakers may be aware of this fact.

10.1.1. Methodological considerations

Besides these main findings, the methodological approaches used in the first part of this thesis provided some interesting additional insights. In chapters 3 to 6, DCE's and a CV study, were reported deriving respondents' stated distributional preferences. The results from each study were used to further refine the subsequent stated preference

experiments. While this stepwise approach may hamper the comparability of the different studies, it contributed to the refinement of stated preference methods to derive distributional preferences and the social (monetary) value of QALY gains. In addition, it underlines the impact different methodologies, designs and assumptions can have on the relative distributional preferences found in empirical studies.

Furthermore, the design of the DCE presented in chapter 3 was built in three phases to investigate the effect of providing additional information to the choice scenarios on the relative weights attached to different attributes. To do so, the attributes of culpability, rarity and having dependents were added in the third phase of the DCE. As a result, the unobserved variance decreased compared to the previous two phases, which only included information on severity of illness and age. Furthermore, the inclusion of additional attributes appeared to affect the relative preferences for severity of illness and fair innings. This indicates that although not all attributes may be considered relevant for decision making from a normative perspective, including them in choice experiments may still contribute to our understanding of societal preferences for each single attribute. This issue is further emphasized in a recent review by Whitty et al. [22]. They argue that the omission of relevant criteria (i.e. information) in a choice experiment may explain some of the inconsistent findings between studies.

Finally, the estimation of latent class models in chapter 6 revealed distinct preference patterns in the data, which stresses the importance of accounting for preference heterogeneity among the public in value-laden issues such as prioritizing health care. This is relevant for both research and decision-making. Analysis of data at the aggregate level may result in recommendations that do not reflect actual preferences in society. Therefore, it is important for future studies to consider multiple models in order to explore possible decision patterns in the data.

10.2. PART B: HAPPINESS IN ECONOMIC EVALUATIONS

The second part of this thesis focused on happiness as a broader outcome measure for economic evaluations, addressing the research questions below:

f) Do respondents focus on health, happiness or both in prioritizing patients in health care?
g) What are the determinants of happiness and the role of health therein?
h) Is there a causal effect of health on happiness?

Chapter 7 answered research question *f* and showed that about half of the respondents in our study were unwilling to discriminate between patient groups based on their health and happiness levels in allocating health care. This seems to reflect a general

aversion against choosing between patients in need of health care. The other half of the respondents in the experiment were willing to make choices between patient groups based on their health and happiness, suggesting that both health and happiness could play a role in priority setting. In this latter group, gains in health received a marginally higher weight than gains in happiness. The finding that about half of the respondents considered both health and happiness in their decisions which patient group to prioritize indicates that broader outcome measures may yield relevant information for allocation decisions in health care [10].

With respect to research question *g*, the results of chapter 8 showed that happiness is a multidimensional concept that is determined by many factors. Moreover, people can differentiate their happiness experienced in different life domains. A strong association was found between health and overall happiness, directly and indirectly (i.e. through domain-specific happiness). However, the domain-specific effects of health on happiness differed substantially; differences in health appeared to have a stronger effect on happiness at school than on happiness at home. The results of this chapter clearly illustrated the broad scope of happiness and the significant role of health therein.

Chapter 9 addressed research question *h*, exploring the causal effect of health on happiness as well as the endogeneity that is present when subjective health measures are used to explore the relationship between health and happiness. Correcting for endogeneity, a stronger association was found between health and happiness. Multiple instruments showed similar results. The sensitivity analyses for the selection of instruments, using a different measure of health, and the dependent variable, using a different measure for happiness, confirmed the findings of the main analysis. The findings of this study are at variance with existing literature since many empirical studies demonstrated that the association between objective health and happiness is weaker than the association between subjective health and happiness [210, 234]. However, the results of our study are consistent with those of Garrido et al. [37], the only other study that also explored the simultaneous relationship between health and happiness. Nonetheless, more research remains warranted to further explore the causal relationship between health and happiness.

The main findings of Part A and Part B provide novel insights with respect to equity weights in resource allocation and happiness as a broader outcome measure for economic evaluations. As a way of bridging these two parts of the thesis, happiness may also function as an equity consideration in economic evaluations and thus contribute to a more equitable distribution of health and healthcare. As shown in chapter 7, respondents were willing to discriminate between groups of people based on different health and happiness levels before treatment and gains in happiness and health. This supports

the idea that people may be willing to prioritize patient groups that are relatively unhappy, or diseases that have a greater impact on happiness. Therefore, the findings are an important contribution to the further refinement of economic evaluations with the overarching aim of reconciling the discrepancy between current recommendations based on the outcomes of economic evaluations and the socially desired allocation of health care resources.

10.3. LIMITATIONS AND RESEARCH IMPLICATIONS

Several limitations of the work presented in this thesis need to be mentioned, next to the limitations already mentioned in the separate chapters.

While working on this thesis many theoretical and methodological choices had to be made that deserve discussion. The design of the studies and the analyses of the results were not solely a technical exercise, but often required fundamental value judgments. These choices were made carefully and were justified in line with the (policy relevant) research questions that needed to be addressed, but obviously they still affected the results presented in this thesis. Therefore, we have described the methodological process as transparently as possible, so that other researches can judge the usefulness of our questions, methods and results for their own research.

In this respect, many techniques have been developed to elicit preferences or WTP values from respondents. DCE and CV studies are considered useful tools and are increasingly popular, as shown in a recent review by Whitty et al. [22]. However, the influence of the context of the choice scenarios and the framing of questions on respondents' preferences complicates the comparability between different studies and restricts their external validity. Therefore, we propose that validity checks judging the usefulness of the results and the quality of the studies should become standard procedure in future studies. Several checklists are already available for DCE's, but these should be expanded to include all studies on this topic, as well as studies that use different techniques [22, 239]. Doing so will increase the ability to compare and interpret results of different studies in a meaningful way and stimulate a more consistent methodological approach. This is required to be able to build a sufficient level of evidence on the distributional preferences in society, lending stronger support to their consideration in decision making.

Furthermore, stated preference methods with hypothetical scenarios were used to determine the preferences in society, using online surveys. Therefore, it is questionable whether respondents understood all the questions and whether their preferences for hypothetical situations resemble actual behaviour. Online surveys make it possible to

reach larger, representative samples efficiently and use more complex and interactive designs, but make monitoring the intelligibility of the questions and the engagement of respondents more difficult. The studies presented in this thesis were all carefully pilot-tested, some included consistency checks, and time limits were defined to exclude respondents rushing through the questionnaire. Nevertheless, the reliability and validity of stated preference data collected online still remain issues of concern and deserve further investigation. Future studies should address these issues carefully as well. For instance, findings could be presented to the study participants to investigate the validity of their preferences.

In this thesis, preferences of the public were collected only from Dutch citizens. Therefore, the geographical and cultural generalizability of the results is limited. An avenue for future research would be to replicate our experiments in different settings. However, given the impact of the context on decisions, it might be preferable to derive equity weights for priority setting that are country and context specific and, hence, locally relevant. With respect to the happiness studies presented in this thesis, it would be interesting to study panel data in addition to the cross-sectional data used here, to investigate the impact of time-invariant factors on happiness measures. This would contribute to the understanding of what constitutes happiness and would further advance the estimation of the causal relation between health and happiness.

While the findings of Part B contribute to the happiness literature, many relevant questions remained unaddressed and new questions have emerged. Chapter 7 showed that a considerable proportion of the general public was willing to consider both health and happiness in their allocation decisions. However, *how* happiness can be incorporated in economic evaluations and subsequent decision making remained unanswered. In order to determine whether happiness can ultimately replace or complement health as an outcome measure for economic evaluations, future research should for instance investigate the extent to which happiness instruments adequately capture all relevant health aspects. Moreover, since happiness measurement is relatively new in health economics, the reliability of current measures for its use in economic evaluation should be further investigated. Another option may be to include both health and happiness measures in economic evaluations and decision making. Then, obviously, double counting needs to be avoided and suitable ways to combine these measures (also in decision making) need to be found.

Future research should also pay attention to the possible contribution of happiness measures to equity weighting. The results of this thesis showed that people might be willing to prioritize patient groups that are relatively unhappy, or diseases that have a greater impact on patients' happiness. How happiness as an equity consideration

should be operationalized to be able to reflect the preferences of the population for a fair distribution of happiness, is an important question for future research. If both health and happiness considerations would be included in decision making, the relative importance of the two should also be further investigated. A DCE might be a useful method for doing so.

Furthermore, as already briefly mentioned in the introduction of this thesis, different definitions of happiness are used interchangeably in the literature and the formulation of the happiness measures and corresponding measurement scales differs between studies as well [32]. This complicates the comparability between different studies, a meaningful interpretation of the study results and hampers their external validity. Future research should aim to improve the consistency in definition and measurement of happiness.

More generally, in order to move forward in this field, additional research is required on both topics addressed in this thesis. Further empirical research is essential to gain more insight into distributional preferences in society, the monetary social value of a QALY, and the role of happiness as an outcome measure for economic evaluations. In doing so, it needs to be emphasized that not all derived preferences are necessarily appropriate from a normative perspective. Therefore, in future studies, close collaboration between researchers and policy makers may be beneficial.

10.4. GENERAL CONCLUSION AND POLICY IMPLICATIONS

This thesis has provided more insight into distributional concerns and the scope of the outcome measures in health care allocation decisions. The results may contribute to improving health economic evaluations and (subsequent) health care decision making.

Different equity principles have been proposed with which equity concerns could be captured in the context of health care allocation decisions. In the literature, the concepts of severity of illness and fair innings have been frequently discussed. Moreover, on a practical level, several countries are attempting to operationalize equity concepts in their decision making frameworks. Well-known examples are the Dutch concept of proportional shortfall and the special treatment of 'end-of- life interventions' of NICE, in the UK. While evidence accumulates about distributional preferences in society, related to a desire for an equitable distribution of health and health care, policy makers should be aware that these may be difficult to describe with one particular set of weights or decision rules. Distributional preferences seem to be context specific and heterogeneous. Further investigation into suitable equity concepts and potentially more complex

weighting schemes remains important. It would also be interesting for policy makers to elaborate further the equity concerns related to the desirability of the health state after treatment and its relevance for policy making.

The underlying differences of opinion in the general public further emphasize the complexity of the task of incorporating the distributional preferences of society in economic evaluations and health care decisions. Achieving full consensus regarding an appropriate equity principle or an appropriate set of equity weights may not only be challenging, but in fact may be impossible to achieve. Nonetheless, it seems worthwhile to take further steps on the road to quantifying preferences in society, providing more insight into the complexity of the decision making process and situations in which equity concerns may conflict. This knowledge is of great value for the further development of equity concepts in decision making frameworks and for better understanding the nature and degree of public support for specific choices. Moreover, policy makers should establish the normative boundaries of defining an equitable distribution of health care resources, since not all empirically derived preferences may be considered relevant or desirable from a normative viewpoint.

With respect to conditional reimbursement, decision makers need to be aware that stopping reimbursement proves to be more difficult than not starting in the first place. Against the benefit of allowing more time to gather appropriate scientific evidence for a permanent reimbursement decision without unduly delaying market access of a technology, there is the risk that engaging in conditional reimbursement schemes may lower the chances of subsequent withdrawal of reimbursement.

This thesis also addressed happiness as a broader outcome measure for economic evaluations. In an era of ageing populations and increasing demand for long term care, the boundaries between health and social care and between health and wellbeing become increasingly blurred. Therefore, considering happiness as an outcome measure seems directly relevant for policy makers attempting to assess all relevant impacts of health interventions. Happiness is a broader outcome measure than health and therefore may be expected to reflect individual welfare better, or at least, to provide relevant complementary information for allocation decisions. When interventions, such as in the long term care sector or in elderly care, aim to improve wellbeing rather than health (alone), such broader measures may be necessary in order to capture the full benefits of these interventions. A more comprehensive representation of welfare then may help to improve decisions about the allocation of health care resources. Moreover, it can lead to a more efficient allocation of resources between sectors of the economy, based on their relative contributions to societal wellbeing. Hence, a full welfare economic assessment of interventions may require a broader set of outcome measures than currently typi-

cally the case. Happiness measures may then be useful, as may other measures, such as capability measures [169], which were not included in this thesis.

Health has been shown to be an important contributor to overall happiness, which may also add to the relevance of happiness as an outcome measure for health care allocation decisions. However, it is important for policymakers to realize that health is only one determinant of happiness. Happiness is a broad, multidimensional concept, which captures many more aspects of human quality of life than health alone. Hence, health care interventions may mainly improve just one element important to overall happiness. Furthermore, in the debate on broader outcome measures, the relationship between health and happiness may suffer of endogeneity bias when based on subjective measures. Policymakers who consider the impact of health on happiness should be aware that they may underestimate this effect when not correcting for endogeneity bias.

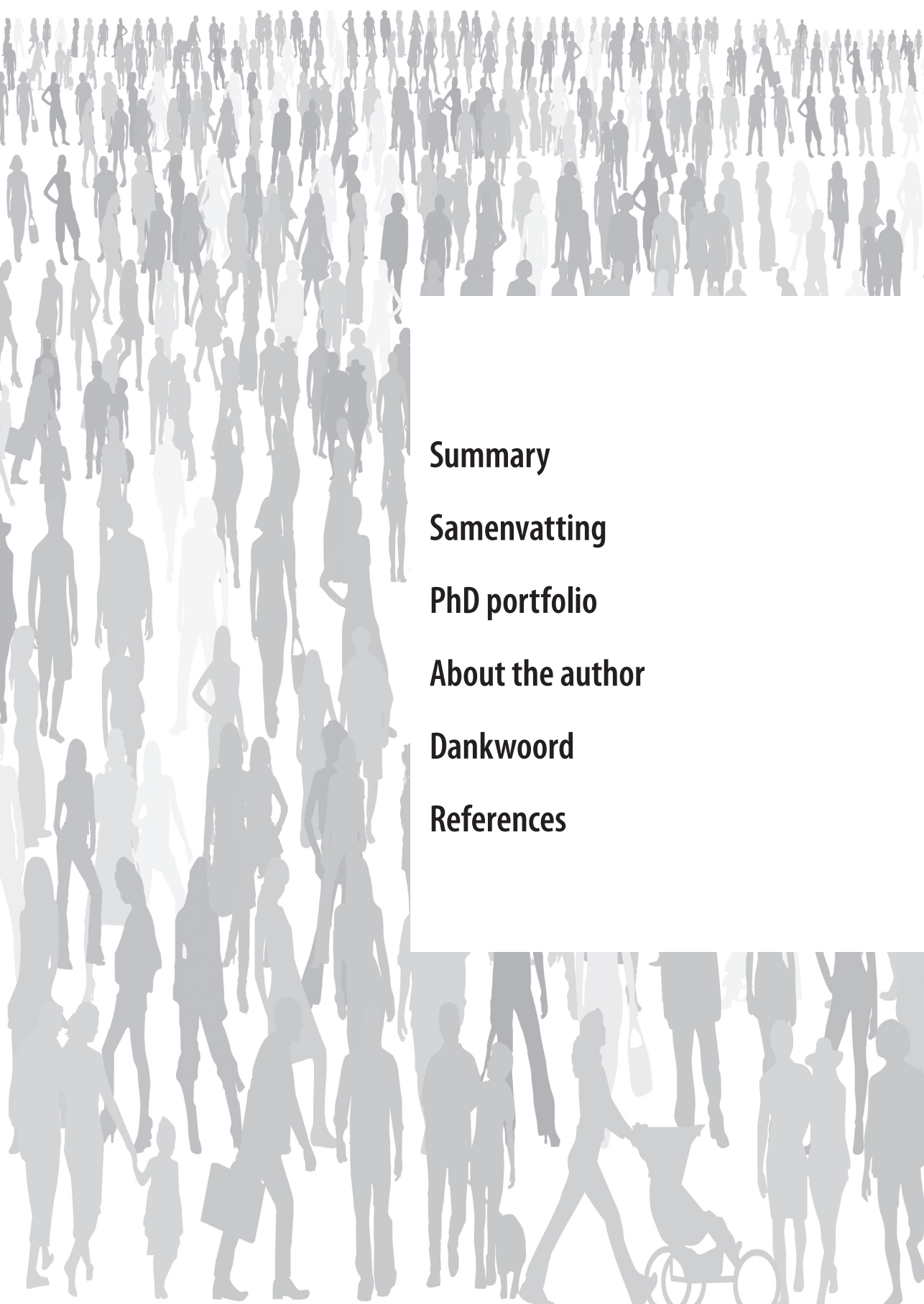
This thesis underlines the additional value of happiness as an outcome value for economic evaluations. However many questions remain. Whether and how happiness should be incorporated in economic evaluations and health care decision making is an important question that requires further research, but also policy debate. Policy makers could consider to start collecting information on happiness alongside information on health in order to improve the understanding of the relationship between health and happiness and the additional value of happiness as a broader outcome measures in economic evaluations. While writing this thesis in November 2015, Statistics Netherland published the first results of a survey of wellbeing in The Netherlands [240]. This underlines the rising social relevance of happiness indicators for the debate on social progress.

10.5. FINAL REMARKS

Given limited health care resources, implicit or explicit decisions regarding the allocation of scarce health care resources are inevitable. When economic evaluations are used to inform such decisions, it is pivotal that they are adequately performed, align with equity concerns in society and the goal of improving social welfare. This thesis has contributed to the theoretical and practical knowledge regarding equity weights and happiness as a broader outcome measure to capture outcomes beyond the QALY. Many relevant questions in these areas remain unanswered. Therefore, sufficient room should be left in the appraisal phase of the decision making process to address those aspects that have not yet been (or perhaps can ever be) adequately captured and quantified in the assessment phase. For instance, sufficient room should be present to judge whether quantified equity weights accurately reflect public preferences or normative principles in particular circumstances. Moreover, continuous attention should be given to further improvement

and refinement of equity concepts and their operationalization. Furthermore, it may be worthwhile to collect happiness information alongside QALYs in economic evaluations, in particular when it concerns interventions not purely aimed at cure.

Hence, it is important to continue developing and refining equity concepts and outcome measures that better reflect and support a socially desired allocation of health care resources. Defining and quantifying the relevant elements in the decision making process where possible may help increase the transparency and consistency of health care allocation decisions. Many steps still have to be taken in improving economic evaluation as an optimal tool for supporting fair and welfare improving policy decisions in health care. This thesis hopes to have contributed to that ultimate goal.



SUMMARY

In health care, economic evaluations are used to inform decision makers about the costs and effects of interventions on the health and welfare of the population. Decision makers use this information to select which treatments will be included in the basic benefits package and which treatments will not. Scientific studies and frequent reports in the media regularly show that there is a discrepancy between the recommendations from economic evaluations and public preferences for the allocation of health care resources. To reduce this discrepancy and so make economic evaluations more useful for health care decision makers, some important methodological challenges need to be addressed. This thesis investigated two of these methodological issues, related to the measurement and valuation of outcomes in economic evaluations. The first part of this thesis addressed the concerns in society with respect to the distribution of health and health care. The public my for instance prefer to give priority to treating more severely ill patients, even if the potential health gains are relatively small, in favour of a more equitable distribution of health and health care. The second part focused on happiness as a broader outcome measure for economic evaluations. Some interventions in the health care sector, for example in elderly care, may not affect the health of patients, but improve their wellbeing. Happiness as a broader outcome measure than health alone may better capture the benefits of such interventions. Addressing these issues we aimed to contribute to the improvement of economic evaluations in health care, reconciliation of recommendations based on economic evaluations with preferences in society, and ultimately to a better allocation of scarce health care resources.

The preferences in society with respect to the distribution of healthcare resources can be integrated in the decision making process by weighting health benefits, measured in terms of Quality Adjusted Life Years (QALYs), according to some equity principle. This equity principle should ideally reflect the relevant distributional preferences in society. Chapter 2 discussed how efficiency and concerns for equity are combined in the Dutch decision making framework and described the difficulties the Netherlands faced when the equity principle proportional shortfall was put in practice. The findings emphasize the need to further develop and refine this equity principle and its use in practice, and provide important lessons for other countries, since many of these issues seem to be universal and irrespective of the equity principle chosen.

In chapters 3 to 6 the distributional preferences of the Dutch population with respect to health care allocation were investigated. Chapter 3 presented results of an experiment deriving the relative weights of the equity principles 'severity of illness' and 'fair innings' and some additional characteristics of patients and their disease. All these characteristics were important for explaining respondents' preferences. In chapters 4 and 5 we inves-

tigated how much society was willingness to pay (WTP) for QALY gains in subgroups of patients which were characterised by specific equity characteristics. Distinct preferences were found for different descriptions of severity as an equity principle, and the willingness to pay for a QALY ranged between €94,700 and €296,756. The distinct preferences and the wide range of WTP values are consistent with the scientific literature and seem to result from the chosen equity principle, the context of the decision and the different methods used.

Chapter 6 investigated the policy instrument conditional reimbursement, and showed that the current reimbursement status of an intervention (i.e. whether or not an intervention is currently included in the basic benefits package) affects the value society attaches to QALY gains. Both the general public and decision makers valued a currently reimbursed intervention higher than an intervention that is currently not yet reimbursed. Hence, policy makers need to be aware of the fact that engaging in conditional reimbursement may lower the chances of subsequent withdrawal of reimbursement.

The second part of this thesis focused on the scope of outcome measures in economic evaluations. Broader outcome measures than health, such as happiness, may better capture all socially relevant benefits of an intervention. The results of chapter 7 showed that about half of the respondents were willing to discriminate between patient groups based on differences in levels and gains in both health and happiness. This supports the idea that broader outcome measures such as happiness may provide additional information that is relevant for the decision how to distribute health care resources.

The last two chapters of this thesis addressed the concept of happiness and the relationship between health and happiness. Chapter 8 provided an extensive overview of what constitutes happiness and emphasized the broad scope of the happiness and the significant role of health therein. The results of chapter 9 indicated that health has a direct effect on happiness, but that the use of subjective measures of health and happiness may bias this relation. In the debate on broader outcome measures, policy makers who consider the effect of health on happiness should be aware that they might underestimate this effect when not correcting for this endogeneity bias.

This thesis contributes to the development and refinement of outcome measures for economic evaluations. However, many relevant questions remained unanswered. Therefore, sufficient room should be left in the decision making process to judge whether outcomes of economic evaluations accurately reflect and support public preferences regarding the allocation of health care resources. Meanwhile, it is important to further develop and refine equity principles and their use in practice. Furthermore, to stimulate and further develop the debate on happiness measures in economic evaluations it seems worthwhile to start collecting happiness information alongside health effects in

economic evaluations. Moreover, still many steps have to be taken to further improve economic evaluations with the overarching aim to make economic evaluations a more useful source of information for health care decision making that is more aligned with distributional preferences in society. This thesis hopes to have contributed to that ultimate goal.

SAMENVATTING

In de gezondheidszorg worden economische evaluaties gebruikt om beleidsmakers te informeren over de kosten en effecten van interventies op de gezondheid en het welzijn van de bevolking. Beleidsmakers gebruiken deze informatie bij het bepalen van welke interventies worden opgenomen in het basispakket, en welke niet. De aanbevelingen op basis van economische evaluaties komen echter niet altijd overeen met de maatschappelijke opvattingen over een eerlijke verdeling van middelen in de gezondheidszorg. Om deze discrepantie te verkleinen en economische evaluaties bruikbaarder te maken voor beleidsmakers dient een aantal belangrijke (methodologische) uitdagingen te worden overwonnen. In dit proefschrift zijn twee van deze vraagstukken nader onderzocht. Beide hebben betrekking op het meten en waarderen van uitkomsten in economische evaluaties. Het eerste deel van dit proefschrift onderzocht de maatschappelijke voorkeuren voor een optimale verdeling van middelen in de gezondheidszorg. Het kan bijvoorbeeld zo zijn dat mensen uit oogpunt van een rechtvaardigere verdeling van gezondheid en gezondheidszorg meer waarde hechten aan de behandeling van patiënten met een ernstige aandoening, zelfs wanneer de gezondheidswinst bij deze patiënten relatief klein is. Het tweede deel van dit proefschrift richtte zich op geluk als een bredere uitkomstmaat voor economische evaluaties. Sommige interventies in de gezondheidzorg, zoals bijvoorbeeld in de ouderenzorg, hebben geen of weinig effect op gezondheid maar verbeteren wel het welzijn van patiënten. Welzijn als bredere uitkomstmaat dan gezondheid lijkt in die situaties beter in staat om alle baten van een interventie te omvatten. Door het behandelen van deze twee vraagstukken hopen we bij te dragen aan het verbeteren van economische evaluaties in de gezondheidszorg zodat zij beter aansluiten bij de maatschappelijke voorkeuren voor een rechtvaardige verdeling van zorgmiddelen. Op deze manier hopen we dat economische evaluaties meer invloed krijgen in het besluitvormingsproces over de aanwending van schaarse middelen in de gezondheidszorg.

De maatschappelijke voorkeuren voor de optimale verdeling van middelen in de gezondheidszorg kunnen geïntegreerd worden in het besluitvormingsproces. Dit kan door de gezondheidseffecten, uitgedrukt in voor kwaliteit gecorrigeerde levensjaren (QALYs), te wegen op basis van een bepaald rechtvaardigheidsprincipe. Hoofdstuk 2 beschrijft hoe doelmatigheids- en rechtvaardigheidsoverwegingen worden gecombineerd in het Nederlandse besluitvormingsproces. Tevens beschrijft dit hoofdstuk de uitdagingen waarmee Nederland geconfronteerd werd bij de operationalisering van het rechtvaardigheidsprincipe 'proportional shortfall'. De bevindingen benadrukken de noodzaak tot een verdere ontwikkeling en verfijning van dit rechtvaardigheidsprincipe en de toepassing ervan in de praktijk. De uitdagingen die gepaard gingen met de ope-

rationalisering van het principe 'proportional shortfall' zijn veelal ook relevant wanneer een ander rechtvaardigheidsprincipe wordt gekozen. De bevindingen van dit hoofdstuk zijn derhalve ook relevant voor andere landen die rechtvaardigheidsoverwegingen in economische evaluaties willen meenemen.

In de hoofdstukken 3 tot en met 6 zijn de rechtvaardigheidsoverwegingen van de Nederlandse bevolking met betrekking tot de verdeling van middelen in de gezondheidszorg empirisch onderzocht. Hoofdstuk 3 presenteerde de resultaten van een experiment naar de relatieve gewichten van de rechtvaardigheidsprincipes 'severity of illnes', 'fair innings' en 'proportional shortfall', en een aantal additionele kenmerken van de patiënt en de ziekte. Alle kenmerken bleken belangrijk bij het verklaren van de voorkeuren van respondenten voor de verdeling van zorg. In hoofdstuk 4 en 5 werd de maatschappelijke betalingsbereidheid per QALY gezondheidswinst onderzocht voor subgroepen van patiënten die werden gekenmerkt door specifieke rechtvaardigheidskenmerken. Uiteenlopende voorkeuren werden gevonden voor verschillende definities van 'severity of illness' als rechtvaardigheidsprincipe, en de maatschappelijke waarderingen per QALY varieerden van €94,700 tot €296,756. De verscheidenheid aan voorkeuren en de grote variatie in betalingsbereidheid per QALY zijn consistent met de wetenschappelijke literatuur, en lijken samen te hangen met het gekozen rechtvaardigheidsprincipe, de context van de besluitvorming, en de verschillende onderzoeksmethoden die worden gebruikt.

Hoofdstuk 6 onderzocht het beleidsinstrument 'conditionele vergoeding', ofwel voorwaardelijke vergoeding van interventies vanuit het basispakket. De resultaten lieten zien dat de huidige vergoedingsstatus van een interventie (d.w.z. of een interventie op het moment wel of niet vergoed wordt) invloed had op de waardering van een QALY gezondheidswinst. Zowel de algemene bevolking als beleidsmakers hechtten meer waarde aan de vergoeding van een interventie die op dat moment al werd vergoed dan een interventie die niet werd vergoed. Beleidsmakers moeten zich zodoende realiseren dat conditionele vergoeding het moeilijker maakt om de vergoeding later te beeindigen.

Het tweede deel van dit proefschrift richtte zich op de reikwijdte van uitkomstmaten in economische evaluaties. Bredere uitkomstmaten dan gezondheid, zoals geluk, zijn mogelijk beter in staat om alle maatschappelijk relevante baten van een interventie te omvatten. De resultaten van hoofdstuk 7 lieten zien dat ongeveer de helft van de respondenten bereid was onderscheid te maken tussen verschillende groepen patiënten op basis van verschillen in huidige gezondheid en geluk en de winst daarin door behandeling. Dit ondersteunt het idee dat bredere uitkomstmaten zoals geluk relevante informatie bevatten voor de besluitvorming over de verdeling van middelen in de gezondheidszorg. Gezondheid was meestal iets belangrijker dan geluk in deze keuzen.

De laatste twee hoofdstukken behandelden het concept geluk en de relatie tussen gezondheid en geluk. Hoofdstuk 8 presenteerde de resultaten van een onderzoek naar determinanten van geluk. Dit onderzoek toonde een breed overzicht van aspecten die bijdragen aan geluk, waarbij gezondheid een belangrijke rol speelde. De resultaten van hoofdstuk 9 lieten zien dat gezondheid een direct effect heeft op geluk maar dat het subjectieve karakter van beide grootheden de relatie kan vertekenen. Beleidsmakers dienen zich te realiseren dat ze het effect van gezondheid op geluk mogelijk onderschatten wanneer ze niet corrigeren voor deze endogeniteit.

Dit proefschrift draagt bij aan de ontwikkeling en verfijning van uitkomstmaten voor economische evaluaties. Veel vragen blijven echter onbeantwoord. Daarom moet er genoeg ruimte blijven in het besluitvormingsproces om te beoordelen of de uitkomsten van economische evaluaties overeenkomen met de maatschappelijke voorkeuren voor de verdeling van middelen in de gezondheidszorg. In de tussentijd is het belangrijk om rechtvaardigheidsprincipes en hun praktische operationalisatie verder te ontwikkelen en te verfijnen. Om het debat over welzijn als uitkomstmaat in economische evaluaties te stimuleren is het belangrijk om naast informatie over gezondheid ook informatie over welzijn te verzamelen in economische evaluaties. De bruikbaarheid van economische evaluaties in het besluitvormingsproces over de aanwending van schaarse middelen in de gezondheidszorg wordt vergroot wanneer zij beter aansluiten bij maatschappelijke voorkeuren voor een optimale verdeling van gezondheid en gezondheidszorg. Om dit te bewerkstelligen is er nog een lange weg te gaan in het verbeteren van de methodologie van economische evaluaties. Dit proefschrift heeft hopelijk een bijdrage geleverd aan het bereiken van dit ultieme doel.

PhD PORTFOLIO

PHD TRAINING

Discrete Choice Modelling, Professor William Greene, University of Southern Denmark

Executive Course Discrete Choice Experiment, Erasmus University Rotterdam

Discrete Choice Modelling, Centre for Microdata Methods and Practice, Institute for Fiscal Studies, London

Advanced Course in Applied Health Economics, Methods for the analysis of categorical dependent variables, Centre for Health Economics, The University of York

Academic Writing, Academic Language Centre, Erasmus University Rotterdam

Course Basic Didactics, Risbo, Erasmus University Rotterdam

Tutor training course 'Problem Based Learning', Institute of Psychology, Erasmus University Rotterdam

Master Health Services Research - Netherlands Institute for Health Sciences (NIHES)

- Study Design
- Classical Methods for Data-analysis
- Modern Statistical Methods
- Pharmaco-epidemiology
- Analysis of Population of Health
- Analysis of Determinants
- International Comparison of HCS
- Economic Evaluation
- Research on Quality of Care
- Evidence-based Policymaking
- Qualitative Research

Erasmus Summer Programme - Netherlands Institute for Health Sciences (NIHES)

- Topics in Meta-analysis

Decision making in Medicine

Prevention Research

Topics in Evidence-based Medicine

Principles of Research in Medicine

Introduction to Public Health

Methods of Health Services Research

Methods of Public Health Research

Health Economics

TEACHING

Patient preferences in the delivery of health care, master program, Institute of Health Policy and Management, Erasmus University Rotterdam

Quality and efficiency in healthcare, bachelor program, Institute of Health Policy and Management, Erasmus University Rotterdam

Integrated care, bachelor program, Institute of Health Policy and Management, Erasmus University Rotterdam

Quality of health care, pre-master program, Institute of Health Policy and Management, Erasmus University Rotterdam

Supervisor and co-evaluator for bachelor and master theses, Institute of Health Policy and Management, Erasmus University Rotterdam

CONFERENCES

European Conference on Health Economics. (2014). Dublin, Ireland. (Work discussed)

European Conference on Health Economics. (2012). Zurich, Switzerland. (Presenter)

European Conference on Health Economics. (2010). Helsinki, Finland. (Poster)

3nd Lowlands Health Economists' Study Group. (2011). Soesterberg, the Netherlands. (Attendant)

2nd Lowlands Health Economists' Study Group. (2010). Egmond aan Zee, the Netherlands. (Work discussed)

9th International Conference on the Mechanisms and Treatment of Neuropathic Pain. (2006). Bermuda. (Attendant)

SCIENTIFIC PUBLICATIONS

Van de Wetering, E., van Exel, N., & Brouwer, W. (2015). Health or Happiness? A note on trading off health and happiness in rationing decisions. Value in Health. In press.

Van de Wetering, E., van Exel, J., Bobinac, A., & Brouwer, W. B. (2015). Valuing QALYs in Relation to Equity Considerations Using a Discrete Choice Experiment. PharmacoEconomics, *33*(12), 1289-1300.

Van de Wetering, E., Van Exel, N., Rose, J. M., & Brouwer, W. (2016). Are some QALYs more equal than others?. European Journal of Health Economics, 17(2), (117-127).

Van de Wetering, E., Stolk, E., Van Exel, N., & Brouwer, W. (2013). Balancing equity and efficiency in the Dutch basic benefits package using the principle of proportional shortfall. The European Journal of Health Economics, 14(1), 107-115.

Van de Wetering, E., van Exel, N., & Brouwer, W. (2010). Piecing the Jigsaw Puzzle of Adolescent Happiness. Journal of Economic Psychology, 31(6), 923-935.

Van de Wetering, E. J., Lemmens, K. M., Nieboer, A. P., & Huijsman, R. (2010). Cognitive and behavioral interventions for the management of chronic neuropathic pain in adults–a systematic review. European Journal of Pain, *14*(7), 670-681.

Hoefman, R., Van Exel, N., Rose, J., Lawerman-van de Wetering, L., Brouwer, W. (2014). A Discrete Choice Experiment to Obtain a Tariff for Valuing Informal Care Situations Measured with the CarerQol Instrument. Medical Decision Making 34 (1), 84-96.

ABOUT THE AUTHOR

Liesbet Lawerman-van de Wetering was born in 's-Gravenhage on July 19th 1983. In 2001 she started her Bachelor in Health Policy and Management at the Erasmus University Rotterdam. During this program she did an internship at Harvard Medical School and developed an interest in scientific research. From 2004 to 2006 she followed a two year research master program in Clinical Epidemiology with a specialization in Health Services Research, at the Netherlands Institute for Health Sciences (NIHES). After obtaining her Master degree she continued her studies by enrolling in the master Health Economics, Policy and Law (HEPL) with a specialization in Health Economics at the institute of Health Policy and Management (iBMG). During her study she started working as a research assistant at the institute for Medical Technology Assessment (iMTA). From 2008 to 2009, she worked on a project concerning the evaluation of disease-managements programs for neuropathic pain. In 2009 she started a PhD project on the valuation and refinement of outcome measures for economic evaluations in healthcare, which resulted in this dissertation.

Liesbet is married to Olaf, they have two daughters: Lune (2012) and Joslin (2014). They are currently expecting their third daughter.

DANKWOORD

Het voelt nog steeds een beetje onwerkelijk dat er nu 'ineens' een proefschrift ligt. Voor mij een mooie afronding van een onvergetelijke periode waarin ik met veel plezier naar mijn 'werk' ging en ontzettend veel geleerd heb. Ik wil dan ook iedereen die direct of indirect heeft bijgedragen aan dit proefschrift hartelijk bedanken.

Werner en Job, wat ben ik blij dat jullie mij de kans hebben gegeven om onder jullie begeleiding te promoveren. Ik had me oprecht geen betere begeleiders kunnen wensen. Ik heb ontzettend veel van jullie geleerd en ik heb de samenwerking altijd als heel prettig ervaren. Het soms wat lange wachten op feedback was altijd het wachten waard ;-). Bedankt voor alle ruimte en vrijheid die ik gekregen heb om mezelf op mijn eigen tempo te ontwikkelen, zodat ik ook zonder moeite al mijn vakanties kon inplannen. Werner, ik heb bewondering voor je creatieve benadering van vraagstukken en de vanzelfsprekendheid waarmee je dat doet. Job, jouw deur stond altijd voor me open en ik kon altijd op je rekenen. Jouw mening, zonder te oordelen, heb ik altijd ontzettend gewaardeerd en ga ik nog vaak missen!

Elly, jou wil ik bedanken voor alle hulp en advies, over zowel 'proportional shortfall' als lastige DCE vragen. Zonder dat je direct betrokken was bij mijn proefschrift wist je altijd wel even tijd voor me vrij te maken.

De promotiecommissie wil ik bedanken voor het beoordelen van mijn proefschrift en voor het opponeren bij de verdediging.

Steef en Renske, jullie hebben tijdens dit traject het dichtst bij mij gestaan en ik ben heel blij dat jullie mijn paranimfen zijn en dus ook vandaag dicht bij me zullen staan. Steef, ik kan je niet genoeg bedanken voor alle hulp, maar ook voor je vriendschap en alle gezelligheid. De tijd samen op de universiteit was goud waard, zo heerlijk vanzelfsprekend om je daar elke dag te zien en even je kamer binnen te lopen! Rens, wij hebben samen voor heel wat uitdagingen gestaan, niet alleen vakinhoudelijk maar ook op onze buitenlandse reisjes… Als kamergenoten en vriendinnen hebben we veel gedeeld samen, veel gelachen, veel thee gedronken en veel gesnoept. Bedankt voor alle momenten en het telkens bijvullen van de snoeppot.

Sofie, gelukkig heb je je plekje op onze kamer al snel weten te bemachtigen. Ik heb het als heel bijzonder ervaren hoe wij samen de laatste loodjes van ons proefschrift hebben afgerond en sorry voor alle babypraat de afgelopen 3 jaar. Ook alle andere collega's van BMG wil ik bedanken voor de ontzettend leuke werksfeer bij BMG.

Lieve vrienden, bedankt voor jullie gezelligheid, afleiding en interesse in mijn proefschrift. In het bijzonder mijn studievrienden Steef, Peter, Kees-Jan en Niels, ik ben nu

echt de laatste die afscheid neemt van de uni. Wat hebben we een mooie tijd gehad en ik ben ontzettend blij met onze vriendschap die daaruit voortgekomen is. Onze etentjes en onze jaarlijkse reisjes zijn voor mij erg waardevol. Lieve Karlijn, ook jou heb ik tijdens onze BMG studie leren kennen. Hoewel jij op een andere universiteit je proefschrift hebt geschreven, hebben we altijd veel samen kunnen delen. Ik heb ontzettend veel bewondering voor je zelfdiscipline, doorzettingsvermogen en nuchtere kijk op dingen. Bedankt voor je vriendschap, alle 'strenge woorden' en aanmoediging die je me gegeven hebt wanneer nodig.

Lieve familie, jullie onvoorwaardelijke liefde en steun hebben mij zeker gebracht waar ik nu ben. Pap, mijn interesse in de wetenschap heb ik van jou geërfd en het deed mij goed om dit met jou te kunnen delen en je altijd om advies te kunnen vragen. Mam, zonder jou had ik het ook nooit gered, alle uren die je op de meisjes hebt gepast hebben mij zo geholpen. Heerlijk om ze zo zichtbaar genietend bij jou achter te kunnen laten! Marleen en Fleur jullie zijn waanzinnige zusjes, wat ben ik blij dat ik jullie heb! Roel en Wil, ik wil jullie ook bedanken voor het oppassen in de laatste weken voor het afronden van mijn proefschrift.

Lieve Olaf, bedankt voor je stimulans om dit proefschrift af te ronden en alle ruimte die ik daarvoor gekregen heb. Wat hebben we het heerlijk gehad in Australië en ik heb ontzettend veel zin in de toekomst met jou en onze meisjes! Lieve Lune en Joslin, ik geniet elke dag weer van jullie aanwezigheid en ik ben blij dat ik me de komende tijd even volledig op jullie kan richten. Deze zomer krijgen jullie er nog een zusje bij. Wat zijn we happy!

REFERENCES

1. Drummond, M. F. (2005). Methods for the economic evaluation of health care programmes: Oxford university press.
2. Dolan, P., Shaw, R., Tsuchiya, A., & Williams, A. (2005). QALY maximisation and people's preferences: a methodological review of the literature. Health Econ, 14, 197-208.
3. Cookson, R., Drummond, M., & Weatherly, H. (2009). Explicit incorporation of equity considerations into economic evaluation of public health interventions. Health Economics, Policy and Law, 4(02), 231-245.
4. Coast, J. (2004). Is economic evaluation in touch with society's health values?. BMJ (Clinical research ed.), 329(7476), 1233-1236.
5. Brouwer, W. B., Culyer, A. J., van Exel, N Job A, & Rutten, F. F. (2008). Welfarism vs. extra-welfarism. Journal of health economics, 27(2), 325-338.
6. Weinstein, M. C., Torrance, G., & McGuire, A. (2009). QALYs: the basics. Value in health, 12(s1), S5-S9.
7. Gravelle, H., Brouwer, W., Niessen, L., Postma, M., & Rutten, F. (2007). Discounting in economic evaluations: stepping forward towards optimal decision rules. Health economics, 16(3), 307-318.
8. Rawlins, M. D., & Culyer, A. J. (2004). National Institute for Clinical Excellence and its value judgments. British medical journal, 329(7459), 224-227.
9. Brousselle, A., & Lessard, C. (2011). Economic evaluation to inform health care decision-making: promise, pitfalls and a proposal for an alternative path. Social science & medicine, 72(6), 832-839.
10. Makai, P., Brouwer, W. B., Koopmanschap, M. A., Stolk, E. A., & Nieboer, A. P. (2014). Quality of life instruments for economic evaluations in health and social care for older people: a systematic review. Social science & medicine, 102, 83-93.
11. Coast, J. (2004). Is economic evaluation in touch with society's health values?. BMJ (Clinical research ed.), 329(7476), 1233-1236.
12. Dolan, P. (2008). Developing methods that really do value the 'Q'in the QALY. Health economics, policy and law, 3(1), 69-77.
13. Stolk, E. A., Pickee, S. J., Ament, A. H. J. A., & Busschbach, J. J. V. (2005). Equity in health care prioritisation: an empirical inquiry into social value. Health policy, 74(3), 343-355.
14. Schwappach, D. L. B. (2002). Resource allocation, social values and the QALY: a review of the debate and empirical evidence. Health Expectations, 5(3), 210-222.
15. Williams, A. (1997). Intergenerational equity: an exploration of the 'fair innings' argument. Health economics, 6(2), 117-132.
16. Dolan, P., & Olsen, J. A. (2001). Equity in health: the importance of different health streams. Journal of health economics, 20(5), 823-834.
17. Shah, K. K. (2009). Severity of illness and priority setting in healthcare: A review of the literature. Health policy, 93(2-3), 77-84.
18. Stolk, E. A., van Donselaar, G., Brouwer, W. B. F., & Busschbach, J. J. V. (2004). Reconciliation of economic concerns and health policy: illustration of an equity adjustment procedure using proportional shortfall. PharmacoEconomics, 22(17), 1097-1107.

19. CVZ. (2006). Pakketbeheer in de praktijk, CVZ Rapport 245; Diemen: College voor Zorgverzekeringen.
20. CVZ. (2001). Breedte geneesmiddelenpakket, Amstelveen: College voor Zorgverzekeringen.
21. RVZ. (2007). Rechtvaardige en Duurzame Zorg, Den Haag: Raad voor de Volksgezondheid en Zorg.
22. Whitty, J. A., Lancsar, E., Rixon, K., Golenko, X., & Ratcliffe, J. (2014). A Systematic Review of Stated Preference Studies Reporting Public Preferences for Healthcare Priority Setting. The Patient-Patient-Centered Outcomes Research, 1-22.
23. Bobinac, A., Exel, N., Rutten, F. F., & Brouwer, W. B. (2012). Valuing QALY gains by applying a societal perspective. Health Economics, 22(10), 1272-1281.
24. Hutton, J., Trueman, P., & Henshall, C. (2007). Coverage with evidence development: an examination of conceptual and policy issues. International Journal of Technology Assessment in Health Care, 23(04), 425-432.
25. Levin, L., Goeree, R., Levine, M., Krahn, M., Easty, T., Brown, A., et al. (2011). Coverage with evidence development: the Ontario experience. International Journal of Technology Assessment in Health Care, 27(02), 159-168.
26. Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. Econometrica: Journal of the Econometric Society, 263-291.
27. Brazier, J., & Tsuchiya, A. (2015). Improving Cross-Sector Comparisons: Going Beyond the Health-Related QALY. Applied Health Economics and Health Policy, 1-9.
28. Dolan, P., Lee, H., King, D., & Metcalfe, R. (2009). Valuing health directly. BMJ, 339.
29. Stiglitz, J. E., Sen, A., & Fitoussi, J. (2009). Report by the commission on the measurement of economic performance and social progress.
30. Helliwell, J. F., Layard, R., & Sachs, J. (2015). World happiness report 2013: Earth Institute, Columbia University.
31. Veenhoven, R. (1991). Is happiness relative?. Social Indicators Research, 24(1), 1-34.
32. Ferrer-i-Carbonell, A., & Frijters, P. (2004). How Important is Methodology for the estimates of the determinants of Happiness?*. The Economic Journal, 114(497), 641-659.
33. Easterlin, R. A. (2005). Is There an 'Iron Law of Happiness?'.
34. Lounsbury, J. W., Saudargas, R. A., Gibson, L. W., & Leong, F. T. (2005). An investigation of broad and narrow personality traits in relation to general and domain-specific life satisfaction of college students. Research in Higher Education, 46(6), 707-729.
35. Lu, L., & Hu, C. (2005). Personality, leisure experiences and happiness. Journal of Happiness Studies, 6(3), 325-342.
36. Dolan, P., Peasgood, T., & White, M. (2008). Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. Journal of economic psychology, 29(1), 94-122.
37. Garrido, S., Méndez, I., & Abellán, J. (2013). Analysing the simultaneous relationship between life satisfaction and health-related quality of life. Journal of Happiness Studies, 14(6), 1813-1838.
38. Wagstaff, A. (1991). QALYs and the equity-efficiency trade-off. Journal of health economics, 10(1), 21-41.

39. Schreyögg, J., Stargardt, T., Velasco-Garrido, M., & Busse, R. (2005). Defining the "Health Benefit Basket" in nine European countries. The European Journal of Health Economics, 6, 2-10.
40. Pronk, M. H., & Bonsel, G. J. (2004). Out-patient drug policy by clinical assessment rather than financial constraints?. The European Journal of Health Economics, 5(3), 274-277.
41. Dakin, H. A., Devlin, N. J., & Odeyemi, I. A. O. (2006). "Yes","No" or "Yes, but"? Multinomial modelling of NICE decision-making. Health policy, 77(3), 352-367.
42. Devlin, N., & Parkin, D. (2004). Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. Health economics, 13(5), 437-452.
43. George, B., Harris, A., & Mitchell, A. (2001). Cost-effectiveness analysis and the consistency of decision making: evidence from pharmaceutical reimbursement in Australia (1991 to 1996). PharmacoEconomics, 19(11), 1103-1109.
44. Rutten, F., & Van Busschbach, J. (2001). How to define a basic package of health services for a tax funded or social insurance based health care system?. HEPAC Health Economics in Prevention and Care, 2(2), 45-46.
45. Cookson, R., McCabe, C., & Tsuchiya, A. (2008). Public healthcare resource allocation and the Rule of Rescue. British medical journal, 34(7), 540.
46. Appleby, J., Devlin, N., Parkin, D., Buxton, M., & Chalkidou, K. (2009). Searching for cost effectiveness thresholds in the NHS. Health policy, 91(3), 239-245.
47. NICE. (2009). Appraising life-extending, end of life treatments. London: National Institute for Health and Clinical Excellence.
48. Dunning, A. J. (1991). Kiezen en delen. Advies in hoofdzaken van de commissie Keuzen in de zorg. Den Haag: Albani.
49. RVZ. (2006). Zinnige en Duurzame Zorg.
50. Bleichrodt, H., & Quiggin, J. (1999). Life-cycle preferences over consumption and health: when is cost-effectiveness analysis equivalent to cost-benefit analysis?. Journal of health economics, 18(6), 681-708.
51. Dolan, P., & Edlin, R. (2002). Is it really possible to build a bridge between cost-benefit analysis and cost-effectiveness analysis?. Journal of health economics, 21(5), 827-843.
52. Williams, A., & Cookson, R. (2000). Equity in health. Handbook of health economics, 1, 1863-1910.
53. Sen, A. (1995). Inequality reexamined: Oxford University Press.
54. McKie, J., & Richardson, J. (2003). The rule of rescue. Social science & medicine, 56(12), 2407-2419.
55. Hadorn, D. C. (1991). Setting health care priorities in Oregon. Jama, 265(17), 2218-2225.
56. Nord, E. (2005). Concerns for the worse off: fair innings versus severity. Social science & medicine, 60(2), 257-263.
57. Oliver, A. (2009). A Fair Test of the Fair Innings?. Medical Decision Making, 29(4), 491.
58. Cookson, R., & Dolan, P. (2000). Principles of justice in health care rationing. Journal of medical ethics, 26(5), 323-329.
59. Boer, B. (2002). Onderzoek op maat: een verkenning van factoren voor het gebruik van Medical Technology Assessment.
60. Zwaap, J., Mastenbroek, C. G., & van der Heiden, L. A. (2009). Pakketbeheer in de Praktijk 2.

61. Brouwer, W. (2009). De basis van het pakket: Urgente uitdagingen voor de opzet en inzet van economische evaluaties in de zorg.
62. Stolk, E. A., van Donselaar, G., Brouwer, W. B. F., & Busschbach, J. J. V. (2004). Reconciliation of Economic Concerns and Health Policy: Illustration of an Equity Adjustment Procedure Using Proportional Shortfall. PharmacoEconomics, 22(17), 1097.
63. Stolk, E. A. (2009). Uitwerking van het pakketprincipe noodzakelijkheid; dimensie Ziektelast.
64. Johannesson, M. (2001). Should we aggregate relative or absolute changes in QALYs?. Health economics, 10(7), 573-577.
65. Green, C., & Gerard, K. (2009). Exploring the social value of health-care interventions: a stated preference discrete choice experiment. Health Economics, 18(8), 951-976.
66. Van de Wetering, E., Stolk, E., Van Exel, N., & Brouwer, W. (2013). Balancing equity and efficiency in the Dutch basic benefits package using the principle of proportional shortfall. The European Journal of Health Economics, 14(1), 107-115.
67. Bobinac, A., van Exel, N., Rutten, F. F. H., & Brouwer, W. B. F. (2012). Inquiry into the Relationship between Equity Weights and the Value of the QALY. Value in Health, 15(8), 1119-1126.
68. Claxton, K., Paulden, M., Gravelle, H., Brouwer, W., & Culyer, A. J. (2011). Discounting and decision making in the economic evaluation of health-care technologies. Health Economics, 20(1), 2-15.
69. Nord, E., Pinto, J. L., Richardson, J., Menzel, P., & Ubel, P. (1999). Incorporating societal concerns for fairness in numerical valuations of health programmes. Health Economics, 8(1), 25-39.
70. Lancsar, E., Wildman, J., Donaldson, C., Ryan, M., & Baker, R. (2011). Deriving distributional weights for QALYs through discrete choice experiments. Journal of health economics, 30(2), 466-478.
71. Winkelhage, J., & Diederich, A. (2012). The relevance of personal characteristics in allocating health care resources—controversial preferences of laypersons with different educational backgrounds. International journal of environmental research and public health, 9(1), 223-243.
72. Norman, R., Hall, J., Street, D., & Viney, R. (2012). Efficiency and equity: a stated preference approach. Health Economics, 22(5), 568-581.
73. Hensher, D. A., Rose, J. M., & Greene, W. H. (2005). Applied choice analysis: a primer: Cambridge Univ Pr.
74. Bliemer, M. C. J., Rose, J. M., & Hensher, D. A. (2009). Efficient stated choice experiments for estimating nested logit models. Transportation Research Part B: Methodological, 43(1), 19-35.
75. Carrasco, J. A., & de Dios Ortúzar, J. (2002). Review and assessment of the nested logit model. Transport Reviews, 22(2), 197-218.
76. Hensher, D. A., & Greene, W. H. (2002). Specification and estimation of the nested logit model: alternative normalisations. Transportation Research Part B: Methodological, 36(1), 1-17.
77. Cropper, M. L., Aydede, S. K., & Portney, P. R. (1994). Preferences for life saving programs: how the public discounts time and age. Journal of Risk and Uncertainty, 8(3), 243-265.
78. Johannesson, M., & Johansson, P. O. (1996). The economics of ageing: on the attitude of Swedish people to the distribution of health care resources between the young and the old. Health Policy, 37(3), 153-161.
79. Diederich, A., Winkelhage, J., & Wirsik, N. (2011). Age as a criterion for setting priorities in health care? A survey of the German public view. PloS one, 6(8), e23930.

80. Drummond, M. F., Wilson, D. A., Kanavos, P., Ubel, P., & Rovira, J. (2007). Assessing the economic challenges posed by orphan drugs. International Journal of Technology Assessment in Health Care, 23(1), 36-42.

81. McCabe, C., Claxton, K., & Tsuchiya, A. (2005). Orphan drugs and the NHS: should we value rarity?. BMJ, 331(7523), 1016.

82. Dellaert, B. G. C., Donkers, B., & Soest, A. V. (2012). Complexity Effects in Choice Experiment-Based Models. Journal of Marketing Research, 49(3), 424-434.

83. Hensher, D. A. (1994). Stated preference analysis of travel choices: the state of practice. Transportation, 21(2), 107-133.

84. Eichler, H., Kong, S. X., Gerth, W. C., Mavros, P., & Jönsson, B. (2004). Use of cost-effectiveness analysis in health-care resource allocation decision-making: how are cost-effectiveness thresholds expected to emerge?. Value in health, 7(5), 518-528.

85. Bobinac, A., van Exel, N., Rutten, F. F., & Brouwer, W. B. (2012). Inquiry into the Relationship between Equity Weights and the Value of the QALY. Value in Health.

86. Claxton, K., Walker, S., Palmer, S., & Sculpher, M. (2010). Appropriate perspectives for health care decisions.

87. Culyer, A., McCabe, C., Briggs, A., Claxton, K., Buxton, M., Akehurst, R., et al. (2007). Searching for a threshold, not setting one: the role of the National Institute for Health and Clinical Excellence. Journal of health services research & policy, 12(1), 56-58.

88. Green, C., & Gerard, K. (2009). Exploring the social value of health-care interventions: a stated preference discrete choice experiment. Health Economics, 18(8), 951-976.

89. Olsen, J. A. (2013). Priority Preferences:"< i> End of Life" Does Not Matter, But< i> Total Life Does. Value in Health, 16(6), 1063-1066.

90. Brouwer, W., van Exel, J., Baker, R., & Donaldson, C. (2008). The New Myth. PharmacoEconomics, 26(1), 1-4.

91. van Exel, J., Baker, R., Mason, H., Donaldson, C., Brouwer, W., & Team, E. (2015). Public views on principles for health care priority setting: Findings of a European cross-country study using Q methodology. Social science & medicine, 126, 128-137.

92. Rawlins, M., Barnett, D., & Stevens, A. (2010). Pharmacoeconomics: NICE's approach to decision-making. British journal of clinical pharmacology, 70(3), 346-349.

93. Nord, E., & Johansen, R. (2014). Concerns for severity in priority setting in health care: A review of trade-off data in preference studies and implications for societal willingness to pay for a QALY. Health Policy, 116(2), 281-288.

94. Bobinac, A., Van Exel, N., Rutten, F. F., & Brouwer, W. B. (2010). Willingness to Pay for a Quality-Adjusted Life-Year: The Individual Perspective. Value in Health, 13(8), 1046-1055.

95. Tsuchiya, A. (1999). Age-related preferences and age weighting health benefits. Social science & medicine, 48(2), 267-276.

96. Whitty, J. A., Rundle-Thiele, S. R., & Scuffham, P. A. (2008). Insights into public preferences for pharmaceutical funding. Marketing, 2(3), 216-234.

97. Van de Wetering, E., Van Exel, N., Rose, J. M., & Brouwer, W. (2014). Are some QALYs more equal than others?. European Journal of Health Economics, 1-11.

98. Hensher, D. A. (2010). Hypothetical bias, choice experiments and willingness to pay. Transportation Research Part B: Methodological, 44(6), 735-752.

99. Bliemer, M. C. J., Rose, J. M., & Hensher, D. A. (2009). Efficient stated choice experiments for estimating nested logit models. Transportation Research Part B: Methodological, 43(1), 19-35.

100. Hole, A. R. (2008). Modelling heterogeneity in patients' preferences for the attributes of a general practitioner appointment. Journal of health economics, 27(4), 1078-1094.

101. Hole, A. R., & Kolstad, J. R. (2012). Mixed logit estimation of willingness to pay distributions: a comparison of models in preference and WTP space using data from a health-related choice experiment. Empirical Economics, 42(2), 445-469.

102. de Bekker-Grob, E. W., Ryan, M., & Gerard, K. (2012). Discrete choice experiments in health economics: a review of the literature. Health Economics, 21(2), 145-172.

103. Skedgel, C., Wailoo, A., & Akehurst, R. (2015). Societal preferences for distributive justice in the allocation of health care resources: a latent class discrete choice experiment. Medical decision making : an international journal of the Society for Medical Decision Making, 35(1), 94-105.

104. Mentzakis, E., Ryan, M., & McNamee, P. (2011). Using discrete choice experiments to value informal care tasks: exploring preference heterogeneity. Health Economics, 20(8), 930-944.

105. Dolan, P., & Tsuchiya, A. (2005). Health priorities and public preferences: the relative importance of past health experience and future health prospects. Journal of health economics, 24(4), 703-714.

106. Shah, K. K., Tsuchiya, A., & Wailoo, A. J. (2015). Valuing health at the end of life: A stated preference discrete choice experiment. Social science & medicine, 124, 48-56.

107. Brazier, J., Rowen, D., Mukuria, C., Whyte, S., Keetharuth, A., Rise, A., et al. (2013). Eliciting societal preferences for burden of illness, therapeutic improvement and end of life for value based pricing: a report of the main survey. EEPRU Research Report, 01/13.

108. Diederich, A., Winkelhage, J., & Wirsik, N. (2011). Age as a criterion for setting priorities in health care? A survey of the German public view. PloS one, 6(8), e23930.

109. Drummond, M. (2012). Twenty Years of Using Economic Evaluations for Reimbursement Decisions.What Have We Achieved?. CHE Research paper.

110. Norheim, O. F., Baltussen, R., Johri, M., Chisholm, D., Nord, E., Brock, D., et al. (2014). Guidance on priority setting in health care (GPS-Health): the inclusion of equity criteria not captured by cost-effectiveness analysis. Cost effectiveness and resource allocation : C/E, 12, 18-7547-12-18. eCollection 2014.

111. Ratcliffe, J., Buxton, M., Young, T., & Longworth, L. (2005). Determining priority for liver transplantation. Applied health economics and health policy, 4(4), 249-255.

112. Schwappach, D. L. B. (2003). Does it matter who you are or what you gain? An experimental study of preferences for resource allocation. Health economics, 12(4), 255-267.

113. Ratcliffe, J. (2000). Public preferences for the allocation of donor liver grafts for transplantation. Health Economics, 9(2), 137-148.

114. Ottersen, T. (2013). Lifetime QALY prioritarianism in priority setting. Journal of medical ethics, 39(3), 175-180.

115. Tsuchiya, A., Dolan, P., & Shaw, R. (2003). Measuring people's preferences regarding ageism in health: some methodological issues and some fresh evidence. Social science & medicine, 57(4), 687-696.

116. Edlin, R., Round, J., McCabe, C., Sculpher, M., Claxton, K., & Cookson, R. (2008). Cost-effectiveness analysis and ageism: a review of the theoretical literature. Leeds Institute of Health Sciences.

117. Nord, E. (1993). The relevance of health state after treatment in prioritising between different patients. Journal of medical ethics, 19(1), 37-42.

118. Nord, E. (1993). The trade-off between severity of illness and treatment effect in cost-value analysis of health care. Health Policy, 24(3), 227-238.

119. Murray, C. J., & Lopez, A. D. (1994). Global comparative assessments in the health sector: disease burden, expenditures and intervention packages. Geneva: World Health Organization.

120. Ubel, P. A. (1999). How stable are people's preferences for giving priority to severely ill patients?. Social science & medicine, 49(7), 895-903.

121. Bowling, A., Mariotto, A., & Evans, O. (2002). Are older people willing to give up their place in the queue for cardiac surgery to a younger person?. Age and Ageing, 31(3), 187-192.

122. Claxton, K., Martin, S., Soares, M., Rice, N., Spackman, E., Hinde, S., et al. (2015). Systematic review of the literature on the cost-effectiveness threshold. Health Technology Assessment(No. 19.14).

123. Miners, A., Cairns, J., & Wailoo, A. (2013). Burden of illnes into value based pricing: Adescription and critique. NICE Decision Support Unit, University of Sheffield.

124. Olsen, J. A., & Donaldson, C. (1998). Helicopters, hearts and hips: using willingness to pay to set priorities for public sector health care programmes. Social science & medicine, 46(1), 1-12.

125. Shiroiwa, T., Sung, Y., Fukuda, T., Lang, H., Bae, S., & Tsutani, K. (2010). International survey on willingness-to-pay (WTP) for one additional QALY gained: what is the threshold of cost effectiveness?. Health Econ, 19, 422-437.

126. Baker, R., Currie, G. R., & Donaldson, C. (2010). What needs to be done in contingent valuation: have Smith and Sach missed the boat?. Health Economics, Policy and Law, 5(01), 113-121.

127. Nord, E., Street, A., Richardson, J., Kuhse, H., & Singer, P. (1996). The significance of age and duration of effect in social evaluation of health care. Health Care Analysis, 4(2), 103-111.

128. Nord, E., & Johansen, R. (2015). Transforming EQ-5D utilities for use in cost–value analysis of health programs. The European Journal of Health Economics, 16(3), 313-328.

129. Petrou, S., Kandala, N., Robinson, A., & Baker, R. (2013). A person trade-off study to estimate age-related weights for health gains in economic evaluation. PharmacoEconomics, 31(10), 893-907.

130. Rowen, D., Brazier, J., Mukuria, C., Keetharuth, A., Hole, A. R., Tsuchiya, A., et al. (2014). Update: Eliciting societal preferences for weighting QALYs according to burden of illness, size of gain and end of life.Policy Research Unit in Economic Evaluation of Health & Care Interventions (EEPRU) Research Report.

131. van de Wetering, L., van Exel, J., Bobinac, A., & Brouwer, W. B. (2015). Valuing QALYs in Relation to Equity Considerations Using a Discrete Choice Experiment. PharmacoEconomics, 1-12.

132. Enthoven, A., & van de Ven, W. (2007). Going Dutch—managed-competition health insurance in the Netherlands. New England Journal of Medicine, 357(24), 2421-2423.

133. Bergstrom, T. C. (2006). Benefit-cost in a benevolent society. The American Economic Review, 339-351.

134. Bhatia, M., & Fox-Rushby, J. (2003). Validity of willingness to pay: hypothetical versus actual payment. Applied Economics Letters, 10(12), 737-740.

135. CVZ. (2006). Richtlijnen voor farmaco-economisch onderzoek, geactualiseerde versie, Diemen: College voor Zorgverzekeringen.

136. Treasury, H. M. (2003). The green book. Appraisal and evaluation in central government.

137. CBS. (2015). Statistics Netherlands: http://statline.cbs.nl/Statweb/?LA=en.

138. Johri, M., Damschroder, L. J., Zikmund-Fisher, B. J., & Ubel, P. A. (2005). The importance of age in allocating health care resources: does intervention-type matter?. Health Economics, 14(7), 669-678.

139. Mason, H., Baker, R., & Donaldson, C. (2008). Willingness to pay for a QALY: past, present and future. Expert Review of Pharmacoeconomics & Outcomes Research, 8(6), 575-582.

140. Pennington, M., Baker, R., Brouwer, W., Mason, H., Hansen, D. G., Robinson, A., et al. (2015). Comparing WTP values of different types of QALY gain elicited from the general public. Health Economics, 24(3), 280-293.

141. Linley, W. G., & Hughes, D. A. (2013). Societal Views On Nice, Cancer Drugs Fund And Value-Based Pricing Criteria For Prioritising Medicines: A Cross-Sectional Survey Of 4118 Adults In Great Britain. Health Economics, 22(8), 948-964.

142. Gyrd-Hansen, D. (2003). Willingness to pay for a QALY. Health Economics, 12(12), 1049-1060.

143. Ryen, L., & Svensson, M. (2014). The willingness to pay for a quality adjusted life year: a review of the empirical literature. Health Economics, 24(10), 1289-1301.

144. Bobinac, A., Van Exel, N., Rutten, F. F., & Brouwer, W. B. (2010). Willingness to pay for a quality-adjusted life-year: the individual perspective. Value in Health, 13(8), 1046-1055.

145. Tunis, S., & Whicher, D. (2009). The National Oncologic PET Registry: lessons learned for coverage with evidence development. Journal of the American College of Radiology, 6(5), 360-365.

146. Walker, S., Sculpher, M., Claxton, K., & Palmer, S. (2012). Coverage with evidence development, only in research, risk sharing, or patient access scheme? A framework for coverage decisions. Value in Health, 15(3), 570-579.

147. Trueman, P., Grainger, D. L., & Downs, K. E. (2010). Coverage with evidence development: applications and issues. International Journal of Technology Assessment in Health Care, 26(01), 79-85.

148. CVZ. (2012). Conditional Reimbursement of health care.

149. Franken, M., le Polain, M., Cleemput, I., & Koopmanschap, M. (2012). Similarities and differences between five European drug reimbursement systems. International Journal of Technology Assessment in Health Care, 28(04), 349-357.

150. Niezen, M., de Bont, A., Stolk, E., Eyck, A., Niessen, L., & Stoevelaar, H. (2007). Conditional reimbursement within the Dutch drug policy. Health policy, 84(1), 39-50.

151. Franken, M., Koopmanschap, M., & Steenhoek, A. (2014). Health economic evaluations in reimbursement decision making in the Netherlands: Time to take it seriously?. Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen, 108(7), 383-389.

152. Mortimer, D., Li, J. J., Watts, J., & Harris, A. (2011). Breaking up is hard to do: The economic impact of provisional funding contingent upon evidence development. Health Economics, Policy and Law, 6(04), 509-527.

153. Menon, D., Stafinski, T., Nardelli, A., Jackson, T., & Jhamandas, J. (2011). Access with evidence development: an approach to introducing promising new technologies into healthcare. Healthcare

management forum / Canadian College of Health Service Executives = Forum gestion des soins de sante / College canadien des directeurs de services de sante, 24(2), 42-56.

154. Stafinski, M. T., McCabe, C. J., & Menon, D. (2010). Funding the unfundable. PharmacoEconomics, 28(2), 113-142.

155. Bleichrodt, H., Pinto, J. L., & Wakker, P. P. (2001). Making descriptive use of prospect theory to improve the prescriptive use of expected utility. Management Science, 47(11), 1498-1514.

156. Koopmanschap, M. A., Stolk, E. A., & Koolman, X. (2010). Dear policy maker: have you made up your mind? A discrete choice experiment among policy makers and other health professionals. International Journal of Technology Assessment in Health Care, 26(02), 198-204.

157. Regier, D. A., Ryan, M., Phimister, E., & Marra, C. A. (2009). Bayesian and classical estimation of mixed logit: an application to genetic testing. Journal of health economics, 28(3), 598-610.

158. Lancsar, E., & Savage, E. (2004). Deriving welfare measures from discrete choice experiments: inconsistency between current methods and random utility and welfare theory. Notice of Re-printing in Health Economics.

159. Martinez-Cruz, A. L. (2013). Implications of heterogeneity in discrete choice analysis: UNIVERSITY OF MARYLAND, COLLEGE PARK.

160. Lancsar, E. J., Hall, J. P., King, M., Kenny, P., Louviere, J. J., Fiebig, D. G., et al. (2007). Using discrete choice experiments to investigate subject preferences for preventive asthma medication. Respirology, 12(1), 127-136.

161. Train, K. E. (2009). Discrete choice methods with simulation: Cambridge university press.

162. Spinks, J., & Mortimer, D. (2015). The effect of traffic lights and regulatory statements on the choice between complementary and conventional medicines in Australia: Results from a discrete choice experiment. Social science & medicine, 124, 257-265.

163. Skedgel, C., Wailoo, A., & Akehurst, R. (2014). Societal Preferences for Distributive Justice in the Allocation of Health Care Resources: A Latent Class Discrete Choice Experiment. Medical decision making : an international journal of the Society for Medical Decision Making.

164. McHugh, N., Baker, R. M., Mason, H., Williamson, L., van Exel, J., Deogaonkar, R., et al. (2015). Extending life for people with a terminal illness: a moral right and an expensive death? Exploring societal perspectives. BMC medical ethics, 16(1), 14.

165. Van Exel, N., de Graaf, G., & Brouwer, W. B. (2006). "Everyone dies, so you might as well have fun!" Attitudes of Dutch youths about their health lifestyle. Social science & medicine, 63(10), 2628-2639.

166. Coast, J., Flynn, T. N., Natarajan, L., Sproston, K., Lewis, J., Louviere, J. J., et al. (2008). Valuing the ICECAP capability index for older people. Social science & medicine, 67(5), 874-882.

167. Grewal, I., Lewis, J., Flynn, T., Brown, J., Bond, J., & Coast, J. (2006). Developing attributes for a generic quality of life measure for older people: Preferences or capabilities?. Social science & medicine, 62(8), 1891-1901.

168. Netten, A., Burge, P., Malley, J., Potoglou, D., Towers, A., Brazier, J., et al. (2012). Outcomes of social care for adults: developing a preference-weighted measure. Health technology assessment, 16(16).

169. van Leeuwen, K. M., Jansen, A. P., Muntinga, M. E., Bosmans, J. E., Westerman, M. J., van Tulder, M. W., et al. (2015). Exploration of the content validity and feasibility of the EQ-5D-3L, ICECAP-O and ASCOT in older adults. BMC health services research, 15(1), 201.

170. Versteegh, M., & Brouwer, W. B. F. (2011). The royal road or the middle way? Patient and general public preferences for health outcomes. Value in Health, 14(7).

171. Cohen, G. A. (1993). Equality of what? On welfare, goods and capabilities. In A. K. S. M.C. Nussbaum (Ed.), The Quality of Life: Clarendon Press, Oxford.

172. Kahneman, D., & Sugden, R. (2005). Experienced utility as a standard of policy evaluation. Environmental and resource economics, 32(1), 161-181.

173. Layard, R. (2006). Happiness: Lessons from a new science: Penguin UK.

174. Dolan, P., & Metcalfe, R. (2012). Measuring subjective wellbeing: Recommendations on measures for use by national governments. Journal of social policy, 41(02), 409-427.

175. Ng, Y. (1996). Happiness surveys: Some comparability issues and an exploratory survey based on just perceivable increments. Social Indicators Research, 38(1), 1-27.

176. Layard, R. (2006). Happiness and public policy: a challenge to the profession*. The Economic Journal, 116(510), C24-C33.

177. Dolan, P., & Kahneman, D. (2008). Interpretations of Utility and Their Implications for the Valuation of Health*. The Economic Journal, 118(525), 215-234.

178. Bruni, L. (2004). The``Technology of Happiness"and the Tradition of Economic Science. Journal of the History of Economic Thought, 26(1), 19-43.

179. Robinson, J. (1962). Economic philosophy: Transaction Publishers.

180. Diener, E., Suh, E., Lucas, R., & Smith, H. (1999). Subjective well-being: Three decades of progress—1967 to 1997. Psychological bulletin, 125(2), 276-302.

181. Ferrer-i-Carbonell, A. (2005). Income and well-being: an empirical analysis of the comparison income effect. Journal of Public Economics, 89(5), 997-1019.

182. Graham, C. (2005). The economics of happiness. World economics, 6(3), 41-55.

183. Kimball, M., & Willis, R. (2006). Utility and happiness. University of Michigan.

184. Frey, B. S., & Stutzer, A. (2002). What can economists learn from happiness research?. Journal of Economic literature, 402-435.

185. Frey, B. S., & Stutzer, A. (2003). Testing theories of happiness. Zurich IEER Working Paper.

186. Powdthavee, N. (2007). Economics of happiness: A review of literature and applications. Chulalongkorn Journal of Economics, 19(1), 51-73.

187. Mahon, N. E., Yarcheski, A., & Yarcheski, T. J. (2005). Happiness as related to gender and health in early adolescents. Clinical nursing research, 14(2), 175-190.

188. Huebner, E. S., Drane, W., & Valois, R. F. (2000). Levels and demographic correlates of adolescent life satisfaction reports. School Psychology International, 21(3), 281-292.

189. Huebner, E. S., Valois, R. F., Paxton, R. J., & Drane, J. W. (2005). Middle school students' perceptions of quality of life. Journal of Happiness Studies, 6(1), 15-24.

190. Gilman, R., & Huebner, S. (2003). A review of life satisfaction research with children and adolescents. School Psychology Quarterly, 18(2), 192-205.

191. Csikszentmihalyi, M., & Hunter, J. (2003). Happiness in everyday life: The uses of experience sampling. Journal of Happiness Studies, 4(2), 185-199.

192. Crossley, A., & Langdridge, D. (2005). Perceived sources of happiness: A network analysis. Journal of Happiness Studies, 6(2), 107-135.

193. UNICEF. (2007). Child poverty in perspective: An overview of child well-being in rich countries: A comprehensive assessment of the lives and well-being of children and adolescents in the economically advanced nations: UNICEF Innocenti Research Centre.

194. Lewis, C. A., Lanigan, C., Joseph, S., & De Fockert, J. (1997). Religiosity and happiness: No evidence for an association among undergraduates. Personality and Individual Differences, 22(1), 119-121.

195. Park, N. (2004). The role of subjective well-being in positive youth development. The Annals of the American Academy of Political and Social Science, 591(1), 25-39.

196. Suldo, S. M., & Huebner, E. S. (2004). The role of life satisfaction in the relationship between authoritative parenting dimensions and adolescent problem behavior. Quality-of-Life Research on Children and Adolescents, 23, 165-195.

197. Veenhoven, R., & Verkuyten, M. (1989). The well-being of only children. Adolescence: an international quarterly devoted to the physiological, psychological, psychiatric, sociological, and educational aspects of the second decade of human life, 24(93), 155-166.

198. Cheng, H., & Furnham, A. (2002). Personality, peer relations, and self-confidence as predictors of happiness and loneliness. Journal of adolescence, 25(3), 327-339.

199. Van Exel, N., Koolman, X., De Graaf, G., & Brouwer, W. (2005). Overweight and obesity in Dutch adolescents: Associations with health lifestyle, personality, social context and future consequences: methods & tables. Rotterdam, The Netherlands: Institute for Medical Technology Assessment (iMTA), 93.

200. Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. Psychological assessment, 4(1), 26-42.

201. Gerris, J. R. M. (1998). Parents, Adolescents and Young Adults in Dutch Families: A Longitudinal Study: a Description of Dutch Family Life in Terms of Validated Concepts Representing Family Living Conditions, Cultural Activities, Pubertal Timing, Child-rearing, Interpersonal Relations and Communication, Personality Dimensions, Adolescent Behavior, Marital Relations, Job Experiences, Household Tasks and Socio-cultural Value Orientations: Institute of Family Studies, University of Nijmegen.

202. de Bruijn, G. J., Kremers, S. P., van Mechelen, W., & Brug, J. (2005). Is personality related to fruit and vegetable intake and physical activity in adolescents?. Health education research, 20(6), 635-644.

203. Frey, B. S., & Stutzer, A. (2000). Happiness, economy and institutions. The Economic Journal, 110(466), 918-938.

204. Stutzer, A., & Frey, B. S. (2004). Reported subjective well-being: A challenge for economic theory and economic policy. Schmollers Jahrbuch, 124(2), 1-41.

205. Pindyck, R. S., & Rubinfeld, D. L. (1998). Econometric models and economic forecasts: Irwin/ McGraw-Hill Boston.

206. Gilman, R., Huebner, E. S., & Laughlin, J. E. (2000). A first study of the Multidimensional Students' Life Satisfaction Scale with adolescents. Social Indicators Research, 52(2), 135-160.

207. Huebner, E. S., Valois, R. F., Paxton, R. J., & Drane, J. W. (2005). Middle school students' perceptions of quality of life. Journal of Happiness Studies, 6(1), 15-24.

208. Demir, M., Özdemir, M., & Weitekamp, L. A. (2007). Looking to happy tomorrows with friends: Best and close friendships as they predict happiness. Journal of Happiness Studies, 8(2), 243-271.

209. Graham, C. (2005). Insights on development from the economics of happiness. The World Bank Research Observer, 20(2), 201-231.

210. Diener, E., & Seligman, M. E. (2004). Beyond money toward an economy of well-being. Psychological science in the public interest, 5(1), 1-31.

211. Van de Wetering, E., van Exel, N., & Brouwer, W. (2010). Piecing the Jigsaw Puzzle of Adolescent Happiness. Journal of Economic Psychology, 31(6), 923-935.

212. Johnston, D. W., Propper, C., & Shields, M. A. (2009). Comparing subjective and objective measures of health: Evidence from hypertension for the income/health gradient. Journal of health economics, 28(3), 540-552.

213. Borsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., et al. (2013). Data Resource Profile: the Survey of Health, Ageing and Retirement in Europe (SHARE). International journal of epidemiology, 42(4), 992-1001.

214. Ferrer-i-Carbonell, A., & Frijters, P. (2004). How Important is Methodology for the estimates of the determinants of Happiness?*. The Economic Journal, 114(497), 641-659.

215. Cameron, A. C., & Trivedi, P. K. (2005.). Microeconometrics using stata: Stata Press College Station, TX.

216. Murray, M. P. (2006). Avoiding invalid instruments and coping with weak instruments. The journal of economic perspectives, 20(4), 111-132.

217. Böckerman, P., Johansson, E., & Saarni, S. I. (2011). Do established health-related quality-of-life measures adequately capture the impact of chronic conditions on subjective well-being?. Health Policy, 100(1), 91-95.

218. Stock, J. H., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. Identification and inference for econometric models: Essays in honor of Thomas Rothenberg.

219. Murray, M. P. (2006). Avoiding invalid instruments and coping with weak instruments. The journal of economic perspectives, 20(4), 111-132.

220. Krueger, A. B., & Schkade, D. A. (2008). The reliability of subjective well-being measures. Journal of public economics, 92(8), 1833-1845.

221. Hyde, M., Wiggins, R. D., Higgs, P., & Blane, D. B. (2003). A measure of quality of life in early old age: the theory, development and properties of a needs satisfaction model (CASP-19). Aging & mental health, 7(3), 186-194.

222. Wiggins, R. D., Netuveli, G., Hyde, M., Higgs, P., & Blane, D. (2008). The evaluation of a self-enumerated scale of quality of life (CASP-19) in the context of research on ageing: A combination of exploratory and confirmatory approaches. Social Indicators Research, 89(1), 61-77.

223. Group, T. E. (1990). EuroQol-a new facility for the measurement of health-related quality of life. Health policy, 16(3), 199-208.

224. Oppe, M., Devlin, N. J., & SZENDE, A. (2007). EQ-5D value sets: inventory, comparative review and user guide: Springer.

225. Beckett, M., Weinstein, M., Goldman, N., & Yu-Hsuan, L. (2000). Do health interview surveys yield reliable data on chronic illness among older respondents?. American Journal of Epidemiology, 151(3), 315-323.

226. Blanchflower, D. G., & Oswald, A. J. (2004). Money, sex and happiness: An empirical study. The Scandinavian Journal of Economics, 106(3), 393-415.

227. Uppal, S. (2006). Impact of the timing, type and severity of disability on the subjective well-being of individuals with disabilities. Social science & medicine, 63(2), 525-539.

228. Lee, P., Smith, J. P., & Kington, R. (1999). The relationship of self-rated vision and hearing to functional status and well-being among seniors 70 years and older. American Journal of Ophthalmology, 127(4), 447-452.

229. Margolis, M. K., Coyne, K., Kennedy-Martin, T., Baker, T., Schein, O., & Revicki, D. A. (2002). Vision-specific instruments for the assessment of health-related quality of life and visual functioning. PharmacoEconomics, 20(12), 791-812.

230. Turner, A., Williams, B., & Barlow, J. (2002). The impact of an arthritis self-management programme on psychosocial wellbeing. Health Education, 102(3), 95-105.

231. East, L., Brown, K., & Twells, C. (2004). 'Knocking at St Peter's door'. A qualitative study of recovery after a heart attack and the experience of cardiac rehabilitation. Primary health care research and development, 5(03), 202-210.

232. Hildingh, C., & Baigi, A. (2010). The association among hypertension and reduced psychological well-being, anxiety and sleep disturbances: a population study. Scandinavian Journal of Caring Sciences, 24(2), 366-371.

233. Blanchflower, D. G., & Oswald, A. J. (2008). Hypertension and happiness across nations. Journal of health economics, 27(2), 218-233.

234. Bjørnskov, C. (2008). Healthy and happy in Europe? On the association between happiness and life expectancy over time. Social science & medicine, 66(8), 1750-1759.

235. French, M. T., & Popovici, I. (2011). That instrument is lousy! In search of agreement when using instrumental variables estimation in substance use research. Health Economics, 20(2), 127-146.

236. Essers, B. A., van Helvoort-Postulart, D., Prins, M. H., Neumann, M., & Dirksen, C. D. (2010). Does the Inclusion of a Cost Attribute Result in Different Preferences for the Surgical Treatment of Primary Basal Cell Carcinoma?. PharmacoEconomics, 28(6), 507-520.

237. Richardson, J. R., McKie, J., Peacock, S. J., & Iezzi, A. (2011). Severity as an independent determinant of the social value of a health service. The European Journal of Health Economics, 12(2), 163-174.

238. Lancsar, E., & Louviere, J. (2008). Conducting Discrete Choice Experiments to Inform Healthcare Decision Making: A Users Guide. PharmacoEconomics, 26(8), 661-677.

239. Bobinac, A., van Exel, N Job A, Rutten, F. F., & Brouwer, W. B. (2012). GET MORE, PAY MORE? An elaborate test of construct validity of willingness to pay per QALY estimates obtained through contingent valuation. Journal of health economics, 31(1), 158-168.

240. CBS. (2015). Welzijn in Nederland 2015.

New expensive technologies and aging populations increase the pressure on health care budgets. To promote an efficient use of limited health care resources, priorities must be set and choices regarding access to health care become inevitable. Economic evaluations are increasingly used to inform such decisions. However, there seems to be a discrepancy between the recommendations from economic evaluations and public preferences for the allocation of health care resources. This thesis investigated two methodological issues related to the measurement and valuation of outcomes in economic evaluations. The first part of this thesis investigated concerns in society for the distribution of health and health care. The second part explored happiness as a broader outcome measure for economic evaluations.

This thesis contributes to the improvement of economic evaluations with the overarching aim to make economic evaluations a more useful source of information for health care decision making, better aligned with distributional preferences in society.