Equity and efficiency in health care priority setting: How to get the balance right?

Rechtvaardigheid en doelmatigheid bij prioritering in de gezondheidszorg: Hoe ontstaat de juiste balans?

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus
Prof.dr. S.W.J. Lamberts
en volgens besluit van het College voor
Promoties.
De openbare verdediging zal plaatsvinden op
vrijdag 2 september 2005 om 13.30 uur

door

Elisabeth Allegonda Stolk
geboren te Bunnik

Promotiecommissie

Promotor:
Prof.dr. F.F.H. Rutten

Overige leden:
Prof.dr. J.P. Mackenbach
Prof. A. Williams
Prof.dr. E.K.A. van Doorslaer

Copromotor:
Dr. J.J. van Busschbach

ISBN: 90-5335-061-6
Financial support by Janssen-Cilag & Pfizer Netherlands bv.
Design, typesetting, and dtp of cover and interior in Scala Sans using Adobe InDesign 1.5 by Pier.

Printed in The Netherlands on recycled paper by Ridderprint offsetdrukkerij bv.

Cover: Damien Hirst, Gold Pharmacy wallpaper

Con.

Contents

Ac.

Acknowledgements

Chapter 2
Cost effectiveness of sildenafil

We thank L Damen and P Rabsztyn, who provided us with details about the treatment scenarios. L van der Hell assisted with the interviews. The research was supported by an unrestricted grant from Pfizer BV in the Netherlands.

Chapter 3
Analysis of the funding debate

We would like to thank an anonymous referee for his valuable comments on an earlier version of this paper.

Chapter 4
Illustration of equity adjustment

We thank Alan Williams and Han Bleichrodt for their helpful comments, and we thank Stefan Pickee and the researchers who provided us with data for their assistance in retrieving the QALY profiles of different patient groups. We acknowledge financial support from the Netherlands Organisation for Scientific Research (NWO) under Project DO 945-10-034 and from the Netherlands Organisation for Health Research and Development (ZONMW) under project VVS-2-7. A part of the empirical work was commissioned and funded by the Dutch Healthcare Insurance Board.

Chapter 5
Definition of equity

We acknowledge financial support from NWO under Project DO 945-10-034 and ZONMW under Project VVS-2-7.

Chapter 6
Elicitation of equity weights

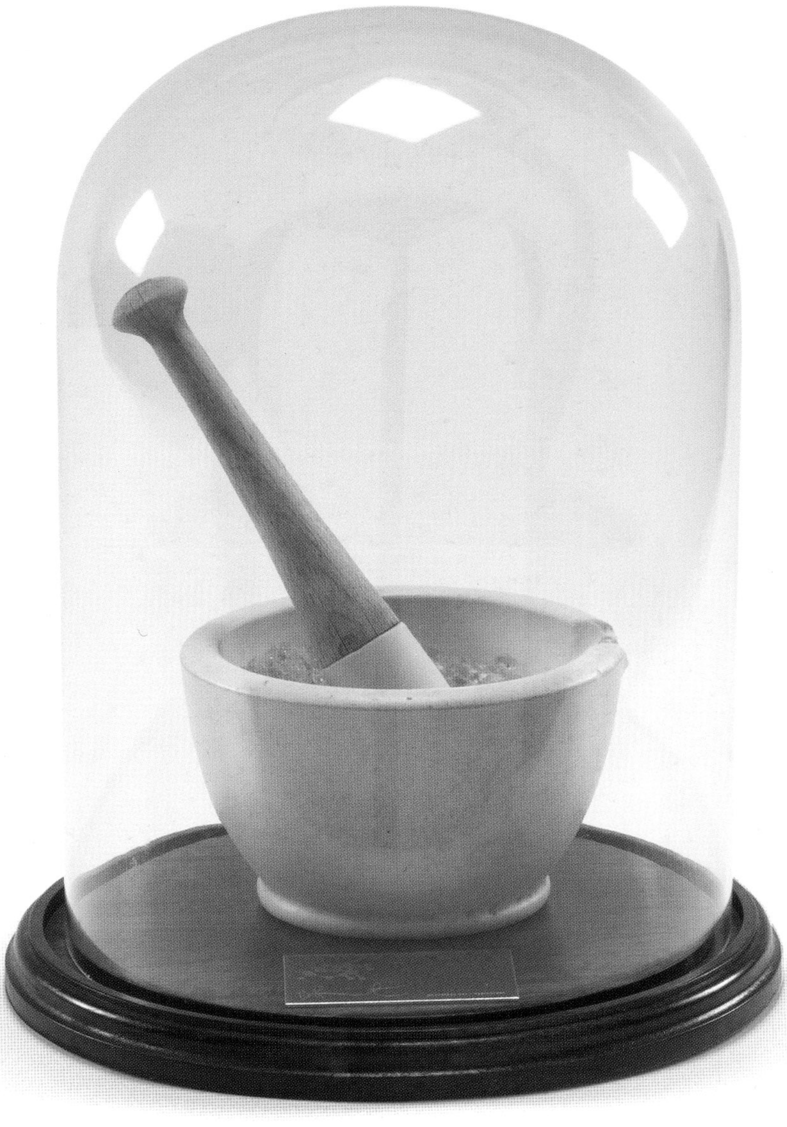
Peter Wakker, two anonymous reviewers and participants of the Workshop on QALYs in Alicante gave many helpful comments.
We acknowledge financial support from NWO under Project DO 945-10-034 and from ZONMW under Project VVS-2-7. Han Bleichrodt also acknowledges a VIDI-Grant from NWO.

Chapter 7
Equity in health policy

In preparing this paper we were helped by Frans Rutten, Jan van Busschbach, and Werner Brouwer, who kindly made useful comments on earlier drafts. We owe them for the way their arguments helped us to focus more sharply on the political and theoretical issues involved.

Damien Hirst: Ashes to ashes

Chapter 1
Introduction:
Ethics and economics, where can they meet?



Chapter 1

Introduction: Ethics and economics, where can they meet?

Based on: Stolk EA, Busschbach JJV. Economics and ethics in health care. Where can they meet? In: Gastmans C, Between technology and humanity. The impact of technology on health care ethics. Leuven: Leuven University press, 2002: Chapter 3, p. 49-66

1.1 Introduction

National health care systems are not just concerned with improving people's health but also with protecting patients against the financial costs of illness (WHO, 2000). In health care this means that cost are spread across members of society using pre-service payment methods like taxes or compulsory insurance, to guarantee equal access for all. But the question is: equal access to what services? Although some may actually feel inclined to fund any treatment that is medically possible, there are limits to public spending and the individual contributions. In these days, the aging population and the medical advances combined with the increased public interest in health and well-being that comes with a strong economy create a desire for more health care interventions than society may be prepared to pay for (Wetenschappelijke Raad voor het Regeringsbeleid, 1997). Already we see that spending on health care is outpacing economic growth in most countries, forcing governments to consider complicated matters of priority setting. What part of national budgets should be allocated to health care, and within health care, how should the available budget be allocated over the different available services if money is lacking to provide them all?

To address the problem that the demand for health care has surpassed the supply, economic evaluations of health care interventions have been introduced. The purpose of economic evaluations is to measure, value and compare the costs and consequences of the alternatives being considered for the allocation of resources, to find out which allocation of resources brings the greatest value for money (Drummond et al., 1997). This approach to priority setting is based on the utilitarian philosophy, a moral theory about social decision-making. This theory suggests that in a situation where different people in society have different interests, which cannot all be met (e.g. because different people have conflicting interests, or because resources are lacking), the social objective should be to maximize total utility from the available resources. The most straightforward way to apply the utilitarian philosophy in health care is to interpret health gains as a measure of utility. Accordingly, economic evaluations typically address the question how much health benefits are produced by different services, and at what costs. With this information it becomes possible to rank different options and to identify the alternatives that maximise the health outcomes at a given cost.

Economic evaluations help to identify what allocations of resources is most efficient. Evidently, an efficient use of resources is an important target in health care. This explains why an increasing number of countries intend to use economic evaluations in priority setting. Nevertheless, the actual effect of economic evaluations on priority setting has been modest. This becomes clear when the practice regarding the application of economic evaluations in health policy is studied. The relationship between resource allocation decisions and the outcomes of an economic evaluation is not always clear. Even in countries where economic evaluations are formally required to inform reimbursement decisions like in Australia and the United Kingdom, there is no clear negative relationship between the level of cost-effectiveness and a positive reimbursement (Devlin and Parkin, 2004; George, 2001). Sometimes, interventions with an unfavourable cost-effectiveness ratio are being reimbursed or vice versa.

Economic evaluations have been criticized for ignoring equity implications. A focus on efficiency assumes distributive neutrality, which implies that the value of a health gain can be determined irrespective of the patient to whom it accrues. However, society may also value the way in which health is distributed across society. It has become increasingly clear that members of a society feel that some patients - especially the worst off - have stronger moral claims on scarce health care resources than others. Apparently efficiency is not the only objective in health care; there is also a concern for equality in the distribution of health. This concern for equality reflects the desire to minimise differences in one or more aspects of health across populations or population groups. If then policy is only targeted at maximising the health outcomes of resources that are invested in health care, the pursuit of equality may be unfulfilled and the resulting allocation of resources is likely to be perceived as unfair. This in turn may explain why reimbursement decisions may be contrary to economic recommendations.

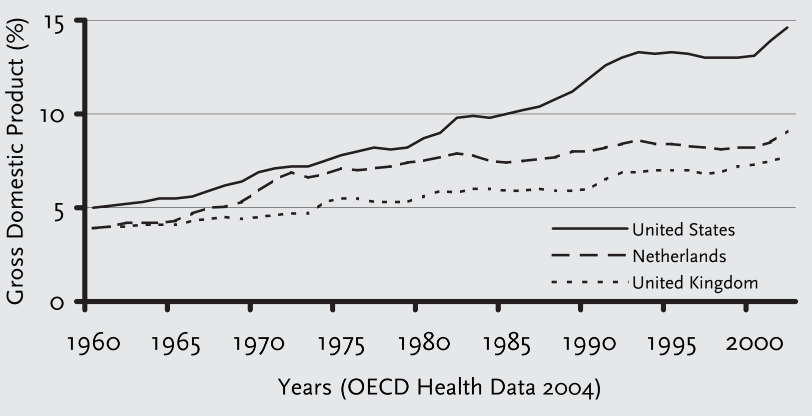
This thesis explores the assumption that the discrepancy between economic evaluation and health policy can be explained on the basis of concerns about fairness. Economic evaluations are currently primarily concerned with finding the allocation of resources that will maximize population health, but equity or fairness is a separate and important concern that health economists and other policy analysts should address. The single-mindedness of economic evaluations is the problem. To resolve this problem, this thesis explores how equity concerns can be addressed in economic evaluations and balanced against efficiency concerns in a systematic way. For this purpose this thesis seeks to widen the economic framework, by integrating value interpretations that describe the fairness of a distribution. This objective poses challenges, both philosophical (which equity concerns represent widely shared believes?) and methodological ones (how to get the balance right?). Before proceeding in this direction, however, the remainder of this chapter will discuss economic evaluations, their theoretical basis, and the discussion they have provoked, to motivate why equity is the concern on which the subsequent chapters of this thesis concentrate.

1.2 Historical perspective on economic evaluations

Presenting health economic results for a clinical audience some twenty years ago could evoke passionate reactions from the audience. It was not unusual for the financial arguments in the health care debate to be received with horror, while the health economists were characterised as cold-hearted money-makers. At those early stages of the debate, differences between health economic analyses and clinical practice seemed unbridgeable. It had at least one advantage: standpoints were clear and typically related to the position of the debaters in the field. On one side were the physicians and patients who defended Hippocrates' Oath that every patient should be treated. They emphasised that the relation between doctor and patients is a unique and private one, which would be endangered by any outside interference. Lawyers who defended the individual right for health care, supported the claims of physicians and patients. Policy makers and health economists took another side in the debate. The policy makers were the first to encounter a complex allocation problem: the budget claims of the health care community rose faster than any other public spending, causing severe problems in government budget control (♠ **Figure 1.1**). After endless negotiation and several policy changes, it became clear that there was no hope that the medical community would be willing or able to stop the ever-expanding budget claims. At that time, the economists stepped in with their expertise: making optimal decisions in situations of scarcity, ignoring the 'no entry signs' surrounding the private doctor-patient relation. The policy makers embraced the expertise of the economists. They also welcomed the recognition by health economists of their policy problem: choices were becoming inevitable given that budgets were running out of control.

The days in which we could afford to rely on the arguments based on our position in the field are now long past. This becomes clear from the decreasing effectiveness of traditional strategies for budget control, central budgeting and provision planning. When health policy makers decentralised their budgets, budget constraints have entered the

♠ **Figure 1.1** Increasing health care costs in the United States, the Netherlands and the United Kingdom



consulting room of the physician, forcing the physician not to waste any money. A clever political move, as policy makers could now share the responsibility for the problems surrounding budget control. As however the gap increased between the care that was medically possible and the budgets would allow, clinicians in turn started to share their problems with policy makers. The policy makers were confronted with clinical quality concerns: waiting lists and insufficient care have become important political issues in the public debate. Furthermore, they were confronted with patients who successfully reinforced their right for health care by law or by public indignation. This means that budget policy must be justified by more arguments than just financial constraints. In summary, both parties gradually adopted a dualistic position in the debate. Therefore, both sides in the debate have much to gain from integrating the expectations of physicians and patients on the one hand and the financial constraints on the other into an acceptable decision-making framework.

In this thesis, I discuss a model, which tries to integrate the most important arguments form both sides into one decision-making framework. I shall argue that this model partly bridges the gap between health economists and physicians and that it improves the understanding of many decisions in which economic and medical arguments seem to contradict. This understanding is attained in both cure and prevention. Decisions in care remain less adequate described because the outcome of care is less well defined.

1.3 The economic paradigm and its critics

Economic evaluation is a tool for establishing the relative efficiency of health care programs and therefore is useful in allocating health resources. To evaluate the efficiency of a health care program, the additional costs of a medical intervention are compared to the effects as a result of treatment. For this purpose the effects are expressed in a generic outcome, like life years or Quality Adjusted Life Years (QALYs) gained. The QALY combines effects on life duration and quality of life in one outcome, by weighing each year for the quality of life during that year. The cost and effects are related as a ratio that can be used to compare different health care programs. Treatments with lower cost-effectiveness ratios are considered more efficient. If efficient use of resources would be the only concern, the best treatment from a health economic point of view is the one that provides the most health at the lowest cost. If a society allocates its resources to funding of the most efficient health care programs, the population will achieve its maximum possible level of health. This sounds like a rational goal for a national health care authority. Indeed, it seems quite reasonable, even logical, to argue that the maximum possible health should be the ultimate goal of the health care system at a national level.

Although health maximisation seems to be a rational goal, there has also been much opposition to the introduction of economic evaluations. People seemed to fear that economic evaluations would not be capable to fully recognize the value of particular interventions, misclassify them as in inefficient and restrict access to such services. An example is nursing care. In nursing care, not much health can be gained because health problems of the elderly are often irreversible. People feared that the true value of nursing care (i.e., to preserve quality of life in spite of bad health, rather than to improve health) would be ignored and relevant other moral arguments to provide services of nursing care would be overlooked. Additionally, people argued that patients who could benefit at the lowest costs are not necessarily the ones who deserve health care the most. For example, is it justified to favour treatment of smokers because of efficiency arguments or should decision makers also take into account that health problems of smokers are self-inflicted? Finally, people were concerned that the worst off patients might be left to suffer. It was argued that irrespective of the efficiency of the intervention needed, a society holds the moral obligation to care for the worse off patients. Since the health maximisation approach overlooked such distributional concerns, it is not surprising that the intention to use economic evaluations in health care policy gave rise to a number of moral questions. The political objective to integrate economic evaluations in health care decision making was viewed with suspicion: it looked like a quick fix for a complex problem.

People feared that outcome measures in economic evaluations would do no justice to the subjective, multiform and complex character of health. Measures like the QALY therefore became the object of scrutiny. But it is not very useful to take the discussion in this direction, because it ignores that we can choose what measure of health is implemented and that we can choose in what sectors of health care decisions will be made on the basis of economic evaluations. Therefore the arguments are not so much opposed to economic evaluations; rather they oppose their current operationalisation with health as a measure of utility, i.e. as the maximand. If people were only able to reach consensus about the kind of distribution that is desirable, economic evaluations could help to achieve this distribution. Indeed, it is more important discuss the reason why people object to economic evaluations. It is likely that much of the opposition to economic evaluations relates to the fear that access to health care would be limited and the solidarity basis underlying health care would crumble away. Dealing with this fear requires a normative discussion about distribution of resources in health care. This discussion should not be different if one considers outcome measures such as the QALY, life years gained or reduction in Hg/mm blood pressure. One only needs to assume that it is possible to develop a method for outcome assessment that is valid. I will therefore ignore the technical aspects in this debate and shift focus to the bigger issue: what do we want to measure and value, and how can policy makers make use of this information?

Indeed, there is recognition that the QALY approach is accompanied by a more fundamental problem than its psychometric properties (Wagstaff 1991; Nord et al., 1999).

A problem is that some patient groups can be identified for whom interventions systematically generate less effect than for other patients. For example, younger or healthier persons usually have a larger ability to benefit from interventions than the disabled and the old, whose potential is limited by their disabilities or shorter life expectancy. High costs interventions are therefore not likely to be cost-effective in such patients groups. Potential victims are the elderly, the chronically ill, the demented, the physically handicapped and patients with a low chance of cure, such as patients at the end stage of life. All these patients are regular users of health care. Although they have a low potential to benefit and their treatments often have unfavourable cost-effectiveness ratios, people seem to feel that disabled people have the same general rights of access to health and social care as other people. Here, striving for efficiency may conflict with our feeling of justice. In this respect, one can understand that, early in the debate, opponents of economic evolution pictured a future in which economist sacrificed the sick and the helpless in order to maintain health care spending within budget (Cohen, 1983). These opponents emphasise that the primary aim of health care is to help the weak, and certainly not to abandon them. For instance, Callahan wrote (Callahan, 1994):

"Our bias, I contend, should be to give priority to persons whose suffering and inability to function in ordinary life is most pronounced, even if the available treatment for them is comparatively less efficacious than for other conditions".

1.4 Economists' response to the critics

Of course, economists were not blind to arguments about the objectives of the health care system and the distribution of health gains. They were and are well aware that allocating health care resources to those patients who are most able to benefit, will result in a distribution of health in which some severely ill patients would be denied treatment. Furthermore, health economists were and are well aware that this allocation would be counterintuitive to the feeling that one should first help those patients who are worse off, irrespectively of their ability to benefit. For two reasons however, they were not all that bothered by this gap between the outcome of the analyses and intuition.

The first line of reasoning tries to counter the intuition that we should always help the worse off, irrespective of their ability to benefit. The classic utilitarian philosophy states that the right thing to do, under any given circumstances, is that which will produce the greatest amount of happiness of the whole. The justification is that all individuals should be treated as equals (Roemer, 1996). Even the tiniest gain in total health would outweigh detrimental effects on equality of the distribution of health. Redistributing outcomes to greater benefits of less advantaged at the cost of total health is considered inappropriate. That is because resources used on the one patient are not available to other patients who might have benefited even more from the same amount of resources, so that suffering is not reduced to a minimum. This is interpreted as an unethical waste of resources. An example might illustrate this position. Consider the situation were budget constraints allow us to treat one patient in condition A or four patients in condition B. Condition A affects an elderly person who would die immediately without treatment and lose 10 life years. Condition B affects younger persons who in a no-treatment situation would continue to live without disabilities for 30 years and after that they die, losing 5 years of their normal life expectancy each. Overall, intervention B will generate more health (20 life years (4*5) versus 10 life years of treatment A). If we choose option A, we deliberately throw away 10 life years causing unnecessary suffering. Assuming that individuals derive utility only from their personal outcomes and that a person's level of well-being has no relevance itself, this alternative distribution cannot be justified.

Also for other reasons distributive justice is often considered not to belong to the scope of economics, even for those who appreciate that people hold social preferences or use other decision rules than utility maximisation. It has for example been argued that a close relation of economics to policy would rely heavily on value judgments that are difficult to legitimate. It is considered appropriate for policy makers to weigh the outcomes of economic evaluations in any preferred way. Economists should not interfere with this political process, because science cannot legitimate normative choices so that results would be arbitrary. This distinction between normative and economic statements seems artificial, however, because both the measurement of efficiency or distributional preferences relies on value judgements. For others the question therefore was not so much if integration of distributive concerns in economic models is warranted, but rather if it is possible. Among those economists who accept that health maximisation is not the only target in health care and that also the distribution of health matters, a much-debated issue is how to obtain appropriate information on utilities related to both the distribution of health and the health outcomes themselves and how to use them in social appraisal. The reason is that the utility concept usually is assumed to reflect individual preferences, and it is not self-evident how interpersonal comparisons required to evaluate distributions can be captured in the economic paradigm. Although such theoretical issues may not constitute a sufficient argument for rejecting particular models for resource allocation, it explains why there was some reluctance in economics to consider distributive issues.

The arguments above represent differences in opinion about the aim of the health care system and/or the aim of the evaluation of its efficiency. To one side were the utilitarians emphasising that the health care gains are the primary target of health care, or at least the primary target of the efficiency analyses. They focus on the production of health, and consider that the main objective of health care should be the maximisation of health in the population. This utilitarian view is disputed from an egalitarian standpoint: health care should focus on the weak and those who are not able to take care of themselves. Therefore the aim of the health care system should be to help those most in need, which concern overrides the interest in maximizing benefits from health care services. A metric way of saying this would be that the utilitarian view aims at the highest possible average health in the population, irrespectively of the variance (the distribution of health), whilst the egalitarians aim to reduce variance, irrespective (at least, to some extent) of the consequences for the average health of the population. But even when economists subscribe to the importance of this alternative distribution, some of them will not change their evaluations: in their eyes economics is a positive science and not normative. But this does not apply to all economists. Within health economics an increasing number of people aim to integrate distributional concerns into the economic framework. This thesis aims to contribute to this work.

1.5 Bridging the gap

An increasing number of economists disagree with the idea that economists should stay far from political interpretation of the results of economic evaluations. An important reason is that many outsiders already consider economics as a normative science, which prescribes how choices must be made and therefore categorically reject its use in policy decisions. This approach, however, is not very subtle and fails to recognize the diversity of utilitarian theories. In the recent economic literature, for example, studies have been published about conditions under which distributive issues can be reconciled with equity concerns (e.g. Bleichrodt, 1997). Other modern economists have argued that utility may not be the only relevant concept in social decision-making and that the utility framework should merely be used as a tool to represent preferences rather than as a formal normative framework (Hurley, 2000). An additional argument that is used frequently to motivate the integration of normative beliefs into the economic framework is that the moral arguments that may outweigh efficiency concerns in reimbursement decisions often remain implicit and arbitrariness could be the result. To deal with the criticisms and to further rationalise decision making, economists increasingly try to integrate normative considerations into their analytical models. The basis of these recent developments is the hypothesis that the value of an intervention is not only determined by the amount of health it generates, but also by the distribution of health that it generates. The value we attribute to different distributions can be measured and compared, by comparing how we value health benefits when they accrue to different patient groups. This value for distributional effects can be subsequently incorporated into the economic analysis.

A first way of incorporating distributional concerns is to measure social preferences for different distributions of community health directly. This involves reflection on the value of a health gain to a particular person, instead of valuation of a health gain per se.

For example, one could ask respondents to trade-off a lesser health benefit for a larger number of people against a larger benefit for a small number of people. Erik Nord was one of the first researchers who advocated this approach to measuring the social value of particular interventions. Nord emphasised that in health policy, choices are not made for patients in isolation but between groups of patients. Making choices between groups of patients means that distributional aspects may play a roll. Theoretically, he argued, it is therefore wrong to base the QALY measurement solely on the quality of life measures that record values for the health state 'of a person like your self'. After all, this means that the value judgement is made for one person in isolation, namely for oneself. Nord therefore wanted to adapt the quality of life measure used in the QALY-approach to incorporate distributional concerns: he explored the potential use of a quality of life measure which was based on a trade-off technique which compares whole groups of patients. In this way he could incorporate distributional considerations when making value judgements about health states (Nord et al., 1999). Unfortunately, investigations of the psychometric characteristics of these so-called person trade-off methods show much less favourable results than the traditional trade-off methods such as time trade-off and standard gamble (Green et al., 2000).

A second way to incorporate distributional considerations into economic evaluations might be to leave the economic evaluation intact in terms of 'cost per life year' or 'cost per QALY', but to adapt the decision model that we apply to these outcomes. Basically, the individual value for a specified health benefit is than converted into a social value equivalent using a weighting that represents the relative weight of treatments for this patient compared to other patients. Wagstaff demonstrated that striving for an efficient distribution of health does not rule out the option of distributional concerns being a factor in health care priority setting (Wagstaff, 1991). Acknowledging that both efficiency and equity are objectives in health care, Wagstaff clarified that it is possible to combine the two by weighting health outcomes for specific equity dimensions, such as ill health, age, or socio-economic condition. These weights then represent the loss in public health, which would be considered acceptable if it would result in a more equitable distribution of health. Wagstaff thus explained that we are dealing with a trade-off: it is possible to obtain a more equitable distribution of health, but only at the cost of a lower average level of health. He called this trade-off the equity-efficiency trade-off.

Elegant about Wagstaff's model is that the QALY remained intact, but a new field was opened up for research: studying the relative importance of specified QALY gains for different patients. In the struggle concerning the question of how to set priorities for health care spending and how to use the outcomes of economic evaluations, this was a major breakthrough. First because it breaks with the view that economics and ethics are incommensurable areas. Second because its shows what aspects of cost-effectiveness analyses are amenable to a searching inquiry into their compatibility with beliefs about justice and fairness. This theory therefore is used as a basis for this thesis.

1.6 Avoiding arbitrariness

To date, a significant number of studies have reported the role of economic and distributional concerns in resource allocation. What emerges from these studies is consistent: evidence is mounting that people in fact do make a trade-off between equity and efficiency when they consider health care priorities. For example, Dolan demonstrated that a specified health gain of 20% for a person in a poor condition was valued equal to a health gain of 40% for a patient who was in a better initial state of health (Dolan, 1998). Other studies gave similar results (Nord et al., 1999; Lindholm et al., 1998). These studies have probed and confirmed the feasibility of measuring a trade-off between equity and efficiency, suggesting that this strategy has the potential to improve the compatibility of economic evaluations with concerns of fairness. However, it is unclear how we should proceed further. A review of studies into the equity-efficiency trade-off showed that different studies emphasised different distributional objectives and that health care policies do not always reflect the social values elicited by empirical studies into the equity-efficiency trade-off (Sassi et al., 2001). To integrate distributional concerns in economic evaluations, we must therefore first determine which equity concerns are important.

It would seem reasonable to assume that the equity concerns, which have been expressed as a reaction to the existing economic evaluation techniques, reflect widely shared social values with which economic evaluations come into conflict. Unfortunately, those who are sceptic towards economic evaluations do not always offer an alternative that better reflects the preferences of individuals and society with regard to the distribution of health effects. For example, it is frequently argued that the classic utilitarian view on health economics is ageist, because the efficiency calculus discriminates against groups that cannot gain many QALYs, like the chronically ill or the old (Harris, 1987). That is because in a cost-effectiveness analysis, the health gains of elderly patients get the same weight as those of younger patients. Given that they do not have as many years to go as a younger person, they will often be the less fortunate ones in priority decisions. Following the methods outlined by Wagstaff, it could be argued that the elderly should be compensated for their shorter remaining life span, by attributing a higher weight to their QALYs. That is, if society believes that the elderly should indeed receive these treatments! Several studies have suggested however that the opposite is true: in a situation of scarcity, society chooses intentionally to prioritise younger patients over the older ones (Busschbach et al., 1993; Nord et al., 1996; Cookson and Dolan, 1999; Tsuchiya, 2000).

Apparently, the moral objections towards economic evaluation do not have the social validity that people may attribute to them. This lays bare what may be the real problem:

our moral views themselves have been insufficiently explored. Whilst financial arguments are often mistrusted because they envisage a limitation of the debate within society, the opposite is true for moral arguments: usually moral arguments outweigh financial ones without further reflection on their appropriateness (Borgmann, 1992). The problem now becomes that arbitrary and capricious moral considerations are perhaps defended under the cloak of justice and allowed to fudge resource allocation issues in ways that would not be acceptable, were their basis exposed (Williams, 1997). It is clear that we are afraid to do injustice only on the basis of the costs, but we should be equally afraid to do injustice on the basis of insufficiently founded or conflicting normative claims!

Probably the best way to open up the debate on what form of equity we are interested in, is by identifying which patients we would give the lowest priority. The decision to reject a treatment is unpopular. The benefits only become apparent in the long run and only for the economy as a whole, whereas identifiable patients are left to suffer in the short run. Equity concerns therefore are often only positive ones, stating reasons why a patient should be treated after all, leaving aside the question from which patients resources might be taken (Van de Vathorst, 2001). Identifying patients who should get low priority, is then a crucial test if we want to get a grip on societal values. Unfortunately, principled approaches of a comparative nature that explicitly priority rank patients are rare. To bridge the gap between health policy and economic evaluations, moral considerations need to be discussed in a meaningful way. The challenge ahead of us is to develop a framework that describes the mix of equity concerns that people subscribe to and how they vary the chosen equity concept from one context to another, so that it becomes possible to discriminate between claims that are based on appropriate and inappropriate moral concerns.

1.7 From theory to practice

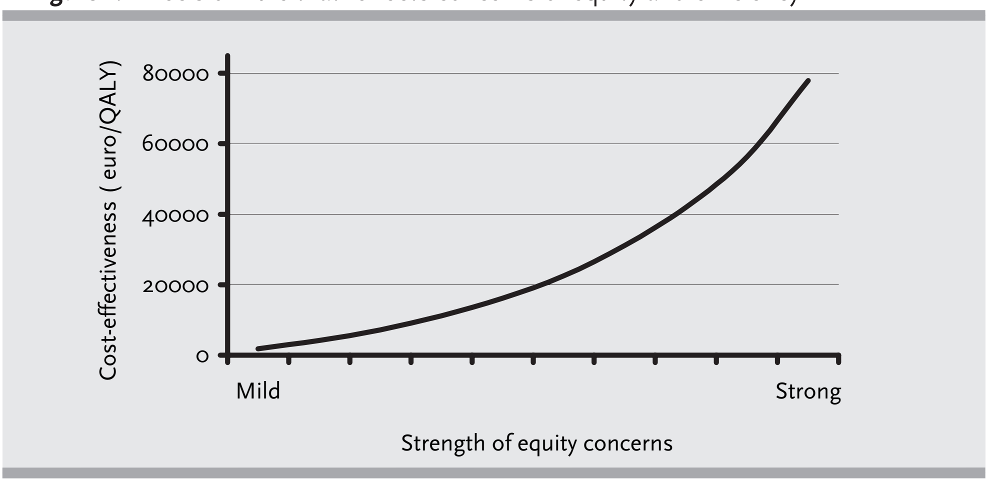
The theory presented so far pictures a policy model for priority setting that makes a trade-off between two targets in health care: the wish to use the resources efficiently, and the wish to distribute the resources fairly. Generally speaking the policy implication of a trade-off between equity and efficiency is that society is prepared to accept higher costs for treatment of patients in a worse condition. Assuming that the nature of our equity concerns can be clarified, the challenge will be to address these concerns in a policy model. To make sure that the criteria are applied in a consistent manner, quantification is necessary.

♠ **Figure 1.2** graphically depicts the general contours of a policy model that balances equity and efficiency concerns. The basic assumption is that health gains for some patients are valued higher than health gains in other patients. This in turn would imply that the society accepts different costs to achieve a given health gain in different patients. In

♠ **figure 1.2** the X-axis describes the strength of patients' claims on health care resources evaluated by an equity concept, the Y-axis reflects efficiency of treatment expressed in cost per QALY. The curved line in the figure shows how the decision rule based on efficiency might be differentiated for equity. In mainstream economic evaluations QALY maximisation is the goal. The treatment with the lowest cost per QALY gets the highest priority. If more budget is available, the next best treatment in terms of efficiency will be reimbursed, and so forth until the whole budget is allocated. In our graph a horizontal threshold for cost-effectiveness would represent this decision rule. The implication of the equity-adjusted threshold reflected in the curved line is that fewer resources are distributed to people in better health compared to those in worse conditions.

To put the theory of balancing equity and efficiency concerns into practice, quantification is needed. Only with quantification we can improve accountability and transparency of the priority setting process and guarantee application of the criteria in a consistent manner. Empirical work to satisfactorily define this framework is urgently needed, yet the road to integration of equity and efficiency concerns is paved with challenges. It is clear that several variants of trade-off techniques could be used to investigate how far society wishes to discriminate between people in a good or bad condition. The problem is to define the questions tightly enough to target the (mix of) equity concerns that are considered relevant in priority setting. If the questions fail to distinguish between relevant equity and efficiency concerns, respondents may interpret questions in unanticipated ways and the answers to the hypothetical questions may poorly predict preferences for real-world policy decisions. Questions should exclude interference from confounding variables, but still be sensitive enough to be inserted in a theoretical framework that describes all factors that efficiency should be traded off against. Otherwise it may be impossible to separate

♠ **Figure 1.2** Decision rule that reflects concerns of equity and efficiency



out weights and potential interactions effects for each factor that is considered relevant (Dolan et al., 2005). The strengths and weaknesses in elicitation procedures are thus closely related to the strengths and weaknesses at the conceptual level. It is fair to say that we have a long way to go in developing methods to elicit preferences that include and balance moral and financial concerns.

1.8 Aims of this thesis

The aim of this thesis is to bridge the gap between those who advocate and those who oppose the use of economic evaluations in health care decision-making on the grounds of fairness. The assumption underlying this work is that although health economics is primarily concerned with finding the most efficient allocation of health resources to achieve a given policy goal of resource allocation, equity or fairness is a separate and important concern that health economists and other policy analysts should address. To ensure that neither equity nor efficiency concerns get disregarded, this thesis explores how both concerns can be balanced in a systematic way. Central questions are:

- How is cost-effectiveness evidence used in health care priority setting, and how are equity concerns taken into account?
- Is it possible to determine what equity concepts are the most relevant ones and measure the accompanying weights for the development of an equity adjustment procedure?
- What is the potential of equity adjustment to regulate costs in health care expenditures?

Guided by these central questions, this thesis moves from the observation that economic evaluations are not always utilized in health care decision-making (Chapter 2) toward explanation of this phenomenon in terms of fairness (Chapter 3). Next several chapters are devoted to resolving the tension between fairness and economic evaluation by expanding economic models so that they can integrate distributional concerns (Chapters 4, 5 and 6). The field of economics offers tools to determine weights for arguments in the distribution, and it identifies what type of operationalisation of different distributional concerns is needed for them to be a feasible support in priority setting. However, reflection on the meaning of equity also requires analyses from an empirical-ethical point of view. These chapters therefore also draw upon analytical philosophy and ethics. This thesis comes full circle with a reflection on the potential value of the presented theoretical and methodological developments (Chapter 7). Finally, chapter 8 contains a discussion of the conclusions derived in this thesis and directions for future research.

Damien Hirst: Pill stool

Chapter 2
Cost utility analysis of sildenafil compared with papaverine-phentolamine injections

Objective: Sildenafil is expected to be more costly to society in treating erectile dysfunction, but more effective than conservative therapy (papaverine-phentolamine injections). To analyze whether the beneficial effects are worth the additional costs, we performed a cost-utility analysis to compare the cost-effectiveness of treatment with sildenafil and treatment with papaverine-phentolamine injections.
Design: We compared two scenarios, the sildenafil scenario (allowing a switch to injection therapy) and the papaverine-phentolamine scenario (no switch allowed). Costs and effects were estimated from the societal perspective. Using time trade-off, a sample of the general public (n=169) valued health states relating to erectile dysfunction. Using these values, we estimated health related quality of life by converting the clinical outcomes of a trial (Goldstein, 1998) into quality adjusted life years (QALYs).
Results: According to the general public, erectile dysfunction limits quality of life considerably: the mean utility gain attributable to sildenafil is 0.11. Overall, treatment with sildenafil gained more QALYs, but the total costs were higher. The incremental cost effectiveness ratio for the introduction of sildenafil was £3639 in the first year and fell in following years. Doubling the frequency of use of sildenafil almost doubled the cost per additional QALY.
Conclusions: The incremental cost-effectiveness ratio suggests that sildenafil is a cost-effective medicine. When considering reimbursement of sildenafil, it should be taken into account that the frequency of use affects this cost-effectiveness ratio.

Chapter 2

Cost utility analysis of sildenafil compared with papaverine-phentolamine injections

Based on: Stolk EA, Busschbach JJV, Caffa M, Meuleman EJH, Rutten FFH. Cost utility analysis of sildenafil compared with papaverine-phentolamine injections. British Medical Journal 2000;320:1165-68

2.1 Introduction

The registration of sildenafil has initiated debate about the socioeconomic aspects of this treatment for erectile dysfunction. Generally, governments are concerned about the affordability of sildenafil (Dinsmore and Evans, 1999). It is not known whether sildenafil is cost-effective. Although the clinical effects of sildenafil have been proved, uncertainty remains about the value of sildenafil to both patients and society.
We performed an economic evaluation of sildenafil according to the usual recommendations (Gold et al., 1996). We used cost utility analysis, a form of cost effectiveness analysis in which clinical outcomes are converted into quality adjusted life years (QALYs) gained. Both costs and effects were measured from the societal perspective. This means that treatment outcomes were valued by the general public and that all costs were considered— that is, medical costs, costs of patients, and costs in other sectors of society. Costs and effects were analysed over five years.

2.2 Participants and methods

We compared the costs of treatment with sildenafil with that of conventional treatment. Before the introduction of sildenafil, injection therapy was the treatment of choice for erectile dysfunction. Many patients, however, were unwilling to receive injection therapy and accordingly did not seek treatment. We therefore assumed that injection therapy was accepted by 10% of patients (Pfizer, Netherlands, personal communication, 1998, based on market research). The vasoactive substance was papaverine-phentolamine and not alprostadil, which is more commonly used, because papaverine-phentolamine is less expensive and equally effective. Papaverine-phentolamine injections are reimbursed in the Netherlands, but no decision has yet been taken about reimbursement for sildenafil.
We estimated utility values for different states of erectile dysfunction. These utilities were applied to the clinical outcomes before and after treatment in a clinical trial of sildenafil by Goldstein et al. (1998). We also estimated the costs of two treatment scenarios for erectile dysfunction and analysed these in a model comprising the probabilities of successful treatment, switching and discontinuation of treatment, and duration of successful treatment. A detailed description of our methods to analyse costs and effects is available (Stolk et al., 1999).

2.2.1 Clinical effects

The study by Goldstein et al. is the largest dose escalation study reported (Goldstein et al., 1998). It was placebo controlled and the patient population consisted of men with erectile dysfunction due to various causes. Efficacy was assessed with the international index of erectile function (Rosen et al., 1997). This instrument contains questions about the two primary end points of erectile dysfunction treatment as defined by the National Institutes of Health—that is, the ability to penetrate and the ability to maintain an erection sufficient for satisfactory sexual intercourse (NIH Consensus development panel on impotence, 1993). These end points were used in the trial. Both questions have five response levels, so together they categorise the patients into 25 (5×5) erectile dysfunction states. These erectile dysfunction states were valued in a separate exercise (described below). The elicited utilities were applied to the health states of the patients in the study of Goldstein et al before and after treatment. The difference between the mean utility before and after treatment (controlled for placebo) is the mean gain in utility. Use of disease specific instruments to calculate QALYs is advocated by Brazier and Dixon (1995) and Drummond et al. (1997).

Because we used previously reported trial data, we had to consider the limitations of these data for use in economic evaluation. Firstly, the trial was designed on an intention to treat basis (Goldstein et al., 1998), which meant that patients for whom sildenafil had no or insufficient effect remained in the trial. As we could not discriminate between patients with a sufficient or an insufficient response, we used the mean utility gain in the trial to calculate the utility gain of sildenafil. Consequently, we underestimated the utility gain in daily practice because only the utility gain of the successfully treated patients should be taken into account. Secondly, results of the international index of erectile function were not available for injection therapy, nor were any other data that allowed calculation of QALYs. We conservatively assumed that the utility gain of sildenafil and papaverine-phentolamine injections would be the same. Given the low acceptability of injection therapy (Mulhall et al, 1999; Althof et al., 1989), this assumption probably overestimates the benefits of injection therapy.

2.2.2 Determining utilities for erectile dysfunction states

From a randomly selected sample of 45,000 people obtained from the Rotterdam telephone directory we recruited 354 people to participate in the valuation task. They were invited by telephone to attend a session of health state valuation and were offered about £10 plus travel expenses. In order to avoid selection bias, the invitation was made without referring to erectile dysfunction. Participants were given the opportunity to withdraw from the valuation sessions without financial consequences after they were informed about the subject of the study.

Participants valued 24 erectile dysfunction states on a scale from 0 to 1 using time trade-off (Drummond et al. 1997). The 25^{th} state described normal erectile functioning and was set at a value of 1.0. Time trade-off was measured relative to the life expectancy of the subjects. Before the valuation task, participants gained experience of the time trade-off method using general health states as defined by the EQ-5D questionnaire (Brooks, 1996).

Time trade-off responses were considered invalid if the participant showed a lexicographic response for the EQ-5D states, had too much missing data either on erectile dysfunction or EQ-5D states, or clearly did not understand the task. A lexicographic response mode means that when a respondent is faced with an option he or she will always choose one particular alternative, no matter how favourable the other might be. Subjects had to value the health states "for a person like yourself." This means, for example, that older people gave values from their own perspective, and people without a sexually active partner would take this into account when performing the valuation task. The exception was that women were asked to imagine being a man with erectile dysfunction. Values are independent of the sexual activity of the respondents because the descriptions of erectile function referred to the relative number of successful attempts at intercourse. For example, a respondent might be asked the following: "If during the past four weeks, your condition was such that you were sometimes able to attain an erection, and you were (almost) never able to maintain your erection, how many years would you be willing to trade off to restore your erectile function?" This also implies that erectile function is valued the same in patients with different levels of sexual activity.

Because we had decided to obtain social valuations we asked a sample of the general public to value the clinical outcomes (Drummond et al, 1997; Gold et al. 1996). The reasoning behind this decision relates to issues of equity and medical ethics (Williams, 1994). Some authors, however, claim that healthy people relatively similar to affected patients should value clinical outcomes (Hadorn, 1991). We therefore explored whether erectile dysfunction is valued differently in different subgroups. We used multivariate analysis of variance to determine whether age, sex, the availability of a partner, having children, sexual activity, and sexual satisfaction influenced the values of the general public.

2.2.3 Costs

All costs are expressed in 1999 British pounds (£1 = €1.62). We used 1999 data to determine the Dutch cost prices. To determine the medical costs, we estimated resource use— for example, consultations and prescription charges (a lump sum charge to refund pharmacy costs and medicines) and multiplied the quantities by the unit prices. We estimated resource use of sildenafil and papaverine-phentolamine injections on the basis of consensus statements on both treatments (Round table conference, 1998). We refined this estimate by developing a low, baseline, and high cost scenario on the basis of clinical experience in two hospitals (University Medical Centre St Radboud, Nijmegen and Hospital St Antoniushove, Leidschendam). Costs outside the healthcare sector and productivity costs were assumed to be negligible.

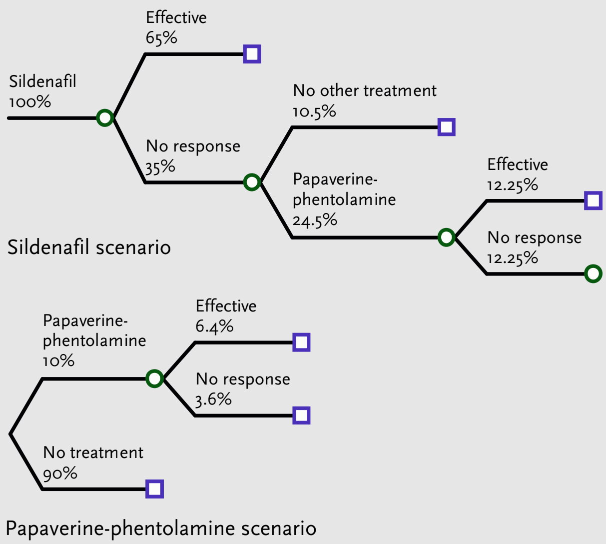
The cost of sildenafil was based on observational data from the first quarter that sildenafil was available in the Netherlands (Foundation of Pharmaceutical Statistics, 1999). A general practitioner or urologist determined the effective dose in an academic or peripheral setting. An appropriate share of the costs of supporting departments was reflected in the cost of a visit to a urologist. The physician's costs were calculated on the basis of the estimated duration of an outpatient visit. The analysis included all costs related to the hospital, such as costs of salaries and supplies, costs of supporting departments, and overhead costs.

2.2.4 Cost effectiveness

We compared two scenarios: treatment with sildenafil and treatment with papaverine-phentolamine (♠ **figure 2.1**). In the sildenafil scenario, we allowed patients to switch to papaverine-phentolamine injections, as these injections may be effective in patients in whom sildenafil has failed. Since sildenafil has already become the treatment of choice, although its cost is not reimbursed in the Netherlands, patients are unlikely to switch from injections to sildenafil. A switch was therefore not allowed in the papaverine-phentolamine scenario.

We compared the sildenafil and papaverinephentolamine scenarios assuming use once a week. The maximum recommended frequency of papaverine-phentolamine injections is once a week, but use of sildenafil is not limited for medical reasons. Nevertheless, we believe that the assumption was appropriate as the utility values were elicited

♠ **Figure 2.1** Patient flows for scenarios of treatment of erectile dysfunction with sildenafil or papaverine-phentolamine injections



independently of the frequency of intercourse. Moreover, the marginal utility gain of increasing intercourse frequency from once to twice a week is unlikely to be as high as the increase between never being able to have intercourse and being able to have intercourse once a week.
The cost effectiveness of the different scenarios was analysed in a model comprising acceptability of treatment, probability of successful treatment, switching or discontinuation of treatment, and the duration of successful treatment. The patient flows in the model were determined on the basis of secondary data—for example, published clinical trials (Meuleman et al., 1998; Jarow et al., 1999; Jackson and Lue, 1998; Korenman, 1998), Dutch observational data (Foundation of Pharmaceutical Statistics, 1999), and clinical experience in the two participating hospitals. We performed an incremental analysis of the costs and effects of sildenafil compared with papaverine-phentolamine. The results are presented as cost per QALY.
The acceptance rate of papaverine-phentolamine treatment could have been influenced by the fact that erectile dysfunction is no longer a taboo subject. In fact, an acceptance rate of 70% has been suggested as feasible (Lycklama à Nijeholt, 1998). We therefore included this variable in a sensitivity analysis. Other variables included in the sensitivity analysis were resource use, values, effectiveness of treatment, and frequency of use. We performed univariate sensitivity analysis to determine which variables have the largest influence on the results. In the multivariate sensitivity analysis we explored to what extent results would change under a (unlikely) worst case scenario.

2.3 Results

2.3.1 Respondents

A total of 184 subjects (52%) failed to attend the interview sessions. This was probably because of extremely bad weather at the time of interview, which made it difficult for participants to reach the university. One person withdrew from the study after he was informed about the subject. A sample of 169 subjects valued the erectile dysfunction states; 89% (150) of the responses were valid. Age ranged from 18 to 80 years (mean age of 45.8 (SD 15.4) years). There were 81 men (54%) and 69 women, which is close to the sex distribution in the general population.

2.3.2 Effects

In Goldstein et al.'s study (1998) the international index of erectile function among men receiving sildenafil rose from 2.0 at baseline to 3.9 at end of treatment for ability to penetrate (placebo group 2.1 to 2.3) and from 1.5 to 3.6 for satisfactory sexual intercourse (placebo group 1.6 to 1.8). ◘ **Table 2.1** gives the mean utilities that were elicited for the 24 erectile dysfunction states described by these two questions. The utilities ranged from

Table 2.1 Social values (SD) for erectile dysfunction (n=150)

Ability to maintain an erection	Ability to attain an erection: Never	Few times	Sometimes	Most times	Always
Never	0.74 (0.18)	0.79 (0.17)	0.82 (0.17)	0.82 (0.15)	0.84 (0.17)
Few times	0.77 (0.18)	0.83 (0.16)	0.85 (0.16)	0.86 (0.15)	0.88 (0.16)
Sometimes	0.79 (0.16)	0.85 (0.14)	0.87 (0.14)	0.90 (0.13)	0.91 (0.13)
Most times	0.81 (0.17)	0.86 (0.15)	0.88 (0.14)	0.94 (0.12)	0.93 (0.13)
Always	0.82 (0.17)	0.87 (0.15)	0.91 (0.13)	0.94 (0.11)	1.00

0.74 to 0.94. When these values are combined with trial data, the mean utility increased from 0.807 at baseline to 0.915 at end of treatment for men receiving sildenafil and from 0.819 to 0.821 for men receiving placebo. Therefore, the mean utility gain attributable to sildenafil is 0.11.

We analysed whether the values of the general public were influenced by age, sex, availability of a partner, having children, sexual activity, and sexual satisfaction. The only relation we found was that participants with children considered erectile dysfunction less of a problem than subjects without children. Since there were no differences between the

Table 2.2 Volumes of resource use

	Units	Sildenafil	Injections	Source
General model	Acceptability of treatment (%)	100	10	1
	Treated by GP (%)	80	10	1
	Treated by urologist (%)	20	90	1
	Pills/ampoules p.w.	1	1	2
Establishing effective dose	Number of visits	3.8	3.7	2
	Mean duration of visits (min.)	8.5	10.3	2
	Number of prescriptions	2.5	2.3	3
	Discontinuing treatment (%)	35	36	4
	Effectiveness switch (%)	50	-	2
Remaining part of 1st year	Number of visits	2.0	2.0	2
	Mean duration of visits (min.)	7.5	7.5	2
	Number of prescriptions	7.8	4.3	3
	Discontinuing treatment (%)	10	14	2
Each following year	Number of visits	1.5	1.5	2
	Mean duration of visits (min.)	7.5	7.5	2
	Number of prescriptions	9.5	5.2	3
	Discontinuing treatment (%)	5	5	2

1 = Pfizer BV market research 1998, **2** = experts participating hospitals, **3** = Foundation of Pharmaceutical Statistics, 1999, **4** = Mulhall et al., 1999; Althof et al., 1989; Meuleman et al., 1998; Jarow et al., 1999; Jackson and Lue, 1998; Korenman, 1998 values of men and women, we used averaged values in the QALY analysis. More extensive description of this analysis is available (Stolk and Van Busschbach, 2003).

2.3.3 Costs

◘ **Table 2.2** and ◘ **table 2.3** show the resource use and the costs attributable to treatment of erectile dysfunction with sildenafil or papaverine-phentolamine injections. Papaverine-phentolamine is cheaper per dose, but it has to be prescribed by an urologist and therefore has higher initial costs (£484 versus £407 for sildenafil). Sildenafil has higher running costs: yearly treatment costs are £254 versus £233 for papaverinephentolamine. The higher initial costs of papaverinephentolamine are recovered after seven years.

2.3.4 Cost effectiveness

Overall, sildenafil creates more benefits and more costs because more patients are treated (♠ **figure 2.1**). Therefore, the main issue is whether the additional effects of sildenafil are worth the additional costs. This question is addressed in the incremental analysis shown in ◘ **table 2.4**. The incremental cost utility ratio of sildenafil compared with papaverine-phentolamine is £3639 per QALY in the first year, decreasing to £2630 per QALY after five years.

2.3.5 Sensitivity analysis

The frequency of use influences the outcomes considerably. Doubling the frequency of use of sildenafil increases the cost per additional QALY by 45% in the first year and 85% in each following year. The initial costs are relatively high because the costs of non-responders are added to the costs of responders. Hence, the effect of the frequency of use on the cost per additional QALY is moderated in the first year. In the long term, however, the main cost driver with sildenafil is the drug.
Assuming a lower utility gain (0.08) than observed in the valuation study, resulted in a 37.5% increase in cost per additional QALY (Stolk et al., 1999). Effectiveness and

◘ **Table 2.3** Unit cost (£)

Units of resource use	Cost price (ex VAT)
Sildenafil tablet	*4.33
Papaverine/phentolamine injection	3.55
Visit to general practitioner (<20 minutes)	10.32
Visit to urologist (weighted mean public/academic hospitals)	16.80+1.29 p.min
Prescription rule (a refund of pharmacy costs)	3.14

* Based on the 'effective dose distribution' across strengths of sildenafil (Foundation of Pharmaceutical Statistics, 1999)

acceptability also influenced the results significantly. The cost per additional QALY increased 38% with a lower effectiveness of sildenafil (50%), but decreased (1%) in each following year. Changes in acceptability had an opposite effect: when acceptability of papaverine-phentolamine injections is increased to 70%, the incremental cost utility ratio is 25% lower in the first year, but 10% higher from the second year onwards. Uncertainty about resource use did not influence the outcomes significantly; in the analysis of different cost scenarios (based on the number of visits and duration of visits), the high cost scenario increased the costs per additional QALY by only 8%.
When the uncertainty of all variables is combined into a worst case model (low utility gain and effectiveness and high costs, dropout, and acceptability), the incremental cost utility ratio is £9343 per QALY in the first year (156% increase), and £4691 in each following year (101% increase).

2.4 Discussion

The mean incremental cost utility ratio of sildenafil compared with papaverine-phentolamine was £3639 per QALY in the first year and improved in the following years. This cost utility ratio is generally favourable, as suggested acceptable thresholds of cost utility vary between £8000 and £25 000 (Laupacis et al, 1992; Goldman et al., 1992). Moreover, many interventions with less favourable cost utility ratios are currently being funded, such as breast cancer screening (£5780 per QALY) and kidney transplantation (£4710 per QALY) (Maynard, 1991). Uncertainty in the data did not hamper interpretation of the results: even in the worst case scenario, the incremental cost utility ratio of £9343 could be considered favourable. Our analysis therefore suggests that the clinical effect is derived at reasonable costs.

Table 2.4 Costs and effects of treatment with sildenafil and papaverine-phentolamine injection and their increment

	Successfully treated (%)		Incremental cumulative		Cost-utility
Year	Sildenafil	Injections	QALYs	Costs (£)	(£/QALY)*
1	77.25	6.4	7.79	£28,368	£3,639
2	69.53	5.5	14.84	£44,773	£3,017
3	66.05	5.23	21.53	£60,356	£2,803
4	62.75	4.97	27.88	£75,161	£2,695
5	59.61	4.72	33.92	£89,226	£2,630
∞*					£2,329

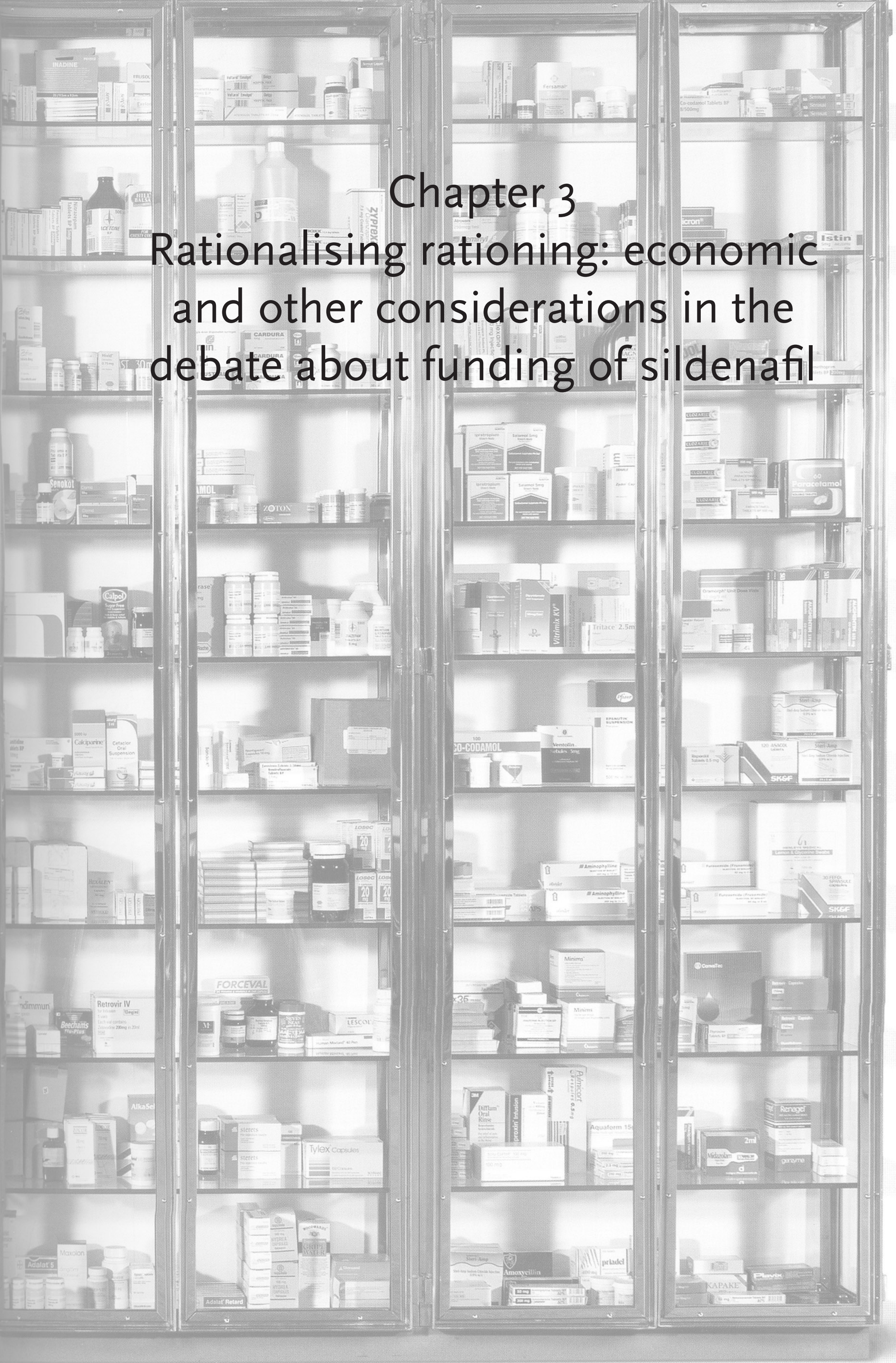
* **ICUR** = incremental cost-utility ratio; The ICUR is £3,639 per QALY in the 1st year and £2,329 in following years. On long term, the influence of the 1st year treatment costs diminishes and the ICUR approaches £2,329 per QALY

We made several assumptions that could be viewed as unfavourable to sildenafil. For instance, we underestimated the effects by not including partner satisfaction and we assumed the effects of oral and injection treatment to be equal. Furthermore, we used a relatively low rate of drop out for injection therapy (e.g. compared to Mulhall et al., 1999; Korenman, 1998), which results in a more favourable cost effectiveness ratio for injection therapy. Although such assumptions might introduce bias, the interpretation of the results is not greatly affected because the assumptions in the economic appraisal of sildenafil were conservative.

The utility values we elicited for erectile dysfunction did not take into account possible comorbidity. As in most cases total disutility is less than the sum of parts (Furlong et al., 1998), we might have overestimated the effect. However, the sensitivity analysis showed that sildenafil remained cost effective with lower utility gains. The subjective nature of the value of erectile functioning again raises issues about whose values should be used in economic appraisal of health care: the values of the general public or those of people at risk (in our case ageing men). However, we found that the utility values for sexual functioning were independent of background variables such as age, sex, and sexual activity. Therefore, neither the limitations in the representativeness of our sample, nor our choice to elicit values from the general public has influenced the results.

These findings should be interpreted in the light of the discussion about the affordability and value of sildenafil to society. Firstly, we have shown that erectile dysfunction limits quality of life considerably, in the eyes of the general public. Furthermore, our study shows that sildenafil is cost effective, and its reimbursement should therefore be considered. However, as frequency of use greatly affects cost, such reimbursement should not be unconditional.

Part of Damien Hirst's The fragile truth



Chapter 3
Rationalising rationing: economic and other considerations in the debate about funding of sildenafil

Although the cost-effectiveness of Viagra for the treatment of patients with erectile dysfunction is favourable, both public and political opinions seem to be inclined not to fund, or merely to partially fund (i.e. by reimbursing only specific patient groups) this medicine. This shows that in funding discussions, cost-effectiveness information is not solely decisive. In a theoretical framework for choices in health care that was developed in The Netherlands (the Dunning report, 1991), two other criteria besides cost-effectiveness were put forward as being important for rationing decisions: 'necessary care' and 'individual responsibility'. Reviewing the Viagra discussion, many of the arguments put forward seemed to be related to these two criteria. However, a clear operationalisation of the criteria necessary care and individual responsibility is lacking, which makes it difficult to use the arguments in funding decisions. In this paper, we try to demonstrate how these criteria were presented in the Viagra discussion and we will indicate how these criteria can be operationalised in relation to the outcomes of a cost-effectiveness analysis.

Chapter 3

Rationalising rationing: economic and other considerations in the debate about funding of Viagra

Stolk EA, Brouwer WBF, Busschbach JJV. Rationalising rationing: economic and other considerations in the debate about funding of Viagra. Health Policy 2002;59(1):53-63.

3.1 Viagra: an exceptional case in debates on rationing

In September 1998, Viagra was introduced on the European market, where it was an immediate success: the sales figures in the first months after introduction were extraordinarily high, especially given the fact that Viagra was a new, non-funded medicine. The question then became whether or not governments should fund Viagra in a public health care system. After all, the budgetary impact of funding of Viagra could be substantial, given this high demand. In such funding decisions, economic evaluation plays an important role in indicating the relative efficiency of the health intervention under consideration. Recently, several studies demonstrated that erectile dysfunction limits quality of life considerably. Hence, the use of an indisputably effective medicine such as Viagra results in a substantial gain in quality of life (Volk et al., 1996; Stolk et al., 1999). These gains are reached at fairly low costs, therefore, the cost-effectiveness of Viagra is very favourable (Smith and Roberts, 2000; Kwok and Kim, 1999; Stolk et al., 2000a).
◘ **Table 3.1** demonstrates that on basis of arguments of cost-effectiveness alone, Viagra is eligible for funding. However, cost-effectiveness information is not the only grounds for basing this decision. Compare this case with another from the table—heart transplants, for example. Heart transplants have much higher cost per QALY, yet their reimbursement is not a matter of debate. In contrast, funding of Viagra is fiercely disputed, indicating that other factors must also play a role in the allocation of health care resources (Stolk and Busschbach, 2000b). Previous experiences with reimbursement decisions already indicated the role of additional arguments, for instance in the case of lung-transplants, where unfavourable cost-effectiveness information was not enough reason not to fund them. Viagra, however, is an exceptional case in that despite a favourable cost-effectiveness ratio, its funding remains in dispute. The public debate about the desirability of funding for Viagra provides us with the rare opportunity to study those arguments favouring denial of reimbursement for a cost-effective medicine. This paper will therefore focus on the additional criteria, besides cost-effectiveness, that play a role in funding decisions. The paper mainly demonstrates that economic evaluation as it stands now, cannot explain why funding is denied or granted in different situations. However, if economic evaluations were broadened to include several societal preferences that are discussed in the paper, it would be a more useful tool in and better predictor of funding decisions. The paper therefore is an effort to contribute to a broader medical technology assessment framework, which incorporates elements other than merely costs and unweighed effects, bridging the gap between current cost-effectiveness analysis and the decision-making process. We will indicate how these criteria may be further operationalised and related to the outcomes of economic evaluations in health care.

3.2 Main arguments against funding Viagra

In the Viagra discussion, two main arguments are put forward, both in the political and in the public discussion, that object to funding the new drug (Brooks 1998; Ramsey-Baggs and Gaskell, 1998; Hayes et al., 1999). First, erectile dysfunction is often accepted as normal part of ageing. Consequentially, treatment can be viewed as an unnecessary luxury, which should not interfere with providing necessary medical interventions (e.g. lifesaving interventions). Second, it is argued that funding of treatment of erectile dysfunction may be denied on the basis of a 'private choice, private responsibility' argument. Having sex in advanced stages of life is perceived to be a personal choice, related more to lifestyle than to health problems. Viagra is thus considered to be a lifestyle drug, not eligible for public funding. In Germany, similarly, it was argued that when only some elderly individuals choose to have sex, while others forgo this pleasure, it is inappropriate to pay collectively for their 'pleasure'.

Although these criteria of necessity and individual responsibility have an intuitive appeal, they lack clear operationalisation. Definitions of for instance luxury health care and serious conditions were not provided in the public or political debate. Also, it was not made clear why Viagra would be an unnecessary medicine. Although erectile dysfunction is not life threatening, clinical need becomes obviously apparent when put in terms of quality of life (Stolk et al., 1999; Jønler et al., 1995; Litwin et al.,1998; Wilke et al.,1997; Gheorghiu et al., 1996). But how does a quality of life score relate to necessity of treatment? A possible answer to that question will be provided in the subsequent sections.

The argument of individual responsibility also lacks a clear operationalisation. How to distinguish between lifestyle purposes and legitimate medical use of Viagra is debated

Table 3.1 Outcomes of Dutch studies in terms of incremental cost per QALY

Intervention	Comparator	$/QALY
GM-CSF in elderly with leukaemia	Daunomycine cytosine	235,958
EPO in dialysis patients	Conservative treatment	139,623
Lung transplantation	Conservative treatment	100,957
End stage renal disease management	No treatment	53,513
Hart transplantation	Conservative treatment	46,775
Liver transplantation	Conservative treatment	44,566
Didronel profylase	Conservative treatment	32,047
PTA with Stent	PTA	17,889
Breast cancer screening	No screening	5,147
Viagra	Androskat	5,097
Surgery for CAM	No treatment	2,778

CAM = congenital anorectal malformation; The costs are valued for the year 1995. This table is adapted from Rutten et al. (2000).

vigorously (Gilbert et al., 2000). In The Netherlands ‘in vitro fertilisation’ has also been labelled as a lifestyle treatment. Nevertheless, the wish to have children is generally widespread and not related to a particularly deviant lifestyle (Anonymous, 2000). Furthermore, those persons wishing to undergo IVF experience great distress and discomfort, leading to an obvious loss of well-being or quality of life. Their only wish seems to be to live and function normally, just like ‘everybody else’. In that sense, it seems rather odd to claim that it is a lifestyle choice to have children. Similar comments can be made about many drugs, which have been labelled ‘lifestyle drugs’ (Gilbert et al., 2000).
In short, the arguments used in the public debate to support denial of funding for Viagra, are unsatisfactory and seem to be used only in an unconvincingly opportunistic manner. The arguments lack generalisability towards other conditions and other therapies, hampering consistency and transparency in decision-making. To progress further in developing a decision-making framework for choices in health care, we need to define the criteria necessary care and private responsibility in a more applicable manner, as done in the next section.

3.3 Operationalisation of other arguments

In spite of their poor current operationalisation, the arguments discussed here about necessity and individual responsibility have intuitive appeal and seem to be based on widely shared underlying moral principles.

3.3.1 Necessity

The criterion of necessary care seems to reflect the idea that patients in a poor state of health are more entitled to health care (if this can provide them with improvement or relief) than those in a better health state. The example of lung transplants may help to illustrate this. Patients eligible for a lung transplant normally are in such poor health states (i.e. low on a QALY scale), that a high cost per QALY is considered acceptable in order to provide them with a ‘last resort medicine’. This phenomenon may be especially pronounced for lifesaving interventions. In contrast, erectile dysfunction is generally considered to be a minor health problem (i.e. occurring high on a QALY scale).
Apparently, the severity of a condition, in terms of its absolute QALY score, determines (at least partly) the necessity of treatment. Necessity of treatment seems to increase when the patient is lower on the QALY scale. In current economic evaluations, however, QALYs are weighed equally regardless of the absolute position of patients on the QALY scale. Thus, only the number of QALYs gained determines priority, while in the decision-making process QALYs that are gained lower on the scale may be given more weight. QALYs are apparently weighed on the basis of necessity, leading to an acceptance of higher cost per QALY in more severe conditions. The process of weighing QALYs will (implicitly) change

the ranking of the different programs in ◘ **table 3.1**. The QALYs gained in lung transplants, for instance, may be considered highly necessary and therefore, be weighed three times higher than is currently done (Waugh and Scott, 1998), resulting in a higher priority for lung transplants. Such weighing seems to reflect an important social preference, as has been argued elsewhere. Especially, Erik Nord has devoted much attention to this phenomenon, trying to estimate the social weights for different QALY gains (Nord et al., 1996; Nord et al., 1999; Scitovsky, 1977). If we would use such weights in economic evaluation, perhaps, the results would better reflect societal preferences concerning the necessity of treatments. In the case of Viagra, this would mean that the QALY gains would receive a relatively low weight in the decision-making process, since the average patient scores relatively high on the QALY scale (increasing the cost-effectiveness ratio indicated in ◘ **table 3.1**). Note that in specific patient groups (e.g., patients with erectile dysfunction due to spinal cord injuries) such an argument does not hold and different conclusions about the necessity of treatment may be drawn.

3.3.2 Individual responsibility

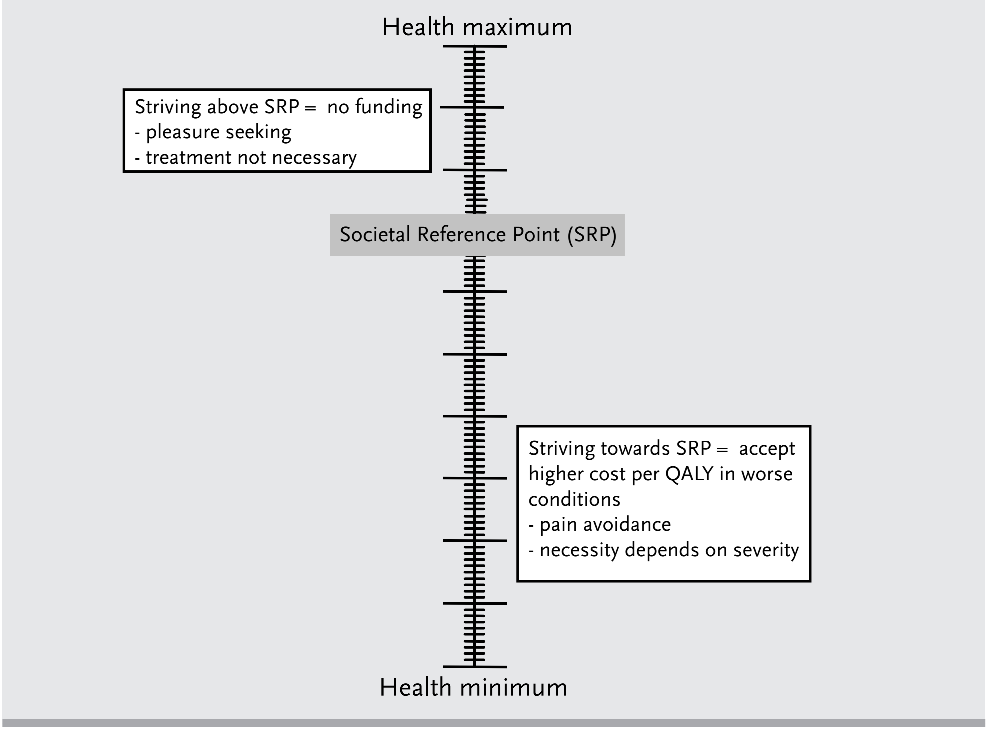
Even if the argument suggests that a condition at the upper end of the scale should have a more favourable cost per QALY to be eligible for treatment than interventions lower on the scale, it could still be argued that Viagra should be eligible for funding, since the costs per QALY are extraordinarily low. Another consideration, however, still might prohibit this funding, namely, the choice between individual and collective responsibility for the costs of a treatment.

The aforementioned argument that erectile dysfunction should be accepted as a part of normal ageing is another prevalent claim that argues against funding of Viagra. This argument seems to reflect some social reference point for health and functioning at certain ages or stages of life (Hayes et al., 1999). It implies that, while QALY gains high on the QALY scale may be considered less necessary than those low on the QALY scale, it may even be possible to determine certain thresholds, below which QALY gains are seen as necessary and above which, unnecessary. In these terms, the choice between individual and collective responsibility could depend on the question whether or not the health gain is viewed as 'pleasure seeking' (or 'luxury health care'), which may—like necessity—be related to the place on the QALY scale where the health gains occur.

To get a better understanding of these thresholds, the theory of Scitovsky could be helpful. The economist Scitovsky distinguishes between two types of utility gain: pleasure seeking and 'pain avoiding' (Scitovsky, 1977). It may be assumed that pleasure seeking in medicine, for instance cosmetic facelifts or liposuction, are those treatments that people want to undergo in order to strive above a (societal) reference point for health at a certain age or life stage. The societal reference point can be seen as some point below perfect health. Pleasure seeking treatments could be left to private responsibility and our collective responsibility refers to pain avoidance, which refers to treatments that are aimed at QALY gains in persons that are below their societal reference point of health and functioning. For instance, treatments to slow down cancer or treatment of hernia to reduce pain or other discomforts are likely to be classified as pain avoiding treatments.

The classification of Viagra as pleasure seeking explains the objection to collective funding of this medication and the urge to let patients pay for their own 'pleasures'. After all, it is likely that collective resource allocation to pain avoiding treatments is preferred over pleasure seeking treatments, e.g., because funding the latter in a situation where not all 'pain avoidance' is accounted for will only increase the individual differences in health status. Then, someone's eligibility for treatment may be defined as the difference between actual health state and a societal reference point, which can both be quantified in terms of quality of life. Note that also for pleasure seeking interventions (such as Viagra for erectile dysfunction) the burden of disease can still be expressed in terms of QALYs. Hence, pleasure seeking or pain avoiding is not defined by absence or presence of burden of illness; the classification of pleasure seeking is thus solely determined by the reference point.

♠ **Figure 3.1** Hypothetical decision-making framework



Different reference points may be used for different patient groups and these societal reference points may shift over time (presumably upwards), as for instance the availability of new treatments will change the perspective on diseases (in 15 or 20 years, having sex at an older age shall perhaps be perceived as normal). Also, society probably uses different reference points for health for the elderly than for younger persons (Williams, 1997; Tsuchiya, 2000; Busschbach et al., 1993), e.g., poorer health states are more acceptable in older persons than they are in younger persons. This is easily illustrated in terms of mobility; only being able to walk a maximum of 3 miles, slowly, may be accepted at the age of 80 but not at the age of 20. For Viagra, this could mean that erectile dysfunction is considered normal or acceptable in older patients, for whom treatment could therefore be classified as pleasure seeking, while it is not normal and unacceptable in younger patients (e.g. younger patients with erectile dysfunction due to diabetes), for whom treatment would be classified as pain avoidance. Finally, the reference points for health may be different for different dimensions of quality of life: experiencing pain may be considered equally unacceptable for everybody, while a decline in mobility after a certain age can be considered more acceptable. In ♠ **Figure 3.1**, this framework is summarised.

3.3.3 Other arguments

The choice between individual and collective responsibility for the costs of a treatment probably depends on three (related) elements. The first element has been discussed extensively above (consisting of the question whether or not the health gain can be seen as pleasure seeking, which is determined by the place on the QALY scale: above or beyond the reference point). However, whether or not to leave the costs of medication up to the patient is probably also related to the possibility to pay the related costs 'out of the patient's pocket' as well as upon whether or not the individual could have prevented the condition requiring the medical attention.

If the costs of an intervention are low, one may choose to shift these costs from the collective to the private responsibility. A major problem in such a shift is that there always will be patients for whom even low costs are hard to pay. Shifting costs would then create differences in health care use between higher and lower income groups, which is considered unacceptable in many countries. An important reason why such a shift could be considered acceptable regardless of the possible differences between income groups that will be a result from the shift, is that the intervention is not considered very important (or necessary). The reason is that by origin the health care system aims to protect people against unforeseeable events with large consequences (either in financial terms or in terms of health effects). The other way around, events with only small consequences do not appeal very much to our feelings of solidarity with those at risk. Also, one may choose not to be insured against relatively affordable (and preferably foreseeable) expenditures.

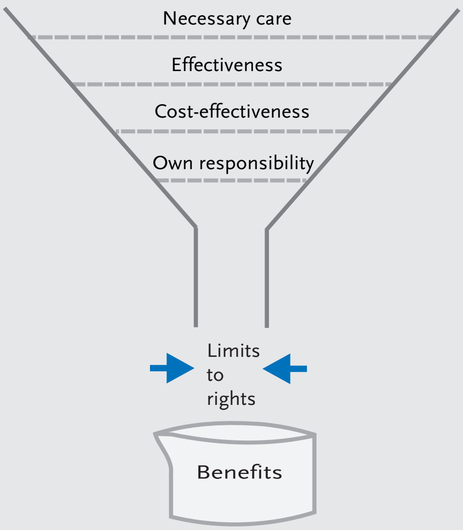
Another argument favouring individual responsibility is derived from the wish to penalise unhealthy behaviour. The origin of the condition can be important in cases where the patient is responsible for the condition. For example, society may be more inclined to pay for Viagra for somebody with an erectile dysfunction due to a spinal cord injury resulting from a traffic accident, than for someone with an erectile dysfunction due to excessive smoking and drinking. However, although this argument is heard frequently for many types of treatment (e.g. lung cancer due to smoking), it is very difficult to put into practice. Also, when the consequences of our actions are severe, it is questionable whether such an argument can or should be decisive.

3.4 Implications for decision making

The arguments regarding necessary care and individual responsibility, in combination with information about cost-effectiveness, seem to outline a decision-making framework in which QALY maximisation is not the only goal. Necessity and individual responsibility also play a role. Below we explore how these findings can change the Dutch model for choices in health care.

The criteria necessary care and individual responsibility have been discussed before. For example, they were put forward, together with effectiveness and cost-effectiveness in the so-called Dunning report (named after its chairman), in a theoretical framework for choices in health care that was developed in The Netherlands (Government Committee on Choices in Health Care, 1992). Basically, in the Dunning report, the four criteria

♠ **Figure 3.2** The Dunning system for priority setting



functioned as a funnel with four sieves, developed to separate care that should receive funding from care that should not be funded. Any intervention that does not make it through all the sieves will not be included in the basic benefit package of health care services (♠ **Figure 3.2**).

The Dunning report has played a key role in the Dutch discussion on rationing health care. Nevertheless, the success of the funnel in terms of removing services from the public health plan or keeping new services out of it has been limited. This may be a result of the poor operationalisation of the criteria necessary care and individual responsibility in the Dunning report. By not relating the criteria to real and measurable outcomes or quantities, working with them becomes rather difficult. Even more so, because all the sieves should lead to a yes or no answer for funding, while decision makers probably weigh the criteria, as we have argued above.

In the operationalisation of the criteria outlined here, the criteria necessary care and individual responsibility no longer function as sieves. Instead, decisions are mostly based on multiple criteria, together guiding the decision. Most interventions will be necessary to some extent (or for certain patient groups) and only those treatments used by patients to strive beyond a societal reference point, may be seen as unnecessary or, in other words, as lifestyle interventions. All other interventions are, in principle, eligible for funding, with necessity being related to the severity of a condition, and higher costs per QALY being allowed in patients who are more seriously ill.

3.5 Conclusions

It becomes apparent from the Viagra funding discussion, that cost-effectiveness information alone cannot provide sufficient information to guide reimbursement decisions. Economic considerations need to be supplemented with the criteria necessary care and individual responsibility. The way these criteria are used in the Viagra debate may suggest that QALY gains are valued differently, depending on the place on the QALY scale where they occur compared to some societal reference point for health. This could explain why funding of Viagra is being denied for several patient groups despite its favourable cost-effectiveness. In short, economic evaluation may benefit from incorporating especially the following preferences present in society:

- QALY gains should be weighed for severity.
- QALY gains should be weighed for necessity in terms of discrepancy between a reference point for health and real health state.

These and other mentioned reflections are a result of analysing the debate on funding only of one particular drug, namely Viagra and therefore, are only a first attempt to get a firmer grip on the arguments underlying funding debates and their relation to cost-effectiveness analyses. It will be interesting to see whether the same criteria will be applied in future debates on new drugs. An important feature of the Viagra case, which may blur the interpretation somewhat, is that this debate may not only reflect attitudes towards rationing in general, but may also incorporate some also 'Calvinistic' attitudes towards sexuality.

Valuing QALY gains differently on the basis of their place on the QALY scale is not common practice in formal economic evaluations, but it may already be performed implicitly by decision makers (Williams and Cookson, 2000). In this respect, it should also be noted that in the health economics and medical literature, there is increasing attention for alternatives for simple health maximisation in which all QALY gains are treated the same (Waugh and Scott, 1998; Brouwer and Van Hout, 1998).

Both the public and the political debate on rationing and the ongoing research efforts of scientists (from various disciplines) indicate the need for a more transparent way of rationing health care that is related to societal preferences for health gains in different patient groups. Developing a publicly supported framework for choices in health care, therefore, is an important goal, though difficult to achieve. In this paper, we have tried to indicate how (a part of) such a framework might be constructed, and indicated some implications of such a framework for the allocation of scarce resources in health care, if it would be consistently used. We hope that this paper may further stimulate the debate on transparent and consistent rationing in health care.

Part of Damien Hirst's The last supper

Chapter 4
Reconciliation of economic concerns and health policy. Illustration of an equity adjustment procedure using proportional shortfall

Economic evaluations have become an important and much used tool in aiding decision makers in deciding on reimbursement or implementation of new healthcare technologies. Nevertheless, the impact of economic evaluations on reimbursement decisions has been modest; results of economic evaluations do not have a good record in predicting funding decisions. This is usually explained in terms of fairness; there is increasing awareness that valuations of QALYs may differ when the QALYs accrue to different patients. The problem, however, is that these equity concerns often remain implicit, and therefore frustrate explicitness and transparency in evidence-based decision making.

Chapter 4

Reconciliation of Economic Concerns and Health Policy. Illustration of an Equity Adjustment Procedure Using Proportional Shortfall

Stolk EA, Van Donselaar G, Brouwer WBF, Busschbach JJV, Reconciliation of economic concerns and health policy: illustration of an equity adjustment procedure using proportional shortfall. PharmacoEconomics 2004;22(17):1097-107.

4.1 Introduction

An increasing number of countries (intend to) use the results of pharmacoeconomic research when making decisions regarding the reimbursement of drugs. Yet even in cost-effectiveness driven reimbursement systems, such as that for pharmaceutical products in Australia and the United Kingdom, no clear negative relationship exists between the level of cost-effectiveness and a positive reimbursement decision. Instead there is a large grey area where the level of cost-effectiveness is a poor predictor for reimbursement decisions of new pharmaceutical products (Rutten and Busschbach, 2001; George et al., 2001; Devlin and Parkin, 2004). An important reason for the relatively modest impact of economic evaluation on decision-making is that cost-effectiveness is not the only factor decision makers are concerned with. One additional aspect decision makers are concerned with is equity, which translates into policies that combat differences in health. In this perception, not all interventions are seen as equally necessary. Treatments that offer patients a last chance of surviving, are usually seen as types of very necessary care, whereas treatment of a relatively mild condition may get lower priority. These types of arguments seem to play an important role in funding decisions, but they often remain implicit and are commonly ignored in economic evaluations (Stolk et al., 2002).

Acknowledging that both efficiency and equity are objectives in health care, health economists have argued for the incorporation of ethical considerations in economic evaluations (Wagstaff, 1991; Nord, 1995; Williams, 1997). The underlying idea is that we need to balance equity and efficiency concerns, which can be done by estimating what loss in public health is considered acceptable if that would result in a more equitable distribution of health. In that way, a trade-off between equity and efficiency can be measured, which may be used to recalculate the value of QALY gains for different recipients. Although a vast body of evidence supports this theory (e.g. Nord et al.,1999; Lindholm et al., 1998), existing economic evaluations do not take these equity concerns into account. This may explain why cost-effectiveness analyses have had a limited impact on priority decisions. A vital condition for reconciliation of cost-effectiveness information and health policy is, therefore, that we formulate some agreed-upon measure of equity and explicate how efficiency and equity concerns should be balanced.

The aim of this chapter is to show how equity concerns can be incorporated into economic evaluations. We applied an equity adjustment procedure, whereby equity was defined in terms of 'proportional shortfall', a concept of equity which combines elements of the two popular equity approaches: fair innings and severity of illness. Next, we discuss some methodological choices in its implementation. Then, we discuss how health policy makers can use this concept and balance it with efficiency concerns. For that purpose the application of the equity-adjustment procedure is illustrated using ten real life interventions. We determined the rank ordering of ten conditions by proportional shortfall. We examined how a trade-off between equity and efficiency would affect funding decisions in the context of a basic benefits package of health services for a tax funded or social insurance based healthcare system, compared with the situation where the only criterion is efficiency.

4.2 How is equity usually defined?

Concerns about the worse-off portion of the population could be incorporated in economic evaluations through a weighting procedure of QALYs, as is illustrated in section 4.4 of this chapter. However, prior to this illustration the question 'who should we be concerned about first?' needs to be addressed. Who are the worse-off? This question is difficult to resolve, because the health achievements of people can be viewed from different perspectives. However, two approaches have attracted much attention: the severity of illness and the fair innings approaches.

The severity of illness approach assumes that the societal value of a health improvement is higher when the patient's initial condition is worse, all other things being equal. In this definition, 'initial health' concerns severity at the time of the intervention as well as the expected health in the case no treatment is provided (Nord, 2005). This approach embodies the feeling that people with severe conditions (e.g. facing immediate death or a severe handicap) must be rescued, whilst this urge to help declines when the acute health problems are less severe. Expressed in QALYs this means that unequal health prospects invoke questions of fairness. Form this point of view, a health improvement in patients with the worst no-treatment QALY-profile gets the highest value (Dolan and Olsen, 2001; Nord, 1999, p.30).

The fair innings approach appeals to the idea that everybody is entitled to a certain amount of life years or QALYs. This implies that health gains in people who get less than their fair innings should be valued higher than health gains in people who are expected to have had their fair innings or more. This approach is consistent with the frequently expressed preference that treatment of the young should get priority over treatment of the old, since indeed the young have not had a chance to enjoy a normal life span, whilst elderly people have already lived a major part of theirs (Williams, 1997). Expressed in QALYs this means that differential levels of total health are considered unfair and thus inequitable. Equity weights will then be higher for patients with a lower expected lifetime QALY total.

Because patients with the poorest no-treatment QALY profile in the future are not necessarily also the ones with the lowest expected lifetime QALY total, the severity of illness and fair innings approaches may produce different results for priority setting. ◘ **Table 4.1** illustrates this. Consider groups A and B, who have the same number of QALYs remaining but differ in age and the number of consumed QALYs. The severity of illness approach makes no difference between these groups, whilst the fair innings approach prioritises the younger group A because of their lower expected lifetime QALY total. Now consider groups C and D, who have the same expected lifetime QALY total, but are confronted with different illnesses at a different age. Fair innings would not distinguish between these groups, but severity of illness would prioritise group C over D because the latter has more QALYs remaining without treatment.

The different results for priority setting have evoked a debate about which of these approaches provides the best way of encapsulating equity concerns in economic evaluations. The trouble with the severity of illness approach may be that substantial differences in health prospects may exist because of different illnesses, but also because of age differences. Hence, unequal health prospects may not always be considered unfair and inequitable. But would we then agree with the way fair innings prioritises over age groups? Probably not; recent experiences in the Netherlands show that interventions with vital consequences have always been funded, for the young and the elderly, even in spite of unfavourable cost-effectiveness (Boer, 2001). Here severity, not age, seems decisive.

In summary, both approaches seem to touch upon relevant equity issues that are morally defensible and receive public support, but neither may fully reflect societal preferences. We therefore hypothesised that the equity concept of proportional shortfall better reflects societal preferences, because it combines elements from the severity of illness and the fair innings approach.

4.3 An intermediate position: proportional shortfall

Fair innings and severity of illness determine who is worse-off using a comparison of 'absolute' health outcomes in terms of total or future health. Alternatively, it could be suggested that comparison of badness of a condition is measured in proportional or relative terms (Sen, 1992, p.90; Clark, 1998; Cuadras-Morató et al., 2001; Johannesson, 2001). In this section we argue that such a proportionate equity concept may offer a way to balance the concerns of the fair innings and severity of illness approaches. For this purpose we describe one particular proportionate equity concept, which assumes that measurement of inequalities in health should concentrate on the fraction of QALYs that

▫ **Table 4.1** No-treatment QALY prospects and total QALY expectations for four hypothetical patient groups

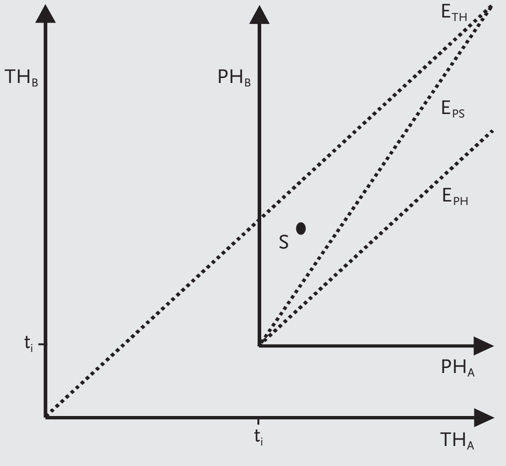
Patient group	QALYs consumed	QALYs remaining	QALY total
A	40	0	40
B	60	0	60
C	60	5	65
D	50	15	65

people lose relative to their remaining life expectancy, and not on the absolute number of QALYs lost or gained. We shall refer to this equity concept as ‘proportional shortfall’.

Proportional shortfall differs from fair innings and severity of illness in the sense that it compares individuals in relative terms to determine who is worse-off. The underlying reasoning is that proportional shortfall assumes that all individuals have the wish to reach the common target for health (as assumed in fair innings), but it also assumes that all individuals have the wish to maximise future health measured from the moment at which they fall ill (as assumed in severity of illness). Treating both concerns as equally important, the proportional shortfall approach makes a tradeoff between the two conflicting desiderata regarding equality in total and future health. In terms of QALYs, proportional shortfall thus suggests that equality exists when the relative distance of the two parties to the common target is the same. To determine who is worse-off one should therefore determine which patients lose the largest share of the QALY expectancy that she/he would have in normal health. Equity weights will then be a function of the proportion of health that a patient is going to lose relative to his or her remaining age-dependent QALY expectation in normal health.

If the proportional shortfall approach is used to encapsulate concerns for the worse-off, the young and the old may get an equal equity weight in spite of different health prospects or health losses. This happens when their QALY expectations without treatment (or QALY losses) represent an equal fraction of their remaining QALY expectancy in normal health. For example, all patients who face a threat of immediate death get equal equity weights,

♠ **Figure 4.1** Graphic representation depicting how an equitable distribution of health between two individuals A and B is defined differently by the severity-of illness, fair innings and proportional shortfall equity concepts



Health is measured in terms of QALYs.
S = the endowment point, i.e. the distribution that will result in case neither of the two patients gets treated.
E = equal, PH = prospective health, PS = proportional shortfall, TH = total health.

irrespective of their current age. After all, they will all lose 100% of their remaining life expectancy. Similarly, if a young patient loses 20 QALYs, and his normal QALY expectation was 40 QALYs, this patients gets the same equity weight as an older patient who loses 1 QALY and had a normal QALY expectation of 2 QALYs: both patients lose 50% of their remaining QALY expectancy. However, if two patients of the same age lose different amounts of QALYs, the highest priority goes to the one with the greatest loss (all else equal).

♠ **Figure 4.1** illustrates how an equitable distribution of health (e.g. measured in terms of QALYs) between two individuals A and B is defined differently by the severity of illness, fair innings and proportional shortfall equity concepts. The taxonomy of this figure is taken from Dolan and Olsen (2001). The axes TH_A en TH_B mark out the total cumulative health, while the subscales PH_A and PH_B mark out the prospective health space of the two individuals at the moment of intervention (t_i). From the location of t_i it is possible to infer that patient A has consumed more QALYs so far than patient B.

According to the fair innings approach an equitable distribution of health exists when persons A and B will have consumed the same amount of QALYs over their respective lifetimes. This is the case when the final distribution of QALYs is located on the line E_{TH} (equal total health) that runs from the origin to the health target. By contrast, the severity of illness approach suggests that an equitable distribution occurs when each individual gets the same amount of future QALYs, regardless of differences in the past. Thus a distribution is considered equitable if it is located at the line E_{PH} (equal prospective health). This approach obviously does not compensate for inequality in the situation at point ‘t_i’; hence total health outcomes may be different for the two patients. It has in common with the fair innings method that the lines E_{PH} and E_{TH} both have a 45° in the relevant evaluation spaces. The reason is that both approaches aim at equalising an absolute number of QALYs for the two patients, in terms of total or future health. Proportional shortfall on the other hand gives the same relative change in QALYs in the future health space the same weight, irrespective of the number of expected QALYs remaining. Therefore the line E_{PS}, which represents equality of proportional shortfall does not have a 45° angle relative to the axes. Instead, it runs from the ‘t_i’ point (analogously to severity of illness) to the health target (as in the fair innings approach).

To demonstrate how ♠ **figure 4.1** may assist in priority setting we have added the endowment point ‘S’ to the figure, which represents the distribution that will result in case neither of the two patients gets treated. The location of S relative to the equity lines indicates how the different equity concepts would value health gains for patients A and B. For example, at point S patient A has consumed more QALYs than patient B; the fair innings argument would therefore give the highest value to the health gains of patient B. By contrast, measured from ‘t_i’ patient B has a better no-treatment QALY-profile. Therefore, using severity of illness the health gains of patient A get the highest weight. If equity is defined in terms of proportional shortfall, again the health gains of patient A will receive the highest weight, because without treatment patient A loses a greater fraction of their remaining QALY expectancy than patient B. Note that these values only concern equity weights and not directly translate into treatment priorities, because then also other factors should be taken into account, like the size of the health gain.

Finally, some comments on ♠ **figure 4.1** are warranted. First, for simplicity ♠ **figure 4.1** assumes that the health target is fixed. This may be considered appropriate in the context of the fair innings approach. The alternative would be a dynamic target level, implying that the target for fair innings is raised when an individual ages. Such an approach, however, would benefit those who are old more than the young and therefore seems in conflict with the objective of fair innings (Williams, 1997; Williams, 2001). However, for proportional shortfall a dynamic health target may be more appropriate, since the aim of proportional shortfall is to respect the wish of all people to live out their remaining life expectancy irrespective of their current age. This implies that the use age-dependent remaining QALY expectations may be more appropriate to define the target level for health. Secondly, it should be noted that in this paper proportional shortfall is simply used to identify who is worse-off. We do not assume any particular definition of the equity weights that can be derived from proportional shortfall. In that respect our approach differs from the approach of Johannesson, who described an approach based on aggregating relative QALYs (Johannesson, 2001). Johannesson assumed a particular set of weights that reflects proportionate inequalities in health. However, it is not obvious that equity weights should be calculated in this manner. Equity weights may not only differ according to the relative size of the health gap, but also according to other characteristics (e.g. their nature, who is affected by them (Williams, 2001) or the distribution of health in a population (Bleichrodt et al., 2004)). It is therefore better to investigate empirically how the weights relate to the proportional health gains or losses. Finally, it is commonly accepted that violations of the Pareto criterion should be avoided. This also holds if an equity adjustment procedure is used to determine the desirability of different resource allocations. A Pareto improvement means that only changes are allowed that can make at least one individual better off, without making any other individual worse-off. Hence, if additional resources will be allocated, only the distributions to the northeast of S are admissible, since they improve the position of A and/or B.

4.4 From theory to practice

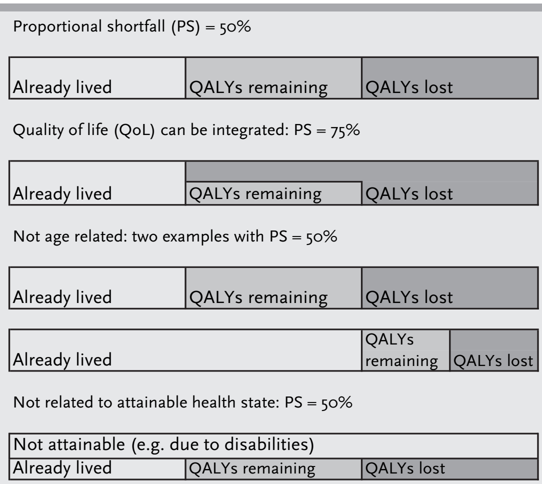
This section illustrates how the application of proportional shortfall would affect priority setting using real life examples. First, however, some issues related to the measurement of proportional shortfall should be discussed.

Proportional shortfall can be measured on a scale from 0.0 (no health loss) to 1.0 (maximal health loss, or immediate death) using ♣ **equation 4.1**. It measures the fraction of the remaining health expectancy that a patient will lose because of the condition for which treatment is considered. In ♣ **equation 4.1** (where PS = proportional shortfall), the denominator reflects the remaining QALY expectation in normal health, which could be determined using age- and sex-specific mortality and quality of life tables (using a dynamic target). The QALY loss caused by a particular condition can be determined by deducting a patient's remaining QALY expectancy without treatment from the remaining QALY expectancy that the patient would have had in normal health. ♠ **Figure 4.2** gives some examples of the calculation of proportional shortfall.

$$PS = \frac{\text{disease related QALY loss}}{\text{QALYs remaining in absence of the disease}} \quad ♣ \; (4.1)$$

In the measurement of proportional shortfall (or other equity concepts) difficulties arise when the measurement concerns the health loss associated with health risks for which preventive treatments are considered. Imagine for example that we were to give preventive medication to all 40-year-olds with hypertension; 90% of these people would not suffer negative health effects without treatment, but a premature death occurs in 10% of the patients at age 60 years. The question then becomes in what subgroup proportional shortfall should be measured: a first option is to measure shortfall only in the patients who actually experience negative health effects. Their remaining life expectancy would be 40 years in absence of the disease (from 40 to 80 years), but they lose 20 years. Thus

♠ **Figure 4.2** Examples in the calculation of proportional shortfall (PS)



PS is defined as the proportion of QALYs that a patient may lose due to illness. Total QALY expectancy in normal health is represented by the full bar, including the sections 'already lived', 'QALYs remaining' and 'QALYs lost'. Quality of life is represented by the height of the bar to illustrate impairments due to the illness for which treatment is considered or because of permanent disabilities.

their proportional shortfall is 50% (20/40), ignoring quality of life losses for the moment. If we also include the proportional shortfall of those patients who were unnecessarily treated because they would not have fallen ill without prevention, the average proportional shortfall is of course smaller. Average proportional shortfall would be 5%, which is the weighted mean of 90% of the patients who have a relative health loss of 0%, and 10% of the patients who lose 50%.

When measuring the cost-utility of a preventive intervention, it is obviously relevant to consider all patients who are being treated. In the measurement of proportional shortfall this may not be the case. An argument against considering all patients is that it would impose double punishment on patients in need of preventive treatment. Already the cost-utility ratio of their treatments is less favourable because those who would actually benefit from the treatment cannot be identified, and therefore already the costs of all patients who received preventive treatment are related to the benefits of the few who would have experienced health loss. It seems unfair that the same treatment characteristic would also result in lower equity weights. And fairness is exactly what we should be concerned about here.

Because equity measures intend to capture our concerns for the worse-off, it seems appropriate to ask ourselves what are the negative health effects that we are trying to prevent? It is clear that preventive treatment of hypertension is not provided to alleviate the direct quality of life effects of hypertension, but indeed to prevent premature deaths caused by cardiovascular events. A possible solution therefore is to focus the equity measurement on the group who actually experience cardiovascular events and measure average proportional shortfall only in this subgroup. This measure then is independent of the prevalence of cardiovascular events in a particular group, and of diagnostic accuracy in identifying the patients at the highest risk. This approach makes sense because our concerns for the worse-off exist independently of the health gains that can be achieved by treatment. Appropriateness of this solution, however, should be a matter of debate.

Table 4.2 Proportional shortfall caused by pneumococcal (PC) pneumonia: Effect of methodological choices

	All persons aged ≥ 65 years	All who eventually get pneumonia	All who eventually get PC pneumonia	All who eventually die from PC pneumonia	All facing imminent death due to PC pneumonia
65-74	0.000199	0.0486	0.122	0.85	1
75-84	0.000359	0.0388	0.095	0.77	1
85+	0.000524	0.0269	0.067	0.62	1
Total	0.000238	0.0328	0.082		1

A related question is when the measurement of proportional shortfall should begin: at the moment preventive treatment is first considered, or at the moment that health problems actually occur? As patients are likely to feel very different about acute death, than about a predicted death over 20 years, we considered it more appropriate to measure proportional shortfall in relation to the moment that treatment is considered. This means that the proportional shortfall for a particular condition differs according to the treatment is provided (preventive or curative). Imagine a patient who suffers from a life threatening condition at the age of 60 years and who expected to live until 80 years. If a curative intervention at the age of 60 years is considered, this approach would assume a proportional shortfall of 100% (the patient may lose 20 out of 20 years). If a preventive treatment would be considered at 40 years of age, then the proportional shortfall is 50% (the patient may lose 20 out of 40 years).

◘ **Table 4.2** illustrates both issues discussed above, using the example of pneumococcal pneumonia, by showing how the estimate of proportional shortfall is affected when the health loss associated with pneumococcal infection is ascribed to different subgroups. Obviously, the proportional shortfall increases when the health loss is ascribed to a smaller and more affected population of patients, or evaluated from a shorter timeframe. The highlighted column gives what we consider the most appropriate interpretation. However, our choices can be disputed and policymakers rather than analysts should perhaps make some of these choices.

◘ **Table 4.2** emphasises that it is insufficient to scrutinize just the general ideas behind an equity concept and that just as much attention should be directed to the methodological choices in adopting each one of them. Although the choices were based on a well defined conceptual framework, there is no scientific reason to exclude other ones. There is a need to examine whether this operationalisation of proportional shortfall reflects societal preferences. To examine the potential of a theoretical concept, it is common to illustrate using hypothetical examples like the ones presented in ◘ **table 4.1**. The ultimate test of the practical value of the concept is, however, a test with real-life examples of studies used in reimbursement decisions. Then we can see if indeed an equity-adjustment procedure improves our understanding of previous such decisions.

4.5 Using an equity-weighted cost-effectiveness threshold

To illustrate the effect of equity adjustment procedures based on proportional shortfall, we calculated the proportional shortfall for ten real-life conditions and examined how a trade-off between proportional shortfall and efficiency would affect funding decisions in the context of a basic benefits package of health services for a tax funded or social insurance based healthcare system, compared with the situation where the only criterion is efficiency. Information about the QALY profiles was available from cost-utility studies. On

Table 4.3 Rank ordering of 10 conditions by PS determined on the basis of the no-treatment QALY profile

Condition	Intervention	Age	LE	HLY	YLD	QoL	LYL	PS
Non-Hodgkin Lymphoma	CHOP-chemotherapy	73	11.5	0	0.5	-50%	11	0.97
Pulmonary hypertension	Lung transplant	40	42	0	3	-57%	39	0.96
Pneumococcal pneumonia	Vaccination	70	13.75	2	<0.25	-38%	11.5	0.82
COPD	Ipatropium	60	20	0	11	-32%	9	0.61
Artherosclerosis	Clopidogrel	50	28.5	8	8	-40%	12.5	0.55
High Cholesterol	Statins	50	30	13	7.5	-21%	9.5	0.28
Hypertension	Antihypertensives	55	31	15.5	7	-25%	8.5	0.26
Symptomatic BPH	Finasteride	70	14	0	14	-8%	0	0.09
Osteoporosis	Oestrogen	70	14.8*	0	14.5	-11%	0.3	0.08
Onychomycosis	Terbinafine	70	14	0	14	-2%	0	0.02

* This estimate differs from other estimates for 70-years-old, because this population consists of women

The interventions are listed because estimates of PS are specific to the treatment that is considered. **LE** = Life expectancy in normal health, **HLY** = health life years, **YLD** = years lived with disabilities, **QoL** = Quality of life, **LYL** = life years lost, **PS** = proportional shortfall, **CHOP** = cyclophosphamide, doxorubicin, vincristine, prednisone

table

our request, the researchers who had been involved in the various economic evaluations reprocessed their data to calculate proportional shortfall. A report of these calculations has been published in Dutch by Stolk et al. (2002). ◘ **Table 4.3** summarizes these data and gives the estimates of proportional shortfall.

Assuming a constant resource constraint, ♠ **figure 4.3** can be used to illustrate how an equity-adjustment procedure would affect reimbursement decisions. This figure includes two decision rules based on economic evaluations, a constant cost-effectiveness threshold and one with equity weighting. In that way, this figure helps us gain insight into the type of interventions that are more or less likely to receive funding using either of the decision rules. The figure plots the proportional shortfall of ten conditions on the X-axis and the cost-utility ratios of corresponding treatments on the Y-axis. Except for treatment of pulmonary hypertension, the cost-utility ratios are all very similar (€20,000 per QALY or less). The proportional shortfall caused by the ten conditions differs substantially, and therefore these datapoints provide a good opportunity to explore the possible consequences of equity weighting for health care reimbursement decisions.

Assuming that no equity weighting procedure is used, the cost-effectiveness threshold would be the same for all interventions: all interventions that are more cost-effective than a certain threshold would be implemented. Where exactly the cut-off point should be to warrant adoption of a technology is a matter of debate. In the figure we used the hypothetical cut off point of about €20,000 per QALY. If this was the threshold value for cost-utility ratios then treatments for all conditions in the figure would be funded except treatment for treatment of pulmonary hypertension.

The decision on several of the treatments might change if an equity weighting procedure was applied, such as lowering the cost-effectiveness threshold when a condition is worse and raising it when a condition is relatively mild. This approach is similar to an equity weighting procedure of QALYs. However, differentiation of the cost-effectiveness threshold has an advantage over equity adjustment of QALYs, because it keeps the equity and efficiency profiles more explicit. The curved line in ♠ **figure 4.3** represents this equity-adjusted decision rule. Application of this decision rule would change the reimbursement decisions for treatments with a cost-utility ratio below (above) the curved line but above (below) the horizontal line. Resources for treatment of the worse-off patients could be expanded (unfortunately the included data points give no example), because less is invested in patients who are in relatively good health already (e.g. patients suffering from benign prostatic hyperplasia, osteoporosis, or onychomycosis).

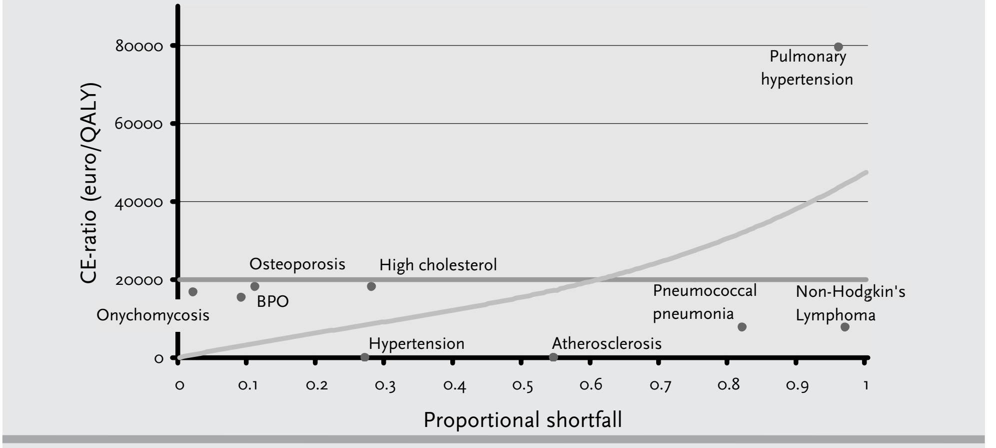
Note that ♠ **figure 4.3** only provides tentative conclusions. As explained in previous sections, we believe that there are good reasons to use proportional shortfall, but as yet this choice is not supported with empirical evidence. If society's equity concerns are better reflected by fair innings, the ordering of the ten conditions on the X-axis would change and different conclusions would be reached. The observations about specific treatments have also little relevance when other cut off points are used. Some explanation of the way we constructed the thresholds in ♠ **figure 4.3** is therefore in order.

The fixed cut off point of €20.000 per QALY was chosen because two Dutch clinical guidelines applied this threshold (osteoporosis (CBO, 2001) and high cholesterol (CBO, 1998)). The curved line was constructed to explain observations from the Dutch reimbursement debates, such as the negative funding decision for treatment of onychomycosis and the positive decision for lung transplants in patients with pulmonary hypertension. We also took into account the expressed concern that treatment schemes for high cholesterol, osteoporosis and benign prostatic hyperplasia cover too many patients with low risks to offer value for money. These concerns were taken to give a tentative idea of societal preferences for equity weighting. However, it could as well be hypothesised that the curved line runs flatter or steeper, or looks like an S-shaped curve. As yet, there is no evidence to support any specific weighting scheme. This is an area for further research.

4.6 Discussion

The aim of this paper has been twofold. First, we have described a proportionate equity concept for prioritising health care interventions, referred to as the concept of proportional shortfall. This equity concept combines elements of two popular but conflicting notions of equity: fair innings and severity of illness. Proportional shortfall assumes that these two notions can be balanced together on the basis of an underlying principle: the wish

♠ **Figure 4.3** Schematic representation depicting how an equity-adjustment procedure would affect reimbursement decisions compared with a constant cost-effectiveness threshold



BPO = benign prostatic obstruction; **CE** = costeffectiveness.

to equalise persons within their own scope of potential for health. Secondly, we have illustrated how equity adjustment procedures could be applied. To that end, we measured proportional shortfall for ten conditions and explored the implications of a trade-off between equity and efficiency in priority setting. Considering real-life interventions and providing tentative conclusions about which treatments are more or less likely to receive funding, we scrutinised the effects of equity-adjustment procedures. The results suggest that the integration of equity concerns into an economic evaluation improves the fit between economic models and reimbursement decisions.

The appeal of the equity concept of proportional shortfall lies partly in its consistency with past healthcare decisions. Noting that interventions with vital consequences for the elderly have always been funded in The Netherlands, even in spite of unfavourable cost-effectiveness, we have argued that the fair innings argument discriminates against the elderly more strongly than policy makers seem to prefer. Since the concept of proportional shortfall partly alleviates the age discrimination implied in the fair innings argument, we hypothesised that it may be a better guide for healthcare policy makers. Nevertheless, several investigations of social preferences regarding the role of age in the distribution of resources suggest that society prefers to allocate resources to the young (Busschbach et al., 1993; Cookson and Dolan, 1999). It is therefore too soon to claim superiority of this (or any other) equity concept yet. Rather than observing the differences, we therefore need to test empirically which equity concept best reflects societal preferences for equity. Such a test has been performed, but the results were inconclusive (Stolk, in press).

Suppose that society agrees with the broad picture that we have provided, that the equity concept should balance our concerns about equality in total health and severity of illness, we could still question whether proportional shortfall offers an appropriate solution. Proportional shortfall assumes a particular trade-off between the two desiderata regarding equality in total and severity of illness. However, there is no specific reason to assume that this trade-off accurately describes societal preferences. People might just as well give more weight to the fair innings than to the health prospects, or vice versa. Indeed, our calculations of proportional shortfall have implications that need to be discussed and tested for empirical support. For example, when a 20-year-old patient who may lose 27 out of an expected 60 QALYs is compared with a 78-year-old patient who may lose three out of six QALYs in normal health, the old patient gets a higher equity weight (proportional shortfall is 0.45 versus 0.50). Not everybody might agree with the consequent priority implications. Age weights may be considered a potential solution to this problem (Williams, 1997; Johannesson, 2001), or perhaps a more complex relation should be modelled between the absolute number of QALYs that are lost and the QALY expectation. For example, the line E_{PS} in ♠ **figure 4.1** may be curved instead of straight.

Moreover, trade-offs may not even be constant, but rather vary in different circumstances, and non-health information may influence trade-offs as well. Such reservations emphasise the need for the scrutiny and comparison of equity concepts.

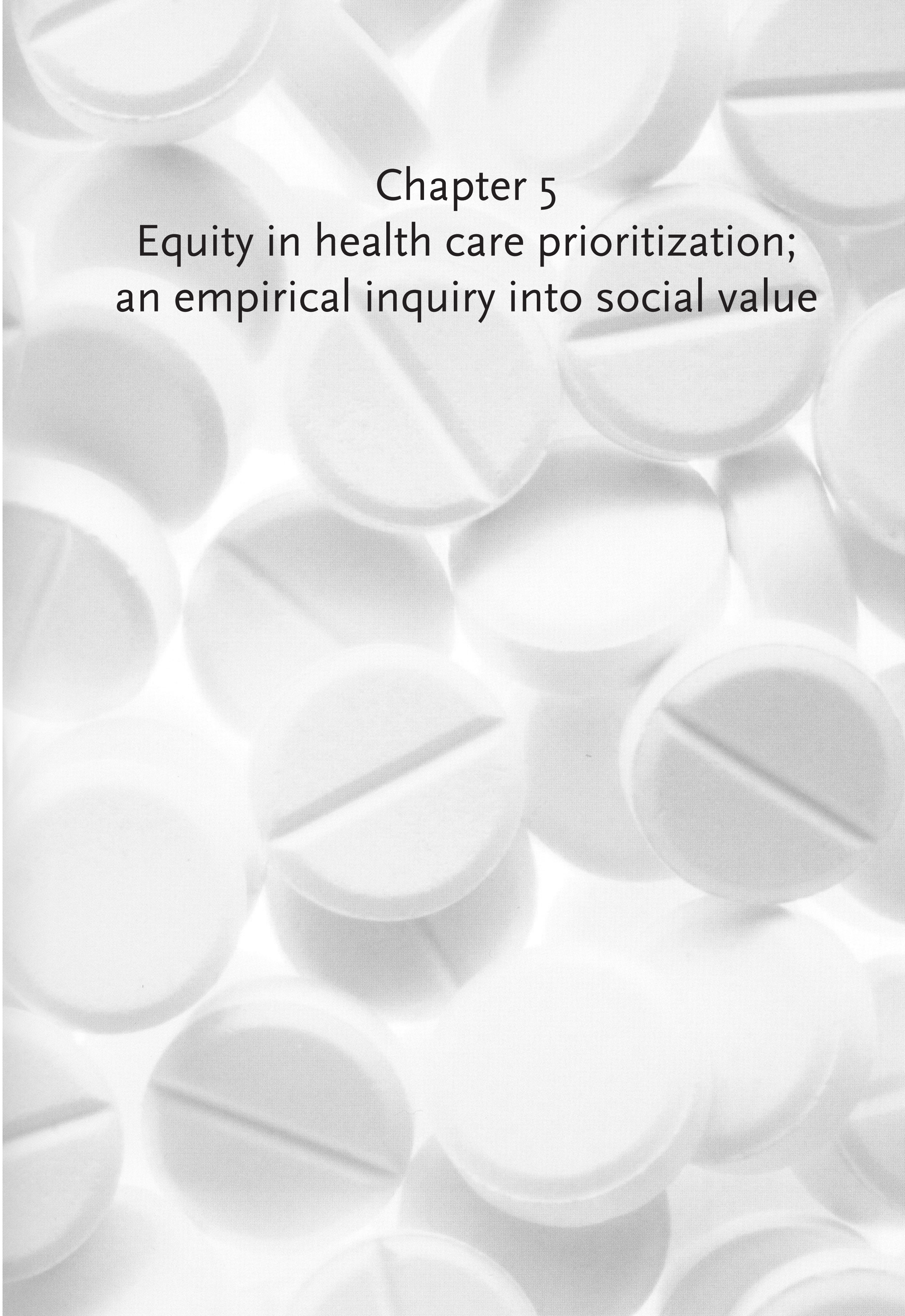
Other refinements may also be desirable. For example, life expectancy differs between males and females or between socioeconomic groups. Should this be reflected in the measurement of proportional shortfall by measuring shortfall relatively to a gender or socioeconomic class dependent life expectancy? Just like the choice of subgroups in which shortfall was measured, this methodological choice has distributional implications. The debate on equity weighting must therefore not be limited to a discussion of the general equity concept, but should also be directed at specifics. The implications of the application of the equity concept of proportional shortfall using real-life interventions presented in this paper are important in that respect.

4.7 Conclusion

This paper has described the equity concept of proportional shortfall and has demonstrated its implications. Measurable interpretations of equity such as the concept of proportional shortfall make it possible to balance equity concerns with concerns regarding the efficiency of the allocation of health care resources. The challenge for the near future will be to test how well this equity adjustment procedure reflects social preferences for the distribution of healthcare resources.

Part of Damien Hirst's Pain killers

Chapter 5
Equity in health care prioritization; an empirical inquiry into social value



The value of QALY gains for different patients may be recalculated using equity weights, but it is unclear which interpretation of equity should be used: severity of illness, fair innings or proportional shortfall. We set up an experiment to analyze which of these equity concepts best reflects people's distributional preferences. Sixty respondents assigned a priority rank to the treatment of ten conditions using the paired comparison technique. We described these real-life conditions by their actual QALY profiles, i.e. in terms of age, disease free period, duration of disease, quality of life, and life years lost. Next we determined the priority rank order of the ten conditions by the three equity concepts, using the weights that each equity concept attributes to the different units of the QALY profile describing the 10 conditions. To explore the social interpretation of equity, we compared the observed and theoretical rank orderings using Spearman correlations. All correlations were significant at a 0.05 level. Fair innings best predicted the observed rank order of the 10 conditions (r=0.95). Weaker correlations were found for proportional shortfall (r=0.82) and severity of illness (r=-0.65). This result calls attention to health policy, because actual health care decisions often reflect concerns of severity of illness. This raises the question if health care decision makers evaluate the claims of different patients for health care by appropriate criteria.

Chapter 5

Equity in health care prioritization; an empirical inquiry into social value

Stolk EA, Pickee SJ, Ament AHJA, Busschbach JJV, Equity in health care prioritization; an empirical inquiry into social value. Health Policy (in press)

5.1 Introduction

The concept of efficiency in the distribution of health is usually taken to imply that people who can gain QALYs in a relatively cheap way are more entitled to treatment than other people are. Societal preferences, however, have demonstrated that other individual characteristics may also affect priority setting, because the principle of health maximization may lead to uneven distributions of health which conflicts with people's equity concerns. To incorporate social concerns about equality in health in economic evaluations several authors have advocated operationalization of an equity-efficiency trade-off (Wagstaff, 1991; Williams, 1997; Nord et al., 1999). This trade-off reflects the willingness to decrease the total amount of benefits from our health care system, if this results in a more equitable distribution of health effects. Unfortunately, equity is an ethical concept that has no precise definition. Ever since the concept of the equity-efficiency trade-off has been introduced, it has been debated which equity concern(s) should be included in the trade-off. To contribute to this debate, we set up an empirical study to explore the social interpretation of equity.

There seems to be agreement that the definition of equity should be found in the health domain, but it is unclear what kind of measure of inequalities in health must be defined. Different authors have tried to persuade others to the use of different equity concepts, like severity of illness (Nord, 1995) and fair innings (Williams, 1997). The severity of illness approach embodies the feeling that people facing severe illness must be rescued, whilst this urge to help declines when the health conditions are less severe (Nord et al., 1999). Therefore, patients in the most critical condition receive the highest priority, e.g. patients facing the threat of immediate death or a severe handicap. In effect, this approach gives highest priority to patients with the poorest health prospects without treatment. There is however no consensus that indeed patients with the poorest health prospects are always the most deserving (Dolan and Olsen, 2001). People who adhere to the fair innings argument believe that the goal of equity adjustment should be to reduce differences in lifetime experience of health instead of reducing differences in future health (Williams, 1997). Not only people's health prospects are then relevant to evaluate their claim on health care, but also the amount of health that they have already consumed. This approach implies that in many cases the young people should get priority over the old, as the old has already had more time alive than the young (and presumably not in such a bad health state that the young will have consumed more QALYs).

Empirical studies revealed public support for both severity of illness and fair innings (Tsuchiya, 1999; Dolan, 1998; Cookson and Dolan, 1999). Therefore it is also possible to argue for equity concepts that combine the two principles together (Cookson and Dolan, 2000). A combination concept may take an intermediate position and help to clarify how priorities are set in the case of a conflict between severity of illness and fair innings.

Recognizing this possibility, Johannesson (2001) and Stolk (2004) have described an equity concept that takes an intermediate position: the concept of proportional shortfall. Proportional shortfall makes a particular trade-off between the goals regarding equality in total and future health. Proportional shortfall has in common with fair innings that the size of the health gap is relevant, but it agrees with severity of illness that also the remaining no-treatment QALY expectation should be taken into account. From this viewpoint equity weights are not simply proportional to the absolute size of the health gap caused by a condition. Rather, equity weights should be determined on the basis of the amount of QALYs that a patient loses proportional to this person's remaining QALY expectancy in normal health (e.g. calculated as the average expected number of QALYs for the population of that age and sex). Higher equity weights then apply if a patient loses a greater fraction of his or her remaining QALY expectation. Proportional shortfall thus values relative changes in expected QALYs, irrespective of the number of expected QALYs concerned (Johannesson, 2001). This reflects the idea that everyone is equally entitled to live out his or her remaining life span, no matter whether the remaining life span is long or short.

In the literature on inequalities such proportionate equity concepts have been discussed frequently (Cuadras-Morató et al., 2001; Sen, 1992; Clark, 1998; Williams and Cookson, 2000), but little is known about the social support for this type of combination principles. Usually the two equity concerns that are combined are evaluated separately. In a recent paper, Cuadras-Morató et al. compared support for the absolute and proportionate equity concepts using axiomatic bargaining theory (2001). Cuadras-Morató et al. recruited respondents to solve resource allocations, whereby the possible solutions specified shares of the available budget that would be allocated to different patients. The solutions represented six different distributive and equity concepts among which the utilitarian position, the fair innings argument and a proportionate equity concept similar to the proportional shortfall approach. Respondents had to indicate which solution they found most attractive. In that way this study explored what equity concept prevailed in circumstances where different views would result in different priorities (Cuadras-Morató et al., 2001). This experiment found no dominant principle, but strongest support was for the proportional solution and fair innings. Which of these two solutions was preferred in the different situations depended on the differences in the capacity to benefit, the health gap, and the context. The authors conclude that more research into social support for equity concepts is warranted, and they advise to explore benefits of realistic examples in future surveys to study on a less abstract level support for different equity concepts.

Building on the studies discussed above we further explored social preferences for equity. In our study we used realistic cases to test support for different equity concepts, as suggested by Cuadras-Morato et al. (2001). We asked our respondents to priority rank treatments of ten conditions using the paired comparison technique. To explore the social interpretation of equity, we compared this observed rank order to the rank orderings expected by the three equity concepts. This study contributes to existing literature by concentrating on the way in which people balance different equity concepts in a series of forced choice questions. The purpose is to see whether combined information on the different choices reveals the underlying decision process and weighting of equity concerns.

5.2 Methods

5.2.1 Respondents

We recruited a heterogeneous sample of students, researchers, and health policy makers (N=65). Students and researchers were recruited at the departments of Health Sciences of the Erasmus University in Rotterdam and the University of Maastricht. Health policy makers were employed at the Dutch Healthcare Insurance Board.

5.2.2 Paired comparison scaling

Respondents had to priority rank ten conditions using the paired comparison technique. The choices were presented on cards in random order (see for an example ♠ **Figure 5.1**). Respondents were asked to indicate which of two patients in a different condition they would treat, if resources were lacking to provide both patients the treatment that they needed. Each respondent compared each condition to each of the remaining conditions, which means that they had 45 choices to make (n(n-1)/2). This design assumes symmetry, i.e. it assumed that asking a person whether A would be preferred to B gives the same result as asking whether B would be preferred to A. Although this assumption not necessarily needs to hold, assuming symmetry was desirable as not to double the number of comparisons that needed to be made. We considered 45 comparisons a feasible number for each respondent to make. Many studies have employed at least thirty-two choices successfully and recent research suggested that an experiment might include over 40 comparisons (Louviere, 2000, p. 134). Moreover, the task was cognitively not complex, because respondents were familiar with issues of health care priority setting, and because many of the choices were straightforward in the sense that they contained a dominant alternative (better in all respects).

To introduce the paired comparison task, we described a hypothetical context in which the respondent would possess a wonder pill, which cures any patient who receives it. Unfortunately, in each pair of options two patients are in need of treatment, but only one pill is available. The respondent had to indicate which patient s/he would give priority. The wonder pill would relieve that patient of all described health problems and bring this person back to normal health; the other patient would not be cured and would have to endure the illness.

We determined a scale value for each of the ten conditions that reflected their position compared to other conditions on the priority scale. To obtain these scale values, the paired comparison scaling technique assumes that the proportion of times that any alternative is chosen over any other alternative reflects discriminal dispersions of the alternatives on the continuum of an underlying unidimensional construct (h.l. priority rank). To determine scale values using paired comparison data, first a matrix must be constructed with the probability that each condition is prioritized over any other. Using the properties of the normal curve, these probabilities can be converted into z-values, corresponding to their location on a normal distribution. These z-scores can be interpreted as measures of priority at interval level. If no extreme values of p (equal to 1 or 0) would be present, the mean z-value for each condition gives its scale value directly. If extreme values do occur, these must be ignored, since the corresponding z-scores approach infinity. The data then should be analyzed using the Thurstone Case V Scaling Model for Incomplete data. See Edwards (1957) for a detailed description of paired comparison scaling.

5.2.3 Health state descriptions

In the paired comparison task we presented realistic decomposed QALY profiles of ten conditions. A QALY profile describes how quality of life develops over time. The specification of these QALY profiles was essential in the context of our experiment,

♠ **Figure 5.1** Example of a choice in the paired comparison task

High cholesterol		**COPD** (narrowed bronchia and damaged lungs)	
Average age	50 year	Average age	60 year
Disease free	13 years	Disease free	0 years
Years lived in disability	7,5 years	Years lived in disability	11 years
Quality of life	-21%	Quality of life	-32%
Life years lost	9,5 years	Life years lost	9 years
Many people with high cholesterol hardly notice this, but they are at risk of cardiovascular events (myocardial infarction, stroke). On average cardiac events happen 13 years after diagnosis. When a cardiac event occurs, patients may experience a few mild effects during the rest of their lives. Life expectancy is shortened.		COPD patients are short of breath and tight in the chest and will often develop chronic bronchitis. The breathing problems can take serious forms and is eventually life-threatening. 28	

because people adhering to different equity concepts will attribute different values to different characteristics of a patient's QALY profile. The presented QALY profiles included all attributes that were relevant according to any of the equity concepts. Hence we provided information about the no-treatment QALY profile (to allow for rank ordering according to severity of illness), and the QALYs foregone measured from a predetermined age-related target level for health (allowing for rank ordering according to fair innings). The proportional shortfall combines the two and required no additional factors to be included in the health state description. Accordingly, the health states were described using the 5 units listed below.

a) mean age of the patients (because that determines the target level for health)
b) time without disability
c) time with complaints
d) average quality of life loss associated with the complaints
e) life years lost

We included ten real-life conditions for which the QALY profile had been established in the paired comparison task, because we assumed that people have clearer preferences for real-life cases. The use of real-life examples therefore might increase the predictive value of the study outcomes for actual decisions in real life. Data regarding the no-treatment QALY profiles of the ten conditions were available from clinical and economic investigations that were performed in The Netherlands. The original researchers were involved in the process of deriving the correct estimates from the models. These data are presented in ◘ **table 5.1**. In addition, the health states were labelled and offered a general description of the health problem, again because we assumed that people's preferences are clearer when explicit labels are used rather than neutral and abstract descriptions of a health state. We learnt from a pilot that the use of labels and descriptions did not introduce biases or framing effects: this pilot in 20 students confirmed that the observed rank order was unaffected by the presentation of labels and health state descriptions.

5.2.4 Establishing the theoretical rank orderings

◘ **Table 5.1** also presents the three theoretical rank orderings and the scores of the ten conditions on the equity scales, which were determined using the weights that each equity concept attributes to the different units of the QALY profile describing the ten conditions. Below we discuss the specifics of these calculations. Note that we simplified the calculations by assuming that no quality of life losses would occur due to other illnesses: the age of a patient equalled the number of QALYs consumed so far, and each life year lost equalled a QALY.

Table 5.1 Scaling by equity concept (rank)

	Parameters					SoI		FIA 70		FIA dynamic		PS	
Condition*	Age	HLY	DLY	QoL	LLY	Scaling	Rank	Scaling	Rank	Scaling	Rank	Scaling	Rank
Onychomycosis	60	0	20	0.98	0	19.60	9	-0.14	8	0.01	10	0.02	10
BPO	70	0	14	0.92	0	12.88	6	-0.18	9	0.01	9	0.08	9
Osteoporosis	70	0	14.5	0.89	0.3	12.91	7	-0.18	10	0.02	8	0.13	8
High Bloodpressure	55	15.5	7	0.75	8.5	20.75	10	-0.08	7	0.12	7	0.33	7
Pneumococcal pneumonia	70	2	0.25	0.62	11.5	2.16	3	-0.03	6	0.14	4	0.84	3
High cholesterol	50	13	7.5	0.79	9.5	18.93	8	0.02	4	0.14	5	0.37	6
Non-Hodgkin Lymphoma	70	0	0.5	0.50	11	0.25	1	0.00	5	0.14	6	0.98	1
COPD	60	0	11	0.68	9	7.48	4	0.04	3	0.16	3	0.63	4
Artheriosclerosis	50	8	8	0.60	12.5	12.80	5	0.10	2	0.20	2	0.55	5
Pulmonary hypertension	40	0	3	0.43	39	1.29	2	0.41	1	0.50	1	0.97	2

* Related treatments are listed in Table 4.3

SoI = Severity of illness, **FIA** = fair innings, **PS** = proportional shortfall, **HLY** = healthy life years, **DLY** = disabled life years, **QoL** = quality of life, **LLY** = life years lost, **COPD** = Chronic Obstructive Pulmonary Disease, **BPO** = Benign Prostatic Obstruction

table

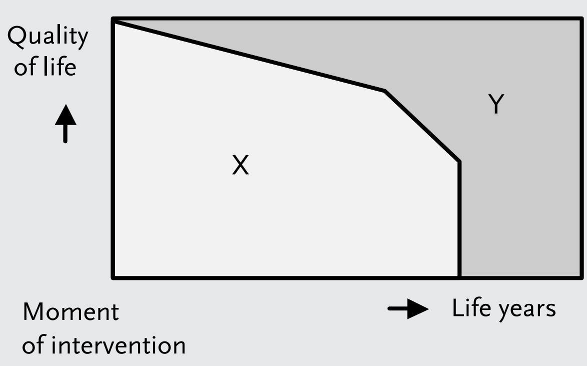
Severity of illness

The severity of illness approach suggests that the patient with the poorest health prospects should get the highest priority. Usually this argument is illustrated with an example on quality of life: studies have shown that a movement from 0.6 to 0.9 receives a lower value than a movement from 0.3 to 0.6 when the size of the health gain is the same (Nord, 1999, p.75). This approach is most concerned about those patients whose inability to function in ordinary life is most pronounced. From this underlying idea one could infer that in terms of QALYs the severity of illness approach would be more concerned with people that have poor QALY prospects than with those who have better prospects if no treatment is provided (assuming equal potential gains). This can be illustrated using ♠ **figure 5.2**, which represents the development of a patient's quality of life during the remaining life years. In the taxonomy of this figure, the size of area 'x' describes severity of illness in terms of the remaining QALY prospects: the smaller area 'x', the greater the severity of illness.

Accordingly, the rank ordering by severity of illness that is presented in ◘ **table 5.1** has been determined on the basis of the no-treatment QALY profile: the higher the absolute number of QALYs that a person will get in spite of his or her condition, the lower his or her priority ranking. In these calculations, all disease free years counted for 1 QALY, QALYs from disabled life years where calculated by multiplying the number of life years by the quality of life weights during those years.

♠ **Figure 5.2** Example of a patient's QALY profile.

This figure describes the QALY-time curve of a hypothetical patient. The vertical line represents quality of life, the horizontal line length of life as expected from at birth. The squared area represents the maximal attainable amount of QALYs. Area 'X' describes the no treatment QALY total of the patient, area 'Y' the QALYs foregone.



Fair innings

Whilst the severity of illness argument considers only inequalities in future health, the fair innings argument is concerned with lifetime health. The fair innings argument requires that everyone be given an equal chance to have a predefined fair innings, and a patient's entitlement for health care drops when s/he gets closer to the fair innings. To put this equity concept into practice, two issues need to be discussed. The first issue concerns past health losses. The fair innings argument bases priorities on the number of QALYs foregone, no matter whether the health losses occurred in the past or will occur in the future. In practice this is not always easy to include in calculations of the fair innings, because the literature usually concentrates on changes in future health that may be brought about by treatment. We therefore simplified the calculations by assuming that all patients have had healthy lives until the moment that they applied for the intervention under consideration. This means that the size of area 'y' in ♠ **figure 5.2** determined the rank ordering by fair innings in this investigation. In this figure 'y' represents health loss: the greater the health loss, the greater priority is placed on treatment of that patient.

Another question in operationalizing the fair innings argument is what the target level of health should be from which the health gap is measured. When Williams illustrated the framework provided by the fair innings, he assumed a 'fair innings' of 70 QALYs (Williams, 2001). Accordingly, the rank ordering by fair innings could be established by dividing the expected lifetime QALYs without treatment by 70. In this approach the fair innings is defined at birth and is equal for everybody. Alternatively, it could be argued that the target level for health innings should be recalculated for survivors according to where they are at present. The reason is that people's health prospects improve when they age (expected age of death is delayed, hence the expected lifetime QALY totals increases with age) and that the years that are 'added' to their life expectancy at birth will also be considered a health loss. In this case, fair innings weights could be determined by dividing the expected lifetime QALYs without treatment by an age specific lifetime QALY expectation

Williams discussed arguments in favour of and against this 'dynamic' version of the fair innings, emphasizing that a dynamic version has more appreciation for health needs of the old. This implies that this method is less powerful to reduce inequalities in total health outcomes over age groups (Williams, 1997). Different demands of equality underlie the two versions of fair innings, reflecting different views as to what aspects of people's health are to be valued in a priority setting context (Williams, 1997) and how people are to be assessed vis-à-vis each other. As there are different ways of seeing this, we decided to put both versions into the experiment: in one version fair innings is fixed at 70 QALYs, in the other version a dynamic fair innings is used, that is calculated using age and gender specific mortality data. Gender was brought in the equation to further individualize the dynamic fair innings, so as not to over- or underestimate the remaining life expectancy of patients in conditions that are gender specific. This means that our dynamic version of fair innings was indifferent to inequalities in life expectancy at birth between males and females. ◘ **Table 5.1** shows several reversals in the tabulated rankings of the ten conditions by the two versions of fair innings.

Proportional shortfall

Whilst both severity of illness and fair innings suggest that equality should be measured in terms of absolute attainments, a third equity concept suggests that comparison may be done in terms of 'proportionate' or 'relative' attainments (Johannesson, 2001). Proportional equality in health corresponds with equalizing persons within their own scope of potential for health, by distributing shares proportionate to people's remaining life expectancy instead of equal shares. Accordingly, different patients are asked to make the same proportional concession to their target level for health, or to accept the same 'proportional shortfall'. Hence equity weights are determined not on the basis of absolute health loss as the fair innings argument assumes, but on the basis of proportional health loss. Since it is the remaining life expectancy that counts and that people want to maximize, this approach calculates remaining life expectancy on the basis of age and gender specific mortality data. This means that proportional shortfall does not compensate for inequalities at birth in the life expectancy between males and females, like the dynamic version of fair innings.

To determine the rank ordering by proportional shortfall first the absolute health loss is measured relative to an age and gender dependent target level for health, similar to the approach in the second version of the fair innings argument. This time however, it is not the absolute loss that counts. Instead, proportional shortfall values the QALY loss relative to the number of QALYs that a patient would receive in the future if he would be in normal health. In other words, proportional shortfall measures the ratio between a patient's QALY expectation if no treatment is to be received ('x' in the taxonomy of ♠ **figure 5.2**), and his or her QALY expectation in absence of the considered condition ('x+y').

5.2.5 Analysis

To explore the social interpretation of equity, we analyzed the congruence between the observed rank order obtained in the paired comparison experiment and the theoretical rank orderings expected by the three equity concepts using Spearman rank order correlations ($p<0.05$). A fisher-z transformation of the correlations was used to test correlation differences ($p<0.05$). First, however, we explored the robustness of the preferences that were elicited in the paired comparison task. For this purpose we applied an internal consistency check. Hereto the difference in the z-value in a pair of conditions

Table 5.2a P-matrix

	Onychomycosis	BPO	Osteo-porosis	High blood pressure	PC pneumonia	High cholesterol	Non-Hodgkin Lymphoid	COPD	Arterio-sclerosis	Pulmonary hypertension
Onychomycosis	0.50	0.94	1.00	1.00	0.97	1.00	0.97	1.00	0.98	1.00
BPO	0.06	0.50	0.94	1.00	0.94	0.98	0.95	1.00	0.98	1.00
Osteoporosis	0.00	0.06	0.50	0.95	0.91	0.94	0.91	0.98	0.98	0.98
High blood pressure	0.00	0.00	0.05	0.50	0.58	0.85	0.66	0.97	0.98	0.97
PC pneumonia	0.03	0.06	0.09	0.42	0.50	0.57	0.89	0.74	0.83	0.98
High cholesterol	0.00	0.02	0.06	0.15	0.43	0.50	0.52	0.89	0.95	0.95
Non-Hodgkin Lymphoid	0.03	0.05	0.09	0.34	0.11	0.48	0.50	0.63	0.78	0.95
COPD	0.00	0.00	0.02	0.03	0.26	0.11	0.37	0.50	0.77	1.00
Arteriosclerosis	0.02	0.02	0.02	0.02	0.17	0.05	0.22	0.23	0.50	0.97
Pulmonary hypertension	0.00	0.00	0.02	0.03	0.02	0.05	0.05	0.00	0.03	0.50

Table 5.2b Z-matrix, eliminating p-values greater than 0.98 and less than 0.02

	Onychomycosis	BPO	Osteo-porosis	High blood pressure	PC pneumonia	High cholesterol	Non-Hodgkin Lymphoid	COPD	Arterio-sclerosis	Pulmonary hypertension
Onychomycosis	0.000	1.542			1.870		1.870			
BPO	-1.542	0.000	1.542		1.542		1.683			
Osteoporosis		-1.542	0.000	1.683	1.319	1.542	1.327			
High blood pressure			-1.683	0.000	0.214	1.020	0.417	1.870		1.870
PC pneumonia	-1.870	-1.542	-1.319	-0.214	0.000	0.174	1.239	0.639	0.957	
High cholesterol			-1.542	-1.020	-0.174	0.000	0.058	1.239	1.683	1.683
Non-Hodgkin Lymphoid	-1.870	-1.683	-1.327	-0.417	-1.239	-0.058	0.000	0.334	0.788	1.683
COPD				-1.870	-0.639	-1.239	-0.334	0.000	0.736	
Arteriosclerosis					-0.957	-1.683	-0.788	-0.736	0.000	1.870
Pulmonary hypertension				-1.870		-1.683	-1.683		-1.870	0.000

BPO = Benign Prostatic Obstruction, **PC** = pneumococcal, **COPD** = Chronic Obstructive Pulmonary Disease was back-transformed to the expected probability that one of the conditions was preferred over the other (Edwards, 1957, p.37-40). When the difference between the expected and the observed data is of the order of 0.05 for most comparisons, the model adequately fits the data (Streiner and Norman, 1995, p. 39-44; Kind, 1996). Additionally, we ran significance tests for the coefficient of consistency (to evaluated consistency in the individual preference orderings) and the coefficient of agreement (to measure agreement among different respondents). Formulas to compute these coefficients can be found in Edwards (1957, Ch. 3).

The rationale underlying the paired comparison technique predicts that the occurrence of inconsistent individual judgments increases as the difference between the compared objects on the underlying continuum decreases. The technique of paired comparison ranking therefore allows individual rank-order inconsistencies. However also other factors might contribute to inconsistencies in individual rank-orderings (poor task comprehension, inability to compare the objects, disinterest in the task), so it is desirable to obtain a measure of the degree of consistency in the responses of each subject. Given that in the paired comparison task each scenario was compared to each other, we were able to test for consistence of a subject's responses. When a respondent is inconsistent, intransitivities occur in the preference ordering. For example, when A is preferred to B, and B is preferred to C, logic predicts that A will be preferred to C. If C on the other hand is considered more favourable these three comparative judgments constitute a circular triad. The number of circular triads is used to find Kendall's coefficient of consistence, which offers a measure of inconsistency in the responses of a particular judge. To test significance of the observed degree of consistency we explored if consistence was greater than can be expected by chance using the χ^2 distribution.

Next we computed the coefficient of agreement, which reflects diversity of preferences among the respondents. When the coefficient approaches 1.0, the subjects have nearly equal orderings. Complete agreement is reached when all respondents make identical choices during the experiment, in which case half of the entries in the preference matrix presented in ◘ **table 5.2a** would be equal to 1.0, while the other half would be zero. Alternatively, if agreement is completely absent among the subjects, all entries will be equal to 0.5. Each time that two test subjects make the same decision in a paired comparison question, we say that we have one agreement regarding this pair. Agreement is measured by counting the number of pairs of test subjects that make the same decision over each pairs of health states that are compared. Again, a χ^2 statistic was used to determine to test agreement among respondents, null hypothesis being that all test subjects cast their preference completely at random.

5.3 Results

5.3.1 Sample

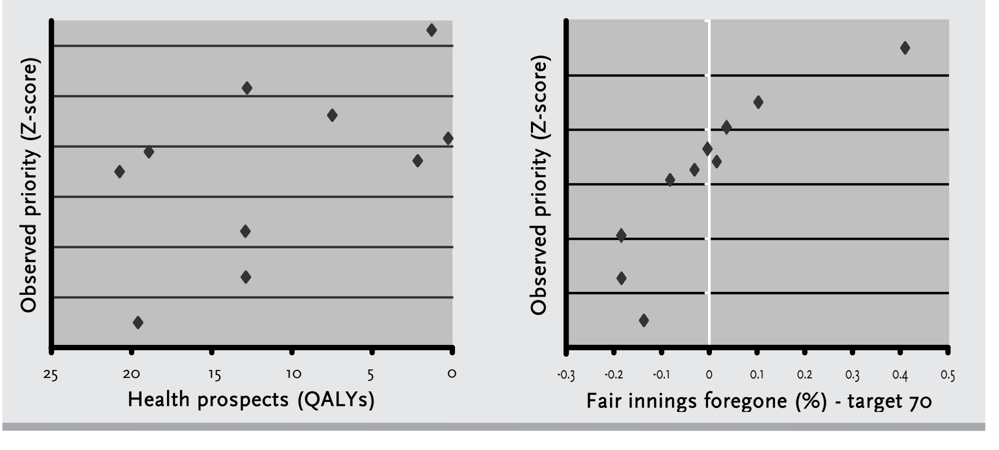
24 students, 24 researchers and 17 health policy makers filled in the questionnaire, which took them on average about 20 minutes. We found no differences in the rank ordering of the three groups. The only exception was that students ranked high cholesterol on place 4 and non-Hodgkin lymphoma on place 5, whilst the other two groups reversed these two positions. Given these marginal differences between the groups we only present the aggregated data.

5.3.2 Paired comparison data

We recorded the frequencies that each alternative was preferred over another alternative. ◘ **Table 5.2a** presents the p-matrix that summarizes the frequency data. The corresponding z-matrix and distances matrix are presented in the ◘ **table 5.2b** and ◘ **table 5.2c**, along with the scale values. These scores indicate the relative distances between the ten conditions on the priority scale.

The internal consistency check demonstrated that the observed proportions agreed well with those to be expected in terms of the derived scale values: the absolute average discrepancy was 0.048. The coefficient of consistence showed a mean value of 0.947 ($p<0.001$), indicating a high level consistency. Participants in the study generally had few circular triads: 30 participants had no triad. The average number of circular triads was 2.1 (SD 3.33) with a maximum of 16 (the maximum number of triads possible was 40). The coefficient of agreement was high: 0.721 ($p< 0.001$). This means that subjects were consistent among each other.

♠ **Figure 5.3** Scatterplots of the observed rank ordering of 10 conditions and four theoretically expected rank orderings



5.3.3 Comparison of observed and theoretical rank orderings

♠ **Figure 5.3** presents the relations between the observed scale values (z-scores) and the scale values expected by the three equity concepts. Spearman correlations show that all theoretical rank orderings were statistically significant correlated to the observed rank ordering. The highest correlations were found between the rank orderings by the two versions of fair innings and the observed rank ordering: 0.985 ($p<0.001$; 95% CI = 0.94 to 1.00) for the dynamic version of fair innings and 0.948 ($p<0.001$; 95% CI = 0.79 to 0.99) for the fair innings with a fixed target of 70 QALYs. The Spearman correlations with proportional shortfall and severity of illness were 0.818 ($p=0.004$; 95% CI = 0.39 to 0.96) and -0.648 ($p=0.043$; 95% CI = -0.91 to -0.03) respectively. Note that the negative sign of the latter correlation is expected, because priority increases when health prospects decrease.

A Fisher-z transformation was used to explore if these correlations differed significantly from each other. This analysis revealed that the correlations of the two versions of fair innings did not differ ($p=0.12$). Also the correlations of proportional shortfall and severity of illness did not differ ($p=0.24$). However, the theoretical rank generated on the basis of fair innings was significantly more correlated with the observed rank ordering than the others. The correlation differences between the two versions of fair innings and both proportional shortfall and severity of illness were all highly significant ($p<0.001$). Visual examination of ♠ **figure 5.3** also demonstrates the most consistent relation between fair innings and the observed value: there were no major exceptions to the rule that the observed value increased with the fair innings foregone. In general, the observed value also increased when the proportional shortfall increased. There were however two

♠ **Figure 5.3 (ctd)** Scatterplots of the observed rank ordering of 10 conditions and four theoretically expected rank orderings

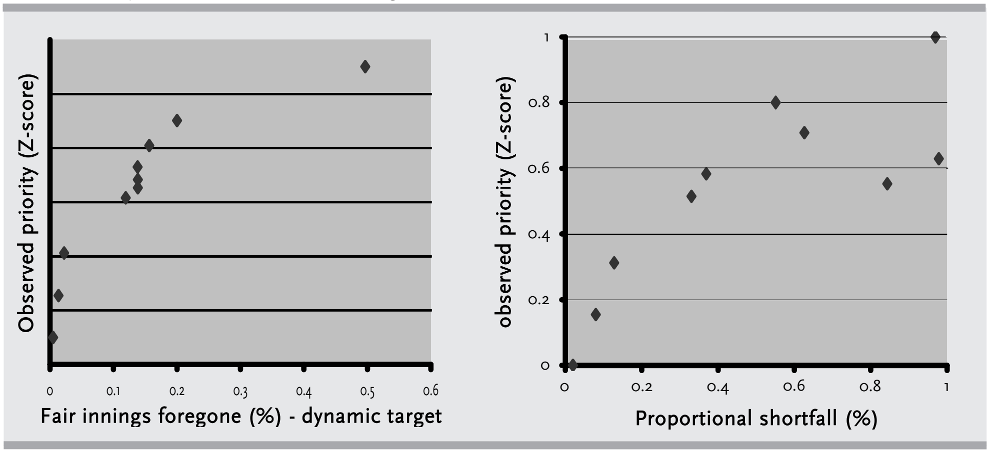


Table 5.2c Matrix of the differences between the z-values, i.e. the scale distances, which were determined by subtracting entries in each column from the corresponding entries in the next column*

	Onychomycosis	BPO	Osteo-porosis	High blood pressure	PC pneumonia	High cholesterol	Non-Hodgkin Lymphoid	COPD	Arterio-sclerosis	Pulmonary hypertension
Onychomycosis		1.542								
BPO		1.542	1.542							
Osteoporosis			1.542	1.683	-0.365	0.223	-0.215			
High blood pressure				1.683	0.214	0.806	-0.603	1.453		
PC pneumonia		0.328	0.223	1.105	0.214	0.174	1.064	-0.600	0.319	
High cholesterol				0.522	0.846	0.174	0.058	1.181	0.444	0.000
Non-Hodgkin Lymphoid		0.186	0.357	0.910	-0.822	1.181	0.058	0.334	0.454	0.895
COPD					1.231	-0.600	0.905	0.334	0.736	
Arteriosclerosis						-0.726	0.895	0.052	0.736	1.870
Pulmonary hypertension							0.000			1.870
Sum		3.598	3.664	5.904	1.317	1.233	2.162	2.753	2.690	4.635
N		4	4	5	6	7	8	6	5	4
Means		0.899	0.916	1.181	0.220	0.176	0.270	0.459	0.538	1.159
Scale values	0.000	0.899	1.815	2.996	3.216	3.392	3.662	4.121	4.659	5.818
Ranking	10	9	8	7	6	5	4	3	2	1

BPO = Benign Prostatic Obstruction, **PC** = pneumococcal, **COPD** = Chronic Obstructive Pulmonary Disease
* To apply this formula the columns must be rank ordered from the least preferred at the left to the most preferred at the right. This information was obtained from the p-matrix.

exceptions: Non-Hodgkin Lymphoma and pneumococcal pneumonia. People were less concerned about the treatment of these two life-threatening conditions for the elderly than expected on the basis of proportional shortfall.

Since the z-score is a relative measure, the absolute values of 'z' have no meaning unless they can be related to an absolute scale. This was only the case for proportional shortfall. In terms of proportional shortfall, the ten conditions were nicely spread across the whole scale (proportional shortfall ranged from 0.02 to 0.98, on a scale ranging 0.00 to 1.00). This means that other indications cannot lose smaller or greater fractions of health than the conditions included in this study already illustrate. To emphasize this, the z-scores were linearly transformed onto a scale from 0 to 1. Since the ten conditions did not cover the whole spectrum of possibilities on the scales of fair innings and severity of illness, the absolute values of 'z' have no meaning for these two equity concepts: higher or lower values may have been found when other conditions would have been included in the experiment. Relevant is then only the fact that the z-scores have interval properties, and that we get an idea of the relative differences between different conditions on an equity rank ordering. For that reason, values of z were not mentioned in ♠ **figure 5.3** for severity of illness and fair innings.

5.4 Discussion

To determine what interpretation of equity should be used in recalculating the value of QALY gains for different patients, we compared the observed rank order of the ten conditions with the rank orders that were expected by the three equity concepts: severity of illness, fair innings, and proportional shortfall. The results showed that the observed rank order of the ten conditions was best predicted by the fair innings concept. Proportional shortfall was also highly correlated with the observed rank order. The severity of illness approach showed a moderate correlation with the observed ranking, suggesting that this concept is less consistent with social preferences for equality in health.

Fair innings had the highest correlation with observed preferences, suggesting that this concept received strongest social support. Moreover, proportional shortfall overestimated the value of treatment of two conditions in the elderly considerably (see ♠ **figure 5.3**). However, it may be to soon to claim superiority of fair innings yet, because effect size might have confounded the results. Respondents assumed that the wonder pill would relieve the recipient of all described health problems. Because the effect size equalled the health gap in the no-treatment QALY profile, a preference for QALY maximization could have boosted the correlation between the observed rank ordering and fair innings. One could wonder why we did not use a fixed effect to prevent this collinearity. However, the use of a fixed effect could not solve this problem of collinearity; it merely moves it. Because the proportional shortfall approach is consistent with the maximization of relative QALYs (Johannesson, 2001), the use of a fixed effect produces collinearity between the observed rank ordering and the rank ordering for proportional shortfall. For example, if the treatment would have assumed a fixed effect (e.g. 5 QALYs) the correlation with proportional shortfall would be affected by the fact that 5 QALYs are a greater share of shorter remaining life expectancies. In future studies the best strategy may therefore be to test preferences for varying effect sizes in different situations depending on the absolute and relative health gap.

The theoretical rank generated on the basis of fair innings was significantly more correlated with the observed rank ordering than the others. The correlation differences between the observed rank ordering and the rank orderings generated by proportional shortfall and severity of illness did not differ significantly. An explanation is that the 95% confidence interval of the Spearman correlations for severity of illness and proportional shortfall were relatively wide. More robust results may be produced when a higher number of cases is included. Moreover, our choice to include realistic cases may have limited robustness of the results to some degree, because cases were not selected randomly or most efficiently in terms of the statistic properties of the design. Case selection was conditional on the availability of the QALY profiles, so that discriminative ability of the design was not maximized. The design was slightly unbalanced in the sense that the cases were not equally well divided over the three theoretical scales. In terms of proportional shortfall, the ten conditions were nicely spread across the whole scale, but for fair innings and severity of illness only half the scales were used (see ♠ **figure 5.3**). Additionally, our case selection did not try to minimize the a priori correlations between the three theoretical rank orderings. The rank orderings by severity of illness and fair innings were not statistically significant correlated ($p<0.05$), but the rank orderings by proportional shortfall was correlated to fair innings (0.717, p=0.020 for fixed target fair innings, 0.800, p=0.005 for dynamic fair innings) and severity of illness (0.855, p=0.002). Since proportional shortfall combines elements of fair innings and severity of illness, a correlation with the other equity concepts is unavoidable. However, a greater level of orthogonality may be obtained when (hypothetical) cases are carefully selected.

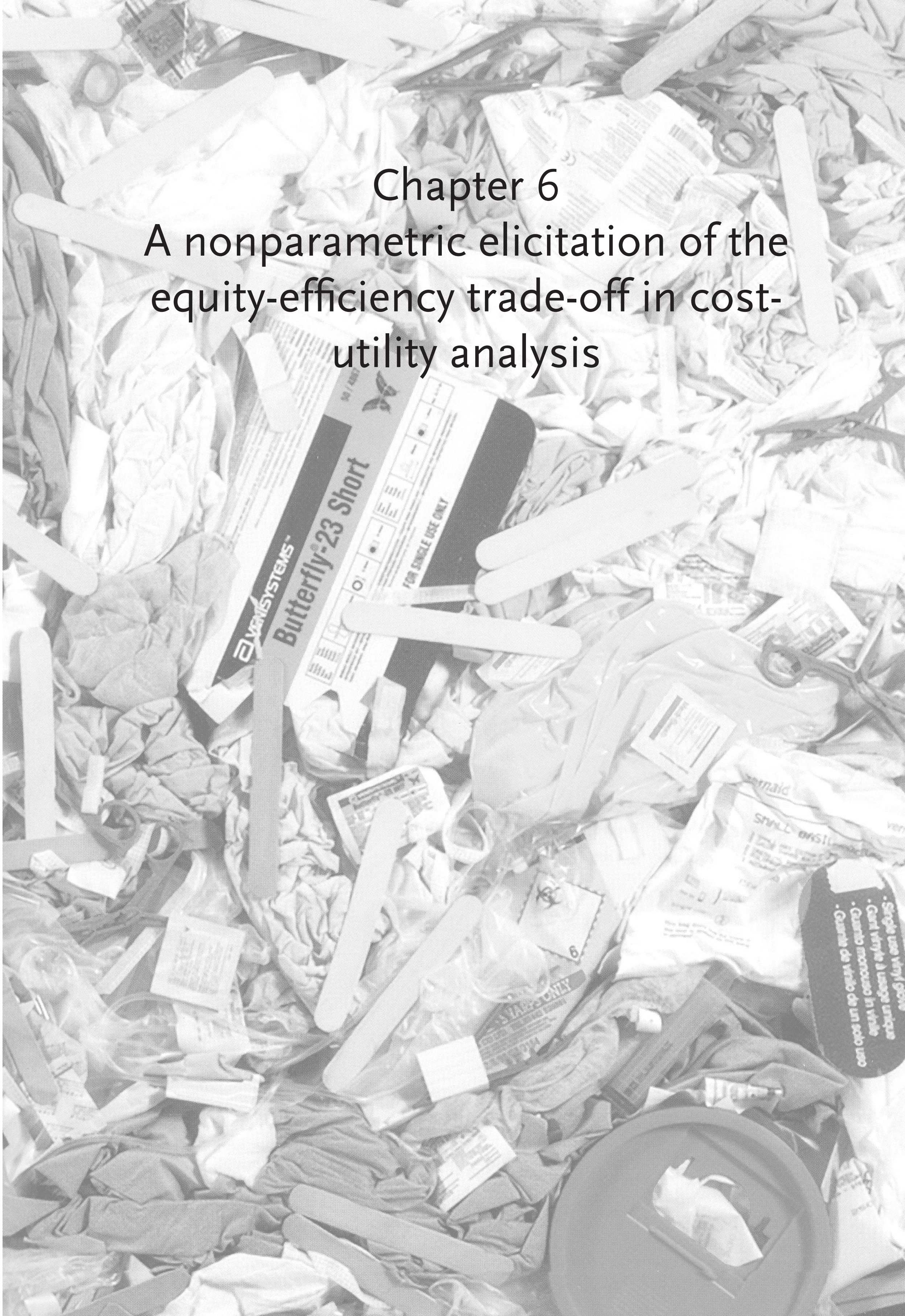
Finally some conceptual issues need to be discussed. First, some researchers may disagree that distributive justice belongs to the scope of economics. This may be traced back to different opinions about the social objective of health care, the type of utilities that constitute social welfare, or the legitimacy of addressing equity issues in economic models (Sassi et al., 2001). Based on the idea that both measures of efficiency and fairness depend on value judgments (Dolan and Olsen, 2002, p. 47), however, we believe that it is appropriate to pay attention to equity issues to improve the descriptive validity of economic models. A second conceptual issue that needs to be discussed is whether all three positions are equally appropriate to inform resource allocation, notwithstanding social support. From a theoretical perspective a strong case can be made for equality in lifetime health, because health is an important condition of human life and a critical constituent of opportunities (Sen, 2002). People who demand equality in lifetime QALYs may therefore find both the severity of illness and the proportional shortfall approach inadequate (Williams, 2001). Nevertheless the public also seems to support principles that focus on health achievements rather than on the opportunities of people to accomplish what they value. As objectivity in ethics is warranted, such 'gut feelings' may need to be criticized and rationalized. Still, 'gut feelings' may also function as eye-opener and bridge towards new ways of moral consideration (Van Willigenburg, 2003). We therefore did not a priori exclude particular equity concepts from this investigation.

The information on equity preferences obtained in this study might have implications for policy. In The Netherlands, interventions with vital consequences for the elderly have always been funded, even in spite of unfavourable cost-effectiveness (Boer, 2001). These policy choices can best be explained by the equity concept of severity of illness, an equity concept that did not reflect the distributive preferences of our sample well. An explanation for the discrepancy between our outcomes and recent policy decisions may be that for the sake of this experiment, a high level of scarcity was assumed. When health policymakers are confronted with terminally ill patients the question is if they perceive the same level of scarcity, or that they will be inclined to increase budget for the health care sector (Van de Vathorst, 2001).

To conclude, this investigation aimed to contribute to the debate about the social interpretation of equity. An important message is that measurable interpretations of equity make it possible to validate people's claims on health care. We conclude that the fair innings argument and proportional shortfall may provide a better basis for determining equity weights for recalculating the value of QALY gains for different patients than the severity of illness approach. Given the apparent conflict between recent health care decisions and the results of our investigation, the question can be raised whether current allocation of the health care budget is in line with societal preferences.

Part of Damien Hirst's Urinal II

Chapter 6
A nonparametric elicitation of the equity-efficiency trade-off in cost-utility analysis



We performed an empirical elicitation of the equity-efficiency trade-off in cost-utility analysis using the rank-dependent quality-adjusted life-year (QALY) model, a model that includes as special cases many of the social welfare functions that have been proposed in the literature. Our elicitation method corrects for utility curvature and, therefore, our estimated equity weights are not affected by diminishing marginal utility. We observed a preference for equality in the allocation of health. The data suggest that the elicited equity weights were jointly determined by preferences for equality and by insensitivity to group size. A procedure is proposed to correct the equity weights for insensitivity to group size. Finally, we give an illustration how our method can be implemented in health policy.

Chapter 6

A nonparametric elicitation of the equity-efficiency trade-off in cost-utility analysis

Bleichrodt H, Doctor J, Stolk EA[1]. A Nonparametric Elicitation of the Equity-Efficiency Tradeoff in Cost-Utility Analysis. Journal of Health Economics (in press)

1 The order of the authorship is alphabetical to reflect equal contributions

6.1 Introduction

The common procedure to aggregate health benefits in economic evaluations of health care is by unweighted aggregation, also referred to as quality-adjusted life-year (QALY)-utilitarianism. This procedure weights the health gains of each individual equally and leads to a maximization of health gains. Several authors have raised concerns about the equity implications of QALY-utilitarianism and have argued that it may be necessary to differentiate between individuals based on, for example, age, health status, or previously enjoyed health (Harris, 1987; Nord, 1995; Williams, 1997; Williams and Cookson, 2000).

Empirical evidence supports these concerns and indicates that people, when choosing between different allocations of health gains, not only consider efficiency, the total amount of health gains, but also equity, the distribution of the health gains (e.g. Nord, 1993; Dolan, 1998; Abellan and Pinto, 1999). These findings suggest that it may be preferable to replace QALY-utilitarianism by some sort of equity-weighted aggregation rule. Unfortunately, the available empirical research offers little guidance as to which rule should be used and how the equity weights could be elicited.

Several authors have proposed theoretical models to incorporate equity considerations into cost-utility analysis (Wagstaff, 1991; Bleichrodt, 1997; Williams, 1997; Dolan, 1998). Both Wagstaff (1991, 1993) and Dolan (1998) proposed to use an iso-elastic social welfare function to allow for a trade-off between efficiency and equity. Within this class of social welfare functions, Dolan (1998) suggested, in particular, to use a Cobb-Douglas function. Wagstaff (1991) and Dolan (1998) did not derive the assumptions underlying their proposed social welfare functions, which complicates an assessment as to why the equity-efficiency trade-off should take the form they propose. They did not explain either how the parameters in their social welfare functions could be assessed.

Bleichrodt (1997) proposed a multiplicative social welfare function, derived the conditions on which it depends, and showed how its equity parameter could be elicited. The range of equity concerns that the multiplicative social welfare function can address is, however, limited. Williams (1997) suggested that individuals should be weighted according to their 'fair innings', the difference between the amount of health they already enjoyed and the amount of health they are entitled to over their lifetime. Williams' proposal suggests that he had in mind some sort of weighted aggregation rule, but he did not specify what form this weighted rule should take nor did he explain how the equity weights could be elicited.

Bleichrodt et al. (2004) recently proposed a new social welfare function to incorporate equity considerations into cost-utility analysis, the rank-dependent QALY model. Their model has several desirable characteristics. First, it is consistent with several social welfare functions that have been proposed in the literature, including QALY-utilitarianism, the Rawlsian social welfare function in which all weight goes to the worst-off individual, and the Gini social welfare function, which is widely used in inequality measurement. The rank-dependent QALY model can also accommodate Williams' fair innings approach. Second, as Bleichrodt et al. (2004) showed, the rank-dependent QALY model depends on assumptions that have normative appeal. A third advantage of the model is that the elicitation of the equity weights is straightforward. Finally, the model is tractable: once the equity weights have been elicited, the model can easily be used in cost-utility analyses.

The aim of this paper is to elicit the equity weights under the rank-dependent QALY model. For reasons explained in Section 2, we used a slightly more general model than the model proposed in Bleichrodt et al. (2004). In Bleichrodt et al. (2004) the social utility function over QALYs is linear, whereas in this paper, we allow for a nonlinear social utility function over QALYs. We refer to this extended model as the nonlinear rank-dependent QALY model. A consequence of using a more general model is that its elicitation becomes more involved, because, in addition to the equity weights, the social utility function over QALYs must be determined.

The structure of the rest of the paper is as follows. In Section 6.2 we describe the nonlinear rank-dependent QALY model. In Section 6.3, we explain the elicitation of the model. To elicit the model, we used an adjusted version of the trade-off method (Wakker and Deneffe, 1996), which was developed to measure utilities under risk. An advantage of the trade-off method is that it is nonparametric: it imposes no assumptions on the utility function or on the equity weighting function. We elicited the nonlinear rank-dependent QALY model both in a sample of students and in a sample of the general population. Section 6.4 describes the designs of the two experiments, Section 6.5 the results. Section 6.6 shows how our method can be implemented in health policy. Section 6.7 offers conclues.

6.2 The rank-dependent QALY model

We consider a health policy maker who has to choose between different QALY allocations. Consider a population of *n* individuals. Let $(q_1, \ldots, q_n)$ denote the *QALY-profile*, which gives q_i QALYs to individual *i*. We will interpret QALYs as measures of health in this paper. Unless otherwise stated, we assume that QALY-profiles are *rank-ordered* so that $q_1 \geq \ldots \geq q_n$. This is, obviously, no restriction because each QALY-profile can be written in a rank-ordered form.

In this paper, we study preferences over QALY-profiles. To describe these preferences, Bleichrodt et al. (2004) suggested using the rank-dependent QALY model. According to the rank-dependent QALY model, the social value of QALY profile $(q_1, \ldots, q_n)$ is equal to:

$$\sum_{i=1}^{n} \pi_i q_i \qquad ♣ \ (6.1)$$

where the π_i are equity weights that are defined as $\pi_i = w(i/n) - w((i-1)/n)$. The function w is a nondecreasing function that has $w(0) = 0$ and $w(1) = 1$.

Under the rank-dependent QALY model, the social value of a QALY allocation is thus expressed in terms of two scales, w and q. The scale q is the familiar one for quality-adjusted life expectancy. The other scale, w, associates with each individual's expected quality-adjusted lifetime, q_i, an equity weight π_i, which reflects the weight the policy maker gives to individual i in the evaluation of QALY-profiles.

Under the rank-dependent QALY model, the equity weight assigned to an individual depends on how well-off he is in terms of QALYs by comparison with the other individuals in society, i.e. the equity weight depends on the individual's rank. A shift in the individual's rank will generally lead to a shift in his equity weight. A detailed explanation of the intuition behind the rank-dependent QALY model is given in Bleichrodt et al. (2004).

It is easily verified that in case the function w is linear, the rank-dependent QALY model is identical to QALY-utilitarianism. If w is convex then the policy maker is averse to inequalities, in the sense that he will always prefer a transfer of QALYs from an individual who has relatively many QALYs to an individual who has less, as long as the rank-ordering of the individuals in terms of the number of QALYs received is not affected. If w is concave then the policy maker is inequality seeking. Because the function w describes attitudes towards inequality, we refer to this function as the equity weighting function.

In this paper, we consider a generalized version of the rank-dependent QALY model, the nonlinear rank-dependent QALY model, in which the value of the rank-ordered QALY profile $(q_1, \ldots, q_n)$ is equal to:

$$\sum_{i=1}^{n} \pi_i U(q_i) \qquad ♣ \ (6.2)$$

The difference with ♣ **equation 6.1** is that in the nonlinear rank-dependent QALY model the utility function U over QALYs need not be linear. An important point to note is that the utility function U in ♣ **equation 6.2** is the policy maker's utility function over QALYs; it reflects the value the policy maker places on different numbers of QALYs experienced by the people in society. Assuming the existence of a social utility function is common in the literature on inequality measurement (e.g. Atkinson, 1970; Ebert, 1988). For health, the approach of defining a social utility function over QALYs has been used by Wagstaff (1991), Bleichrodt (1997) and Dolan (1998). Note that ♣ **equation 6.2** could be made consistent with individuals valuing their own QALYs in a nonlinear manner by substituting $u_i(q_i)$ for q_i, where u_i is an individual utility function over QALYs. We do not pursue such an extension in this paper.

The reason to allow for nonlinear utility over QALYs is that the elicitation of social preferences is a descriptive task and it is not a priori clear that a linear utility function over QALYs describes preferences over QALY-profiles well. If it does not, a preference for a more equal distribution of QALYs may be the product of two conceptually different factors: a preference for equality per se and diminishing marginal utility for QALYs. Diminishing marginal utility reflects that the policy maker's valuation of additional QALYs decreases with the amount of QALYs. For example, a policy maker may consider receiving 80 QALYs and receiving 90 QALYs as close because in both cases an individual has a long and healthy life, whereas he considers the difference between receiving 50 and 60 QALYs as larger. On the other hand, the policy maker might also prefer a more equal distribution regardless of his valuation of QALYs. Such a preference for equality is reflected in the equity weights π. The nonlinear rank-dependent QALY model allows to separate these two types of concern for equality and can, therefore, shed more light on what drives people's preferences over QALY-profiles.

As mentioned above, by taking a utility function over QALYs, our approach is consistent with Wagstaff (1991), Bleichrodt (1997) and Dolan (1998). In fact, Dolan's Cobb–Douglas model is a special case of ♣ **equation 6.2** in which the utility function over QALYs is logarithmic. Hence, our elicitation of the utility function over QALYs allows for a test of the Cobb–Douglas social welfare function proposed by Dolan.

Because the nonlinear rank-dependent QALY model, ♣ **equation 6.2**, is more general than the rank-dependent QALY model, ♣ **equation 6.1**, it shares the advantage of encompassing many of the social welfare functions that have been proposed in the literature. The nonlinear rank-dependent QALY model is also easy to use in practice once the utility function and the equity weighting function have been elicited. The elicitation of the nonlinear rank-dependent QALY model is, however, more involved because both the utility function and the equity weighting function must be assessed. We now turn to the issue of elicitation.

6.3 Elicitation

We elicited the nonlinear rank-dependent QALY model in two stages. In the first stage, the social utility function over QALYs was elicited. That is, we put subjects in the position of health policy makers and determined how they valued the amounts of QALYs received by others. This approach of putting subjects in the position of health policy makers is common in the literature on the equity-efficiency trade-off (e.g. Nord, 1993; Dolan, 1998; Rodrigues-Miguez and Pinto-Prades, 2002). The elicited social utilities were then used as inputs in the second stage, in which the equity weighting function was elicited.

6.3.1 Elicitation of the utility function

We first selected two gauge outcomes R and r and a starting value x_o. We took $x_o > R > r$. Let (x, p, y) denote the rank-ordered QALY-profile that gives x QALYs to proportion p of the population and y QALYs to proportion 1-p of the population, $x \geq y$. We determined the number of QALYs x_1 that made a subject indifferent between (x_1, p, r) and (x_0, p, R). Because more QALYs are preferred to less, we must have $x_1 > x_0$. In terms of ♣ **equation 6.2**, the indifference between (x_1, p, r) and (x_0, p, R) means that

$$w(p)U(x_1) + (1-w(p))U(r) = w(p)U(x_0) + (1-w(p))U(R) \qquad \clubsuit\ (6.3a)$$

or

$$U(x_1) - U(x_0) = \frac{1-w(p)}{w(p)}(U(R) - U(r)) \qquad \clubsuit\ (6.3b)$$

After x_1 had been elicited, the number of QALYs x_2 was determined such that the subject was indifferent between (x_2, p, r) and (x_1, p, R). This indifference implies by ♣ **equation 6.2** that

$$U(x_2) - U(x_1) = \frac{1-w(p)}{w(p)}(U(R) - U(r)) \qquad \clubsuit\ (6.4)$$

Combining ♣ **equation 6.3b** and ♣ **equation 6.4**, we find that

$$U(x_2) - U(x_1) = U(x_1) - U(x_0) \qquad \clubsuit\ (6.5)$$

We can continue in this fashion and elicit indifferences between (x_j, p, r) and (x_{j-1}, p, R), in the process eliciting a standard sequence $x_1, \ldots, x_k$ such that the utility intervals between successive elements are all equal. That is, $U(x_i) - U(x_i-1) = U(x_j) - U(x_{j-1})$ for all i and j between 1 and k.
The origin and the unit of the utility function can be chosen freely. We selected $U(x_0) = 0$ and $U(x_k) = 1$. It then follows that $U(x_j) = j/k$ for all j between 1 and k.

6.3.2 Elicitation of the Equity Weighting Function

In the first stage of the elicitation procedure, the proportion p was kept constant to be able to elicit the utility function. To elicit the equity weighting function, the proportion p will be varied across questions. We used the following types of questions to elicit the equity weighting function. For low proportions p, we elicited the amount of QALYs z that made subjects indifferent between (x_k, p, x_0) and (x_i, p, z), $0 < i < k$, $x_i \geq z$, where the x's are elements of the standard seqence that was elicited in the first stage. Using the scaling $U(x_0) = 0$ and $U(x_k) = 1$, this indifference implies under the nonlinear rank-dependent QALY model:

$$w(p)U(x_k) + (1-w(p))U(x_0) = w(p)U(x_i) + (1-w(p))U(z)$$

$$\leftrightarrow w(p) = w(p)*(i/k) + (1-w(p))U(z)$$

$$\leftrightarrow w(p) = \frac{U(z)}{1+U(z)-(i/k)} \qquad \clubsuit\ (6.6a)$$

For high proportions p, we elicited indifference between (x_k, p, x_0) and (z, p, x_j), $0 < j < k$, $z \geq x_j$, where the x's, again, are elements of the standard sequence elicited in the first stage. By the nonlinear rank-dependent QALY model we obtain:

$$w(p)U(x_k) + (1-w(p))U(x_0) = w(p)U(z) + (1-w(p))U(x_j)$$

$$\leftrightarrow w(p) = w(p)U(z) + (1-w(p))*(j/k)$$

$$\leftrightarrow w(p) = \frac{j/k}{1+(j/k)-U(z)} \qquad \clubsuit\ (6.6b)$$

We could also have determined the equity weighting function by eliciting indifference between $(z, 1)$ and (x_k, p, x_0). This immediately gives $w(p) = U(z)$. Ubel et al. (2001) showed, however, that people tend to overstate their preference for equality when one of the options involves no inequality. We, therefore, avoided this type of questions.

Our procedure for determining the equity weights has three potential drawbacks. First, the outcomes z will generally not belong to the standard sequence elicited in the first stage and, therefore, their utility has to be approximated. This approximation may introduce bias. In the analysis of the results we used a linear approximation. Over small intervals the utility function does not deviate much from linearity and a linear approximation will be reasonable as long as successive elements of the standard sequence are close. To test the robustness of our results, we also approximated the utilities of z assuming three nonlinear parametric utilities, as will be described in Section 6.4.

Second, our procedure imposes bounds on the elicited equity weights. In ♣ **equation 6.6a** the equity weight can vary only between 0, which occurs when $z = x_0$ and i/k, which occurs when $z = x_i$. If the outcome z exceeds x_i then the QALY profile (x_i, p, z) is no longer rank-ordered. Its rank-ordered analogue is $(z, 1\text{-}p, x_i)$ and the indifference between (x_k, p, x_0) and $(z, 1\text{-}p, x_i)$ gives by ♣ **equation 6.2**

$$w(p)U(x_k) + (1-w(p))U(x_0) = w(1-p)U(z) + (1-w(1-p))U(x_i) \qquad \clubsuit\ (6.7)$$

That is, an equation with two unknowns, $w(p)$ and $w(1\text{-}p)$, which cannot be solved in a unique manner. Similarly, in ♣ **equation 6.6b** the equity weight can only vary between j/k, which occurs when $z = x_j$, and 1, which occurs when $z = x_k$. In Section 6.4, we explain how we handled the potential boundedness problem.

Finally, our method may suffer from error propagation. ♣ **Equation 6.6a** and ♣ **equation 6.6b** determine equity weights by a ratio. Error propagation for ratios can be problematic if the denominator is close to zero, so that small errors in the numerator lead to large errors in the ratio. Such problems do not occur in ♣ **equation 6.6a** and ♣ **equation 6.6b** because the denominator is far from zero, in fact, more so than the numerator. Moreover, in both expressions, the numerator and the denominator are positively correlated because of a common term, which further reduces the overall error in the ratio. These observations suggest that error propagation will not be problematic in our design.

6.4 Experiments

6.4.1 Subjects

subjects

We performed two experiments to elicit the nonlinear rank-dependent QALY model. The subjects in the first experiment were 69 students at Erasmus University, Rotterdam. The subjects in the second experiment, which was run one month after the first, were 208 members from the general population. These subjects were recruited through a marketing agency from a representative sample of the Dutch population between 16 and 70 years old. ◘ **Table 6.1** describes the characteristics of the sample from the general population split according to sex, level of education and age. Women were over-represented in our sample and people with a low level of education were slightly underrepresented. Subjects in the general population sample were paid €17.50 for their participation, subjects in the students sample were paid €12.50. Prior to the actual experiments we performed nine pilot sessions, using students, to test and fine-tune the questionnaire.

6.4.2 Procedures

The experiments consisted of a computer-based questionnaire. In the student sample, the experiment was carried out in personal interview sessions. In the general population sample, the experiment was carried out in group sessions with a maximum of 15 subjects

◘ **Table 6.1** Characteristics of the sample from the general population

	Population	Proportion (%)
Sex	Male	39.7
	Female	60.3
Education	Low	18.2
	Middle	45.0
	High	36.8
Age	11-20	7.7
	21-30	15.3
	31-40	20.6
	41-50	20.1
	51-60	24.4
	61-70	12.0

per session. There were 22 group sessions in total and, hence, the average number of subjects per session was slightly less than 10. In the group sessions, the experiment was introduced classically. The questionnaire was then administered individually. There were three interviewers present during the group sessions to help subjects with any problems. Before the experiment started we explained to the subjects why it is important for health policy to have information on people's preferences concerning the allocation of health and that their responses would help to make better-informed resource allocation decisions. We then explained to them in intuitive terms the concept of a QALY. The QALY-explanation that was read to the subjects can be found in Appendix 6A.

The decision problem in the actual experiment was the following. Subjects were asked to consider a cohort of newborns who suffer from some disease. The disease was left unspecified to avoid a possible framing effect. We deliberately selected a cohort of newborns to avoid that people thought they might themselves belong to the cohort and consider the decision problem as a decision under risk. In that case, preferences for equity would be confounded by risk attitude.

Subjects were told that there exist two treatments for the disease. The treatments have identical costs but differ in their effects. The treatments were labeled A and B to avoid possible framing effects. The outcomes of the treatments were integer numbers of QALYs. The treatments gave one part of the cohort, the "better-off group" as we will call them henceforth, more QALYs than the other part, the "worse-off group". Subjects were asked to make a choice between the two treatments. An example of the questions that subjects faced is given in Appendix B.

Following the explanation of the decision problem, subjects were given a practice question. In the student sample, we asked subjects to explain their answer to this question. In the general population sample, the interviewers asked some of the subjects to explain their answer. We used the explanation to check whether subjects understood the experimental task. In case we were convinced that subjects understood the task, we asked them to move on to the actual experiment.

Elicitation was by means of a sequence of choices. We opted for a choice-based elicitation procedure, because empirical evidence suggests that choice-based procedures are more consistent and less susceptible to biases than other elicitation procedures, such as matching (Bostic et al., 1990). We used the parameter estimation by sequential testing (PEST) procedure to elicit responses (Luce, 2000, pp. 291–292). PEST is an adaptive elicitation technique that determines the stimulus value for each new question by the subject's response to the previous one. PEST has the advantage of being able to home-in on an indifference value without the subject being aware that this is happening, thus, preventing the subject from forming a conscious numeric indifference. Such mental "matches" have been shown to lead to biases (see, Luce, 2000, for a review). Another advantage of the PEST procedure is that it tests for inconsistencies in subjects' responses, by repeating questions, and only converges to an indifference value when the responses become consistent. The PEST algorithm determined indifference to the nearest QALY integer value. The PEST algorithm is described in Appendix C, which also includes an illustration of the method.

In the first stage of the experiment, the utility function over QALYs was elicited. We elicited a standard sequence of six elements. So $x_k = x_6$ in our study. The starting value x_0 was set equal to 10 QALYs and the two gauge outcomes *R* and *r* were set equal to 8 and 5 QALYs, respectively. We avoided the outcome 0 QALYs, because this might invoke strong emotions which could distort the elicitation. The proportion p was set equal to 1/2. So in the first stage of the experiment half of the cohort was in the better-off group and half was in the worse-off group and the outcome x_j was elicited so that indifference held between (x_j, ½, 5) and (x_{j-1}, ½, 8). We learned from the pilot sessions that these stimuli led to a standard sequence x_1, ..., x_k whose successive elements were relatively close.

In the elicitation of the utility function, we varied only the outcome x_j to reach indifference. To try and avoid that subjects would focus too much on this outcome, and ignore the other stimuli, we included two filler questions in which all stimuli varied after each choice question.

To elicit the equity weighting function, the proportion of the cohort that belonged to the better-off group was varied. The elicitation of the equity weights was preceded by a practice question. By asking subjects to explain their answer to this question we were able to check whether they realized that the proportion had changed. We used five proportions in the elicitation of the equity weighting function: $p_1 = 1/6$, $p_2 = 1/3$, $p_3 = 1/2$, $p_4 = 2/3$ and $p_5 = 5/6$. The proportions were chosen so as to achieve a good spread over the [0, 1] interval and so that subjects could easily compute which treatment gave more QALYs. In the pilot sessions, we experimented with different proportions. It turned out that using smaller proportions than 1/6 or higher proportions than 5/6 led to unstable estimates. We, therefore, avoided using such low and high proportions in the actual experiment.

The stimuli varied with the proportion used. The first column of ◘ **table 6.2** shows the question that we employed for each proportion. In the questions p_1, p_2, and p_3, we

◘ **Table 6.2** Questions used to determine the Equity Weights

Proportion	Question	Interval
$P_1 = 1/6$	(x6, 1/6, 10) vs. (x2, 1/6, z1)	[0, 1/2]
$P_2 = 1/3$	(x6, 1/3, 10) vs. (x3, 1/3, z2)	[0, 2/3]
$P_3 = 1/2$	(x6, 1/2, 10) vs. (x4, 1/2, z3)	[0, 5/6]
$P_4 = 2/3$	(x6, 2/3, 10) vs. (z4, 2/3, x2)	[1/6, 1]
$P_5 = 5/6$	(x6, 5/6, 10) vs. (z5, 5/6, x2)	[1/6, 1]

determined the outcome z that yielded indifference in the comparison between $(x_6, p, 10)$ and (x_i, p, z_m), $(i, m) = (2, 1), (3, 2), (4, 3)$, where x_i and x_6 were taken from the standard sequence that was elicited in the first stage, and we used ♣ **equation 6.6a** to compute the equity weights. If a subject was about to make a choice that implies that z_m exceeds x_i, in which case (x_i, p, z_m) is no longer rank-ordered and the analysis of Section 6.3 cannot be applied, the computer increased x_i to x_{i+1}. For example, in the question for p_1, x_2 was raised to x_3 when z_1 was about to exceed x_2. In the questions for p_4 and p_5, we elicited the outcome z_m that made the subject indifferent between $(x_6, p, 10)$ and (z_m, p, x_2), $m = 4, 5$, and we used ♣ **equation 6.6b** to compute the equity weights. In case a subject was about to violate rank-ordering of (z_m, p, x_2), which occurs if z_m is less than x_2, the computer decreased x_2 to x_1.

In Section 6.3, we explained that our elicitation method imposes bounds on the values that the equity weights can assume. The third column of ◘ **table 6.2** shows for each of the five proportions the interval within which the weight given to the better-off group is forced to lie. For example, the first entry of the column shows that the weight given to the better-off group when the size of the better-off group is 1/6 of the size of the cohort could never exceed 1/2. It would, of course, be better to have a higher upper bound than 1/2, which could be achieved by replacing x_2 by a "higher" element of the standard sequence, i.e. x_3, x_4 or x_5. We learned from the pilot sessions, however, that this made the estimates less stable and more sensitive to response error.

An example, using the data from one of our subjects, may explain the problem of sensitivity to response error. Let the standard sequence $\{x_1, \ldots, x_6\}$ be {15, 21, 29, 39, 54, 68}. Suppose that x_4 were used instead of x_2 to determine $w(1/6)$. We would then elicit the outcome z_1 that made the subject indifferent between (68, 1/6, 10) and $(39, 1/6, z_1)$. Suppose that the subject's true equity weighting function is strictly convex, i.e. $w(p) < p$ for all p in (0,1). To have $w(1/6) < 1/6$, z_1 must be smaller than 12. Suppose, as is likely, that the subject's choices are subject to response error. It is unlikely that a value of z_1 will be elicited that is lower than 11, because the subject will realize that by dominance (68, 1/6, 10) is better than (39, 1/6, 10). On the other hand, we may well elicit a value higher than 13. Hence, there is more room for errors "on the right" of 12 than "on the left". This asymmetric error pattern may bias $w(1/6)$ upwards. By using x_2, $w(1/6) < 1/6$ corresponds to a value of z of 14 or less. So there is more room for error on the left and the problem of asymmetric error is less urgent. Only for those subjects who threatened to violate rank-dependency did the computer change x_2 into x_3. But the choices of these subjects implied $w(1/6) > 1/3$ and it is unlikely that these subjects' true value of w(1/6) is less than 1/6, so the problem of asymmetric error did not occur for these subjects. To reduce the possibility of asymmetric errors affecting the results, we did not use proportions lower than 1/6 (or higher than 5/6) either. The final selection of the stimuli reflected what we believed to be the most finely tuned balance between stability of the estimates and restrictiveness of the bounds. In the pilot sessions, the bounds caused no problems: the implied equity weights were always at a safe distance from the bounds.

Because the PEST procedure requires a series of choices to find the indifference value, we were able to mix questions for different proportions. Both the outcomes and the proportions, therefore, changed across questions. We hoped that this would encourage subjects to focus on all the stimuli. The order in which questions appeared was random.

6.4.3 Analysis

We classified a subject's utility function as concave, linear, or convex depending on how the slope of his elicited utility function changed across points of the standard sequence. Let Δ^j_{j-1} denote the difference between $(x_j - x_{j-1})$ and $(x_{j-1} - x_{j-2})$, $j = 2, \ldots, 6$. It is easily verified that a concave utility function corresponds to Δ^j_{j-1} positive, a linear utility function corresponds to Δ^j_{j-1} zero, and a convex utility function corresponds to Δ^j_{j-1} negative. We observed five values of Δ^j_{j-1} for each subject. To account for response error, we classified a subject's utility function as concave (linear/convex) if at least three values of Δ^j_{j-1} were positive (zero/negative).

To compute the equity weights, we needed the utilities of z_m, $m = 1, \ldots, 5$. These were determined through linear interpolation. To test the robustness of the results, the utility of z_m, $m = 1, \ldots, 5$, was also computed allowing for curvature of utility. We examined three parametric specifications for the utility function: the power function, the exponential function and the expo-power function.

Let $y = (x - x_0)/(x_6 - x_0)$, where x is in $[x_0, x_6]$. The power function is defined by y^r, if $r > 0$, by $\ln(y)$ if $r = 0$, and by $-y^r$ if $r < 0$. The exponential family is defined by $(e^{ry}-1)\ /\ (e^r - 1)$ if $r \neq 0$ and by y if $r = 0$. The power and exponential family are widely used in economics and (medical) decision analysis. Dolan's (1998) Cobb-Douglas social welfare function is the special case of ♣ **equation 6.2** where the utility function is logarithmic.

The expo-power family was introduced by Abdellaoui et al. (2002) and is a variation of a two-parameter family proposed by Saha (1993). The expo-power family is defined by $(1-\exp(-y^r/r))\ /\ (1-\exp(-1/r))$ with $r > 0$. We included the expo-power family because it can accommodate some important preference patterns that are incompatible with both the power and the exponential family (see Abdellaoui et al., 2002, for a discussion). The three utility functions were estimated by a distribution-free iterative procedure that minimized the sum of squared residuals, using the elements of the standard sequence and their corresponding utilities as data inputs.

To analyze equity weighting at the individual level, we examined how the slope of a subject's equity weighting function evolved. Let Δ^j_{j-1} be equal to the difference between $(w(p_j) - w(p_{j-1}))$ and $(w(p_{j-1}) - w(p_{j-2}))$. For $j = 2, \ldots, 6$, with $p_0 = 0$ and $p_6 = 1$, the function w is concave if Δ^j_{j-1} is positive for all j, linear if Δ^j_{j-1} is zero for all j, and convex if Δ^j_{j-1} is negative for all j. Again, we allowed for response error in classifying subjects' weighting functions. The classification criterion used was motivated by a pattern observed in the data and will be explained in the next section.

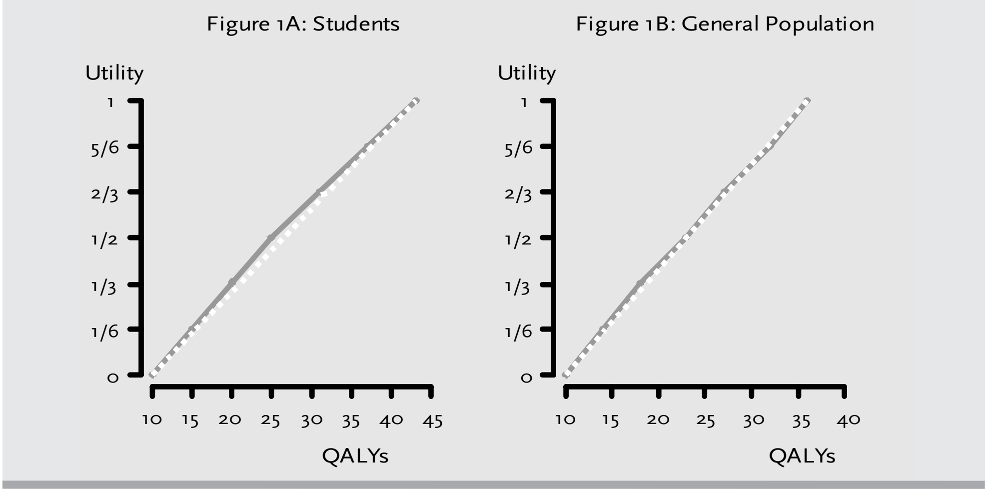
6.5 Results

Four subjects had to be excluded from the student sample. Three of them did not reach convergence because they did not value additional QALYs above some level, one subject violated rank-ordering of QALY-profiles in the second stage even after the computer had adjusted the stimuli. This left 65 subjects in the analysis of the student sample.

In the general population sample, 29 subjects had to be excluded: 14 subjects violated rank-ordering of QALY-profiles even after adjustment of the stimuli, 9 subjects did not reach convergence because they did not value additional QALYs above some level, 4 subjects found the task too difficult, the computer of one subject crashed and 1 subject refused to start the experiment. This left 179 subjects in the analysis of the general population sample.

Whereas the other exclusions are unlikely to have affected the results, the exclusions due to violations of rank-ordering may have had an effect on the results. In the questions for p_1, p_2 or p_3, the violations of rank-ordering may have reflected a desire to make the profile (x_{i+1}, p, z_m) more attractive. If so, $w(p_i)$, $i = 1, 2, 3$, would have to exceed its imposed upper bound and the exclusion of these subjects leads to a downwards bias in the estimated equity weights. In the questions for p_4 and p_5, the violations of rank-ordering may have reflected a desire to make the profile (z_m, p, x_{j-1}) less attractive. If so, $w(p_j)$, $j = 4, 5$, would

♠ **Figure 6.1** The elicited utility functions



have to fall short of its imposed lower bound and the exclusion of these subjects leads to a downwards bias in the estimated equity weights. Unfortunately, we do not know in which questions the subjects violated rank-ordering. We have some indication, however, that the effect of the violations was negligible, as we discuss in Section 6.7.

6.5.1 Elicitation of the utility function

♠ **Figure 6.1** shows the elicited utility functions over QALYs for both samples, based on the median data. The functions look similar when we use the mean data. Both utility functions were close to linear: the utility function for the student sample was slightly concave, whereas the utility function for the general population sample might be described as “linear with random error”.

The above observations were confirmed when we looked at the parametric estimates of the utility function, which are displayed in ◘ **table 6.3**. The linear utility function is the special case of the power function when the power coefficient is equal to 1, and it is the special case of the exponential function when the exponent is equal to 0. ◘ **Table 6.3** shows that in both samples, the mean and median power parameters were close to 1 and the mean and median exponential parameters were close to 0, suggesting that the assumption of linear utility over QALYs was reasonable at the aggregate level. The inter-quartile ranges show that individual coefficients varied considerably and that the above conclusion did not necessarily hold at the individual level.

The power coefficient was significantly different from zero, the case where utility is logarithmic, suggesting that Dolan’s (1998) Cobb–Douglas social welfare function did not fit our data well. For all three estimations, no significant differences were found between the coefficients in the student sample and those in the general population sample. We found in neither sample a significant difference in goodness of fit between the three parametric specifications.

◘ **Table 6.4**, which displays the results of the analysis of the individual data, shows that there was no predominant shape of the social utility function. In both samples, the proportion of subjects with a concave utility function, which corresponds to diminishing marginal utility, was slightly higher than either of the two other categories, but not

◘ **Table 6.3** Parameter estimates

	Parametric Families								
	Power			Exponential			Expo-Power		
	Median	Mean	IQR	Median	Mean	IQR	Median	Mean	IQR
Students	0.90	0.97	0.21	-0.32	-0.21	0.67	1.18	1.25	0.22
GP	0.96	1.07	0.32	-0.06	-0.01	0.98	1.25	1.35	0.34

IQR = inter-quartile range, **GP** = general population

Table 6.4 Classification of Subjects in Terms of the Shape of the Utility Function

	Concave (%)	Linear (%)	Convex (%)
Students	21.5	18.5	12.3
General Population	20.7	17.9	17.3

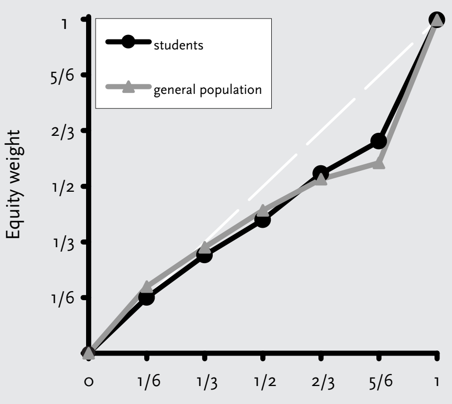
significantly so. One reason why there were relatively many subjects whose utility function could not be classified is that we used a rather strict classification criterion. For example, if a subject's standard sequence was equal to {10, 15, 19, 24, 28, 33} then he was classified as mixed even though his utility function was almost perfectly linear. We could of course have used a weaker classification criterion, e.g. the sign of Δ^j_{j-1} plus or minus the standard deviation of the responses, but this would have allowed for the possibility that, for a given Δ^j_{j-1}, a subject's utility function was both classified as convex and as concave, which seemed undesirable.

An easy heuristic for subjects to use in answering the utility elicitation questions would be to let $x_j - x_{j-1} = R - r$. This might have inflated support for the linear utility function. There were three subjects in the student sample and one in the general population sample who had such an answer pattern. This suggests that a large majority of our subjects did not use such a heuristic.

6.5.2 Elicitation of the equity weighting function

The number (proportion) of cases in which the computer had to adjust the stimuli to avoid a violation of rank-ordering was 15 (21.7%), 7 (10.1%), 0, 5 (7.2%) and 3 (4.3%) in

Figure 6.2 The elicited equity weighting functions



questions 1–5 in the student sample and 75 (41.9%), 35 (19.6%), 15 (8.4%), 38 (21.2%) and 29 (16.2%) in the general population sample. The conclusions did not depend on whether we used linear utility, power utility, exponential utility, or expo-power utility to compute the utilities of the z_m, $m = 1, ..., 5$. We, therefore, only report the results under the linear approximation.

♠ **Figure 6.2** shows the median equity weighting functions for both samples. The shape was similar: it was largely convex except for the first part which was linear for the student sample and slightly concave for the general population sample. Recall from Section 6.2 that a convex (linear/concave) weighting function corresponds to inequality aversion (neutrality/seeking). ♠ **Figure 6.2**, therefore, suggests that subjects were predominantly averse to inequalities in health, except when the size of the better-off group was small.

One possible reason why subjects may not have been uniformly inequality averse is that they did not properly take into account group size. There is a vast psychological literature showing that when people are dealing with relative frequencies, like proportions, they distort them in recognizable ways: people tend to overestimate small proportions and underestimate high proportions (e.g. Tversky and Kahneman, 1992; Gonzalez and Wu, 1999). We will label this type of behavior insensitivity to group size. In our study, insensitivity to group size would make that people perceive the better-off group as larger than it actually is when the proportion in the better-off group is small, say 1/6, and perceive the better-off group as smaller than it actually is when the proportion in the better-off group is high, say 5/6.

◘ **Table 6.5** shows the results of the individual analyses of the equity weighting functions. The hypothesis of insensitivity to group size, formulated above, would support concavity of the equity weighting function when the proportion of the better-off group is close to 0 and convexity of the equity weighting function when the proportion is close to 1. In other words, insensitivity to group size is inconsistent with convexity of the equity weighting function when the proportion of the better-off group is close to 0, and is inconsistent with concavity of the equity weighting function when this proportion is close to 1. To account for both insensitivity to group size and response error, we classified a subject's equity weighting function as concave if at least three values of Δ^j_{j-1} were positive and $(1 - w(5/6)) \leq 1/6$, i.e. there was no "downwards jump" in the equity weights near 1, as convex if at least three values of Δ^j_{j-1} were negative and $w(1/6) \leq 1/6$, i.e. there was no "upwards jump" in the equity weights near 0, and as linear if at least three values of Δ^j_{j-1}

◘ **Table 6.5** Classification of Subjects by the Equity Weighting Function

	Concave (%)	Linear (%)	Convex (%)	Insensitivity (%)
Students	7.7	0	41.5	38.5
General	3.9	0	31.3	54.2

were zero and not both $w(1/6) > 1/6$ and $(1 - w(5/6)) > 1/6$. To test the robustness of our findings, we also used the classification concave if at least three values of Δ^{j}_{j-1} were positive and Δ^{6}_{5} was not negative, linear if at least three values of Δ^{j}_{j-1} were zero and not both Δ^{6}_{5} negative and Δ^{2}_{1} positive and convex if at least three values of Δ^{j}_{j-1} were negative and Δ^{2}_{1} was not positive. The results were similar to those reported in ◘ **table 6.5**.

◘ **Table 6.5** shows that few subjects had a concave or linear equity weighting function. The proportion of subjects with a convex equity weighting function was much higher although still lower than 50%. The final column of ◘ **table 6.5**, which shows the proportion of subjects for whom both $w(1/6) > 1/6$ and $(1 - w(5/6)) > 1/6$, suggests that the behavior of a sizeable number of our subjects was consistent with insensitivity to group size. The above analysis suggests that the equity weights that we obtained were the product of both insensitivity to group size and what we may call “true” concerns for equality. To try and separate these two factors, we estimated the following parametric form for the equity weighting function:

$$w(p) = \frac{\delta p^{\gamma}}{(\delta p^{\gamma}) + (1 - p)^{\gamma}} \qquad ♣ \; (6.8)$$

This specification was first proposed by Goldstein and Einhorn (1987) for decision under risk. Gonzalez and Wu (1999) gave an interpretation for the parameters γ and δ, which

♠ **Figure 6.3** The elicited equity weighting function after correction for insensitivity to group size

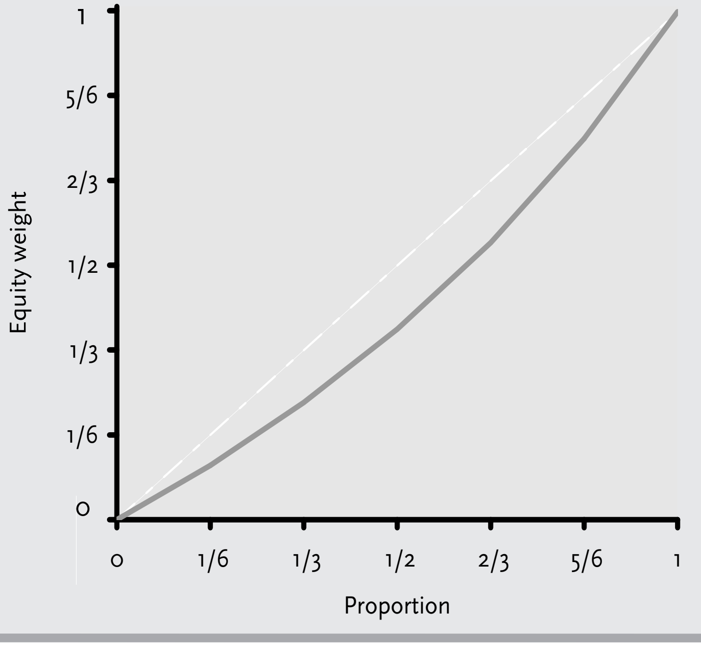


Table 6.6 Parameter Estimates for the Equity Weighting Function (♣ equation 6.8)

	Parameters					
	γ			δ		
	Median	Mean	IQR	Median	Mean	IQR
Students	0.68	0.69	0.32	0.59	0.69	0.32
General Population	0.56	0.61	0.44	0.63	0.73	0.47

IQR = inter-quartile range

with some modifications, also applies to the social decision context that we consider here. The parameter γ determines the curvature of $w(p)$ and, hence, the sensitivity to group size. Values less than 1 indicate insensitivity to group size and the lower γ is, the less sensitive the individual is to changes in group size. The parameter δ indicates the attractiveness of giving health gains to the better-off group, and thus measures preferences for equality. The lower δ is, the more equality-minded people are. Values less than 1 correspond to inequality aversion.

♣ **Equation 6.8** was estimated by a distribution-free iterative procedure that minimized the sum of the squared residuals. ◘ **Table 6.6**, which displays the results of the estimation, shows that insensitivity to group size and preferences for equality jointly determined the equity weights. Insensitivity to group size was stronger in the general population sample. The difference in the estimate of γ is significant by the nonparametric Mann–Whitney test ($p = 0.004$), but only marginally so by the independent samples t-test ($p = 0.067$). Aversion to inequality, measured by δ, was similar in the two samples.

Because the parameters γ and δ are largely independent, we can use ♣ **equation 6.8** to correct for the impact of insensitivity to group size on the equity weights by setting $\gamma = 1$. ♠ **Figure 6.3** shows the equity weights when $\gamma = 1$ and $\delta = 0.6$, the case which corresponds to our median data. The figure shows, for example, that the weight given to the better-off group was 0.375 when the size of the better-off group was equal to half the cohort.

6.6 Implementation in health policy

To illustrate the implications of our findings, we computed equity-adjusted cost-utility ratios for 12 treatments. To perform these computations, we made two assumptions. These assumptions are not innocuous and we therefore urge the reader to interpret the equity-adjusted cost-utility ratios with caution. These ratios serve as an illustration of how our method can be applied in practice, not as a guide to policy making. The first assumption is that we can extrapolate outside the domain of estimations. We found that the social utility function over QALYs was roughly linear on the interval [10, 40]. The data suggest that linearity also held on [5, 10]. Linearity on [10, 40] means that $U(x_j) - U(x_{j-1})$, $j = 1, \ldots, 6$, is about 5. From ♣ **equation 6.4** and ♣ **equation 6.5**, we know that $U(x_j) - U(x_{j-1})$

is equal to $((1-w(p)) / w(p)) (U(8) - U(5))$. We also found that $w(1/2)$ was about equal to 0.4. This implies that $U(8) - U(5)$ was close to 3, which is consistent with linearity. However, we do not know whether U was also linear on [0, 5] and on [40, →). In fact, when we looked at those subjects for whom x_6 exceeded 50 years then we found more concavity than in the general sample, suggesting that the assumption of linearity of the social utility function on [40, →) need not hold. Similarly, we did not estimate any equity weights on (0, 1/6) and on (5/6, 1). We had to assume that the estimated pattern of equity weighting on [1/6, 5/6] can be extrapolated to these two subdomains.

Our second assumption is that it is better to use the equity weights that are corrected for insensitivity to group size than the uncorrected ones. That is, we will use ♣ **equation 6.8** with $\gamma = 1$. We used the corrected equity weights because we believe that insensitivity to group size, which arises because of people's limited cognitive abilities, is a bias in people's preferences that ought to have no impact on health policy. We realize that this assumption is controversial. After all, some of what we are correcting for may be true equity preference. Nevertheless, we believe that the corrected equity weights were closer to subjects' true equity weights than the uncorrected weights.

Most of the selected treatments were taken from Stolk et al. (2004); data on the remaining conditions were obtained through personal communication. To adjust cost-utility ratios for equity considerations we computed the distribution of QALYs within the Netherlands, on the basis of mortality figures (CBS, 2003) and quality of life estimates (Toenders, 2002). The distribution is displayed in the first two columns of ◘ **table 6.7**. We then computed the equity weights for a patient in each of the groups. The equity weights were

◘ **Table 6.7** Distribution of QALYs and Equity Weights

		Equity Weight	
Lifetime QALYs	Proportion (%)	$\gamma = 0.56, \delta = 0.63$	$\gamma = 1, \delta = 0.6$
<1	0.55	26.81	1.56
1-15	0.27	12.29	1.55
15-30	0.73	8.87	1.54
30-40	1.06	6.18	1.52
40-50	3.15	4.04	1.48
50-55	4.28	2.66	1.41
55-60	6.40	1.88	1.32
60-65	11.07	1.36	1.19
65-70	20.38	1	1
70-75	26.54	0.88	0.79
75-80	21.40	1.24	0.64
80-82.5	3.32	2.82	0.57
>82.5	0.85	8.23	0.56

computed using ♣ **equation 6.8**. The third column shows the equity weights when we used the parameter values that best fitted our data in the general population sample, $\gamma = 0.56$ and $\delta = 0.63$, the fourth column shows the equity weights after correction for insensitivity to group size, $\gamma = 1$, and using $\delta = 0.60$. We rescaled the equity weights so that the weight given to a patient with expected lifetime QALYs between 65 and 70 was equal to 1. This scaling is based on Williams (1997), who suggested that a person's fair innings was approximately 70 QALYs. The third column of the table shows the effect of insensitivity to group size: individuals who are in the tails of the QALY distribution get more weight than those who are closer to the middle of the distribution. The fourth column shows that this, counterintuitive, effect disappears after correction for insensitivity to group size. Then, the weights are monotonically decreasing.

◘ **Table 6.8** displays the results of adjusting the cost-utility ratios for equity concerns. The first column describes the conditions that we studied, the second the treatments for these conditions. The third column shows for each treatment the costs per QALY gained when no equity weighting was applied, i.e. under the common procedure of aggregating QALYs. The fourth column shows the ranking of the treatments in terms of cost-effectiveness when no equity weighting was applied. As the table shows, surgery for congenital anorectal malformation was the most cost-effective treatment and lung transplantation for pulmonary hypertension was the least cost-effective treatment.

The fifth column shows for each disease the number of expected lifetime QALYs that the average patient obtains without treatment. The sixth column gives the rescaled equity weights that were obtained after correction for insensitivity to group size. These weights can directly be read off from ◘ **table 6.7**. The seventh column shows the cost-utility ratios adjusted by these equity weights. The final column shows the ranking of the treatments in terms of equity-adjusted cost-utility ratios. As expected, there were some shifts in ranking in favor of treatments aimed at patients with lower expected lifetime QALYs. For example, the cost per QALY of statins was higher than that of terbinafine when no equity weighting was applied, but statins were more cost-effective than terbinafine when cost-utility ratios were adjusted for equity concerns.

6.7 Discussion

6.7.1 Main findings

In this paper, we have elicited, both in a sample of students and in a sample from the general population, the trade-off between equity and efficiency in the allocation of health. We assumed the nonlinear rank-dependent QALY model, a model that encompasses many of the social welfare functions that have been proposed in the literature. A correction for utility curvature was applied but we found that, on the aggregate level, social preferences were approximately linear in QALYs. People were generally inequality averse, except when

Table 6.8 Equity adjusted cost-utility ratios

Condition	Treatment	Cost per Qaly	Rank	Life-time QALYs	Equity weight*	Eq. Adj. Cost per QALY	Rank
CAM	Surgery	2,482	1	9.4	1.55	1,601	1
Erectile Dysfunction	Sildenafil	5,656	2	77.0	0.64	8,838	3
Non-Hodgkin Lymphoma	Chemotherapy	7,771	3	73.3	0.79	9,837	4
Artherosclerosis	Clopidogrel	11,629	4	54.9	1.41	8,248	2
BPO	Finasteride	12,788	5	80.7	0.57	22,435	7
Onychomycosis	Terbinafine	16,843	6	83.7	0.57	29,549	9
Osteoporosis	Oestrogen	18,151	7	83.2	0.57	31,844	11
High Cholesterol	Statins	18,151	7	56.1	1.41	12,873	5
Metastatic Breast Cancer	Chemotherapy	22,441	9	56.1	1.41	15,916	6
Heart Disease	Heart transplant	38,206	10	42.2	1.48	25,815	8
End-stage Renal Disease	Kidney Repl.	44,607	11	57.8	1.41	31,636	10
Pulmonary Hypertension	Lung Transplant	79,412	12	41.6	1.48	53,657	12

* Equity weigth for γ=1; δ=0.6; **CAM** = Congenital Anorectal Malformation, **BPO** = Benign Prostatic Obstruction, **Kidney repl** = kidney replacement by dialysis or transplant

table

the better-off group was small. The reason why we found no global inequality aversion may be insensitivity to group size. Global inequality aversion was observed when we corrected for insensitivity to group size. Few differences were observed between the sample of students and the sample from the general population.

6.7.2 Possible biases

As noted in Section 6.5, the exclusions due to violations of rank-dependency may have affected the results. We tested for the effect of these exclusions by making the extreme assumption that the excluded subjects violated rank-dependency in every question. This assumption means that these subjects had the highest equity weights of all subjects in the questions for p_1, p_2 and p_3, and the lowest equity weights in the questions for p_4 and p_5. Such a preference pattern is unlikely and the assumption is almost certainly too extreme, which means that the actual bias will be smaller, but the analysis gives an indication of the maximum effect of the exclusions on the median equity weights. Under the assumption, the median equity weights for 1/6, 1/3, 1/2, 2/3 and 5/6 changed by 0, +0.003, +0.006, -0.007 and 0, respectively, in the student sample and by 0, +0.015, +0.026, -0.022, -0.025 in the general population sample. Hence, even under an extreme assumption about the effect of the exclusions due to violations of rank-ordering, the effect of these exclusions was small.

A frequently encountered problem in preference assessment tasks is that people have a tendency to respond in round numbers, often multiples of five, which can lead to bias. Because a choice-based procedure was used, round answers were less likely in our study. In fact, the proportion of round answers (multiples of five) was 21.7% in the student sample and 21.3% in the general population sample, which are not significantly different from 20%, the proportion of round answers expected when people do not have a tendency to use round answers.

It may have been possible that some subjects did not understand the concept of a QALY properly, leading to additional response error. It would have been easier to perform the experiment with years of life instead of QALYs. We opted to use QALYs, because policy makers and researchers are most interested in the trade-off between equity and efficiency as measured by QALYs. Upon questioning by the experimenter, most subjects seemed to understand the concept of a QALY well. To complete the exercise, they generally assumed that people in the cohort lived in relatively good health for the largest part of their life, and that the largest QALY loss was related to life-years lost.

Our findings depend on the validity of the nonlinear rank-dependent QALY model, ♣ **equation 6.8**. Even though ♣ **equation 6.2** is quite general, it may in some cases be too restrictive. The model assumes, in particular, that the equity weights depend only on individuals' relative positions, their rank, and not on absolute differences between the amounts of QALYs received. If this assumption does not hold then our results may no longer be valid. Another violation would occur if there is no separability between the equity weights and the utility for QALYs. In that case, the elicitation of the utility for QALYs might depend on the proportion used. We could have used a more general model than ♣ **equation 6.2** to take these possible violations into account. This would, however, have led to a model that is more difficult to apply in practice. The question is whether violations of the nonlinear rank-dependent QALY model, if any, are sufficiently widespread and serious to justify giving up the tractability of the model.

Finally, it is possible that, even though we tried to control for it, asymmetric errors may still have affected the results. If this were true, then these errors will have had most effect on $w(1/6)$ and $w(5/6)$, biasing $w(1/6)$ upwards and $w(5/6)$ downwards. The effect on the other three weights that we elicited is probably negligible, because in these estimations the stimuli were not close to the bounds and there was enough room for error "on both sides". Our main finding of a generally convex equity weighting function, i.e. aversion to inequality, is confirmed when we only look at $w(1/3)$, $w(1/2)$ and $w(2/3)$, giving grounds for confidence in the results.

6.7.3 Final remarks

Our study suggests that people are averse to inequalities in health. If people's societal preferences ought to have a place in health policy, then our findings connote that QALYs should be weighted for equity concerns. We have shown that the rank-dependent QALY model can be used for this: we have presented a method to elicit the equity weights under the model and we have shown how these equity weights can be implemented in health policy. We repeat that the purpose of the latter exercise was illustrative; before more robustness checks are performed, restraint should be exercised in using the data we presented in actual policy making.

Finally, a few words about the equity concept we used are in order. Because we studied people's preferences over allocations of lifetime QALYs, our study focused on differences in lifetime health expectancy between groups of newborns. This setup implicitly assumed that the desirability of a distribution depends on people's (expected) lifetime health. In that sense, our approach is close to Williams' fair innings approach. Several authors have discussed other concepts of equity and have argued that equity may also be concerned with other issues, such as patients' actual health state and when and how health losses occur (Culyer and Wagstaff, 1993; Cuadras-Morató et al., 2001; Dolan and Olsen, 2001). Our empirical results have little bearing in case such equity concerns are adopted. How these other equity concerns can be operationalized, remains, therefore, an open question.

Appendix 6A. Explanation of QALYs

In this experiment, health is described in terms of quality-adjusted life-years (QALYs). Quality-adjusted life-years are a measure of health and can be calculated by multiplying life-years by a numeric value that reflects quality of life during those years. A year in full health counts as 1 QALY. A year in which people are confronted with health problems counts as less than 1 QALY. For example, I consider myself to be in full health. As long as I stay in full health each year I live counts as 1 QALY. But suppose that I had arthritis then each year would count as less than 1 QALY. If we assume, for example, that pain and mobility reduce my quality of life by 50%, then each year that I live in this health state counts as 1/2 QALY. The questionnaire specifies how many QALYs a subgroup of a cohort will get. If the number of QALYs is high, you can be sure that the people live long and that heir quality of life is good. If the number of QALYs is low, then this number of QALYs can be the result of either a long life with severe disability or a short life with no disability.

Appendix 6B. Presentation of the experimental questions (♠ figure A.6.1)

Consider a population of newborns that suffer from a particular disease. You must choose one of two treatments for the cohort which have identical costs, but a different distribution of QALYs.

Treatment A gives

½ of the cohort	40 QALYs	and	½ of the cohort	10 QALYs

Treatment B gives

½ of the cohort	30 QALYs	and	½ of the cohort	15 QALYs

>> You are doing fine. Please continue.

- ■ I prefer treatment A
- ■ I prefer treatment B

Continue

Appendix 6C. Explanation of the PEST procedure

The PEST procedure obeys the following four rules:

1. On every reversal of step direction, halve the step size.
2. The second step in a given direction, if called for, is the same size as the first.
3. The fourth and subsequent steps in a given direction each double their predecessor, except that large steps may be disturbing to a human observer and an upper limit on permissible step size may be needed.
4. Whether a third successive step in a given direction is the same as or double the second depends on the sequence of steps leading to the most recent reversal. If the step immediately preceding that reversal resulted from a doubling, then the third step is not doubled, while if the step leading to the most recent reversal was not the result of a doubling, then this third step is double the second. Doubling occurs on the first three responses in the same direction.

Consider the following example:

A. 1/2 the cohort gets *X* QALYS and 1/2 the cohort gets 5 QALYs;

B. 1/2 cohort gets 30 QALYs and 1/2 the cohort gets 8 QALYs.

The initial increment for change was 4 QALYs. The stopping rule occurred when an incremental change in QALYs in option A is less than 2 QALYs. The first step is to select a random starting value of *X* in some interval, say (30, 100). This interval depended on the stimuli in the question. Suppose that $X = 70$. ◘ **Table A.6.1** illustrates the PEST procedure.

suppose

Note that the PEST procedure can correct for errors. In the example above we began zeroing in at Trial 5. However, if a subject got to Trial 7 and had made some errors, he could break out of the convergence by choosing A A A or B B B during the next several trials. As mentioned in the main text, we included two random 'filler' trials after each real trial so that the subject did not know convergence was happening.

Table A.6.1 Illustration of the PEST procedure

Trial	X	Choice	Comment
1	70	A	Random selection
2	66	A	First change
3	62	A	Rule 2
4	54	A	Rule 4
5	38	B	Rule 3
6	46	A	Rule 1
7	42	B	Rule 1
8	40	A	Rule 1
9	41	B	Stopping rule

Part of Damien Hirst's The sleep of reason

Chapter 7
Criteria for determining a basic health services package. Recent developments in the Netherlands

The criterion of medical need figures prominently in the Dutch model for reimbursement decisions as well as in many international models for health care priority setting. Nevertheless the conception of need remains too vague and general to be applied successfully in priority decisions. This contribution explores what is wrong with the proposed definitions of medical need and identifies features in the decision-making process that inhibit implementation and usefulness of this criterion. In contrast to what is commonly assumed, the problem is not so much a failure to understand the nature of the medical need criterion and the value judgments involved. Instead the problem seems to be a mismatch between the information regarding medical need and the way in which these concerns are incorporated into policy models. Criteria-medical need, as well as other criteria such as effectiveness and cost-effectiveness-are usually perceived as hurdles, and each intervention can pass or fail assessment on the basis of each criterion and therefore be included or excluded from public funding. These models fail to understand that choices are not so much between effective and ineffective treatments, or necessary and unnecessary ones. Rather, choices are often between interventions that are somewhat effective and/or needed. Evaluation of such services requires a holistic approach and not a sequence of fail or pass judgments. To improve applicability of criteria that pertain to medical need we therefore suggest further development of these criteria beyond their original binary meaning and propose meaningful ways in which these criteria can be integrated into policy decisions.

Chapter 7

Criteria for determining a basic health services package
Recent developments in The Netherlands

Stolk EA, Poley MJ, Criteria for determining a basic health services package. Recent developments in The Netherlands. European Journal of Health Economics 2005; 6(1):2-7

7.1 Introduction

After long periods of relative abundance the Dutch healthcare system must now cope with limited resources. The government therefore has ceased or limited funding of several treatments as from 1 January 2004 (e.g., oral contraceptives, in vitro fertilization, adult dental care, and physiotherapy), and more exclusions have followed since than. Remarkably, many of these exclusions have been discussed before. However, previous governments were unable to enforce negative reimbursement decisions. Therefore these decisions reflect an important change in the attitude towards priority setting.

The will to succeed in priority setting may be a result of ever-increasing budget pressures, but also proposed changes towards market regulation of the Dutch health care sector seem to have stimulated the debate on priority setting. Assuming that competing insurers and health care providers will be in a better position than the government to improve efficiency, the imminent reforms aim to limit the regulating role of the state and to introduce competition among insurers and health care providers (Ministry of Health Welfare and Sports, 2004). Even though a movement towards more competition is propagated all over Europe, it seems a risky transition. The key question is whether market orientation can indeed promote efficiency. It could well be hypothesized that this objective will not be met because of obstacles to full and fair competition (Andersen et al., 2001). The consequences of market failure could be dramatic. The reason is that to reform the system, the government will have to abandon several traditional mechanisms to regulate health care expenses (e.g., central planning and budgeting). Its main means to influence health care expenditures will be through decisions about the composition of the benefit package and shifts in the nature of health care financing. Unfortunately, methods for prioritization or installation of user charges have been applied with little success in the past. Therefore, pessimistically, it may be that the reforms produce the opposite effect: costs may actually rise because the government is giving up its means to contain expenditures.

It is not surprising that the imminent reforms placed priority setting back high on the political agenda. This has resulted in new efforts to improve the applicability of the criteria that are to guide priority decisions in The Netherlands. The purpose of this contribution is to reflect on the shifting attitude towards priority setting, to inform about the latest developments in The Netherlands, and to reflect on the question whether current developments will indeed result in more effective priority setting.

7.2 Dutch history of priority setting

Since the early 1990s the Dutch approach to priority setting has concentrated on four criteria that were introduced by the Government Committee on Choices in Health Care (more commonly known as the Dunning Committee after its chairman): to be eligible for reimbursement, care must be necessary, effective, efficient and cannot be left to the individual's own responsibility (Government Committee on Choices in Health Care, 1992). These criteria reflect two values. Effectiveness and efficiency represent a coherent approach to priority setting using evidence-based medicine. The other criteria explicate the political viewpoint that not all (cost-effective) medical interventions must be paid for by collective means. Instead it was argued that only interventions that improve or maintain normal functioning (i.e., necessary care), and that cannot be left to one´s own responsibility should be paid for collectively. Public debate has followed the proposal of these criteria, revealing broad support. The explicitness of criteria and the involvement of the public are both distinctive and appreciated features of the Dutch approach toward priority setting.

Unfortunately, attempts to apply the criteria in health care decision making have made clear that it was not easy to reach agreement on their meaning nor to apply them to specific procedures (Van de Vathorst, 2001). There are several examples of governments excluding certain provisions from insurance to include them again only shortly after (e.g. dentures). Frequently even only the announcement about excluding services from reimbursement caused so much commotion that plans were postponed. Such was recently the case with the rollator. In vitro fertilization is an example of a therapy on which opinions were deeply divided as to whether it constitutes necessary and effective care. Are involuntarily childless couples impaired in normal functioning? Similar problems were posed by contraceptives, where it was questioned whether treatment can be left to individual responsibility, and whether own payments would be acceptable given that socioeconomic health inequalities should be avoided. Because of heterogeneous views many similar issues were undecided, and priority setting reached an impasse.

Gradually one started to become convinced that the idea of devising a simple set of rules for priority setting was flawed and would not work in practice. The government therefore seemed to shift its focus from limiting coverage of the benefit package to ensuring that services would be provided appropriately, i.e., to those patients most likely to benefit. After all, the (cost-)effectiveness of health care is less a characteristic of a treatment per se than of its application by physicians to patients with particular diagnoses. As a result many policy measures were implemented to steer the decisions of practitioners, health care institutions, and insurers, through either financial or nonfinancial incentives. Examples are stimulation of guideline development, conditional reimbursement strategies, the installation of expert committees to audit physicians´ practices, record keeping of provided treatments, and changes in the financing of hospital care.

However important it is to strengthen participation of professionals at different levels of health care priority setting, it does not imply that explicit choices—with transparent underlying value judgments—can be avoided. If all choices in health care would have to be made on a decentralized basis, a system of 'covert rationing' would emerge. Heterogeneous views on what constitutes effectiveness or necessity will continue to exist, but they will be hidden from view (Ham and Robert, 2003; Norheim, 1999). In such circumstances patients' wishes can be fulfilled only through an accumulation of private decisions, local initiatives, and perhaps sheer chance. Although explicating value judgments will inevitably evoke protests (because persons are not accustomed to limits being set to their entitlements to health care), priority setting based on transparent, explicit criteria is still to be preferred by far (Robinson, 1999). Although there will always be still other factors that influence the chance of reimbursement (e.g., budget impact and uncertainty around clinical and economics effects (Devlin and Parkin, 2004; Al et al., 2004), it is important to explore how well criteria for reimbursement decisions match the objectives that are seen as relevant to decision making.

7.3 Necessary care

Of the four criteria for reimbursement decisions (necessity, effectiveness, cost-effectiveness, and own responsibility), the criterion with the largest unused potential seems to be 'necessary care'. This is commonly attributed to the problem that a clear standard to evaluate necessity or need is lacking. This problem obviously refers to measurement issues, but also issues of fairness are at stake. The meaning of necessary care seems straightforward on the surface. If you are sick, whatever makes you well again is medically necessary. If you are in good health, all that keeps you well is medically necessary. However, between these extremes there may be a wide range of services that can be considered more or less necessary, depending on circumstance (Commission on the Future of Health Care in Canada, 2002). This may explain why medical necessity is so hard to define, and why drawing a strict line between necessary and not necessary care may be perceived as unfair: minor differences in the perceived need between people who qualify for reimbursement and those who do not would have serious financial consequences.

Underlying the Dutch attempt to improve functioning of necessary care was the recognition that the Dunning Committee used necessary care to determine a hierarchy of conditions for which treatment is needed, while health policymakers used it as a binary criterion. Necessary care has been used only to locate a cutoff point that distinguishes between treatments that should be included in the basic benefit package and those that should not. Assuming, however, a hierarchy of necessity, this policy model ignores a wealth of information. It was therefore proposed to adopt a continuous scale (an intervention is more or less necessary) instead of a binary one (an intervention is either necessary or not). If policy decisions were to reflect gradations of necessity, the measurement issue is likely to be easily solved. Intuitively it is a small step from the immeasurable concept of 'necessary care' to the concept "burden of disease" that can be expressed in quality of life and life years.

The real problem is the application to policy. In the last 4 years much attention of policy makers has been devoted to the integration of a hierarchy of medical need in policy models. The guiding principle was that the relative need should in some way be reflected in the costs that society is willing to pay to treat a patient (Toenders, 2001). Two applications have been considered. First the relative definition of medical need could be related to efficiency. This means that the relative efficiency criterion should be applied differently when the disease problem is more or less disabling, by varying the cost-effectiveness threshold in reimbursement decisions according to burden of disease. Simultaneous assessment of efficiency and medical need may improve functioning of both criteria. However, it is difficult to screen the whole benefit package in terms of necessity and cost-effectiveness. There are about 11,000 different medicines on the market in The Netherlands (Ministry of Health Welfare and Sports, 2002). Establishing a relationship between medical need and financial constraints would be easier if economic evaluations of all covered services could be avoided. The second application therefore proposes to vary user charges on the basis of medical need. This strategy is used, for example, in Belgium (Annemans et al., 1997).

7.4 Varying the cost-effectiveness threshold

The first application of necessary care as a relative criterion entails using the severity of a patient´s condition to determine the applicable cost-effectiveness threshold. This approach has attracted attention of health economists, who noted that policy makers are reluctant to base resource allocation decisions on efficiency alone because distributional aspects of health are then ignored. Health economists therefore have explored the idea of a tradeoff between efficiency and severity, directing research towards answering the questions of how the severity of a condition should be established, and what cost-effectiveness threshold applies to treatments that target conditions of different severity. ♠ **Figure 7.1**, for example, shows how empirically derived estimates of a tradeoff between equity and efficiency translate into varying cost-effectiveness thresholds that may apply to different patient groups that are characterized by the number of quality-adjusted life-years which patients will attain if no treatment is provided (see Bleichrodt et al., in press). Governments that seek to apply a relationship between medical necessity and efficiency will therefore have access to late-stage development policy models.

It will be difficult, costly, and time consuming to gather economic evidence for all interventions that are currently included in the basic benefit package. A solution is then to apply the tradeoff between severity of a condition and cost-effectiveness only at the intake of new treatments in the benefit package. In The Netherlands such an intake

procedure can be established rather easily. From 2005 onward a pharmacoeconomic study is formally required for drugs seeking reimbursement. However, it has been made clear that efficiency relates to probability of funding, but that there will not be a single threshold. Legislation gives the minister of health the discretionary power to evaluate economic evidence with respect to "importance to public health". It looks as if this discretionary power will be used to balance concerns of efficiency against other concerns like severity of a condition (i.e., relative need), budget impact, and therapeutic value.
Adoption and explication of this policy model would improve transparency and perceived coherence of reimbursement decisions. However, it should be recognized that the potential for actual cost containment might be limited. Policy makers already seem to make more use of the necessity criterion than they are fully aware of. A recent study analyzed the gatekeeping function of the Dutch reimbursement system in the pharmaceutical sector. This study related the satisfaction of explicit (clinical and economic) and implicit (grouped under the heading "importance to public health") criteria to services being accepted or rejected for inclusion in the benefit package. In this study severity of the disease emerges as a prominent decisive yet implicit criterion (Pronk and Bonsel, 2004). A relationship between medical necessity and cost-effectiveness thus already has been observed in The Netherlands, which is also true for other countries (Devlin and Parkin, 2004).

7.5 Necessity related to user charges

A second application involves categorization of the health care benefit package by relative need, introducing a system of various levels of user charges alongside it. There is ample room for such systems to have a substantial effect on the health care expenditures in

♠ **Figure 7.1** Relationship between cost-effectiveness and disease severity measured as the patients no treatment quality-adjusted life-year (QALY) total, varying around the values 20000 euro per QALY (lower curve) and 40000 euro per QALY (upper curve)

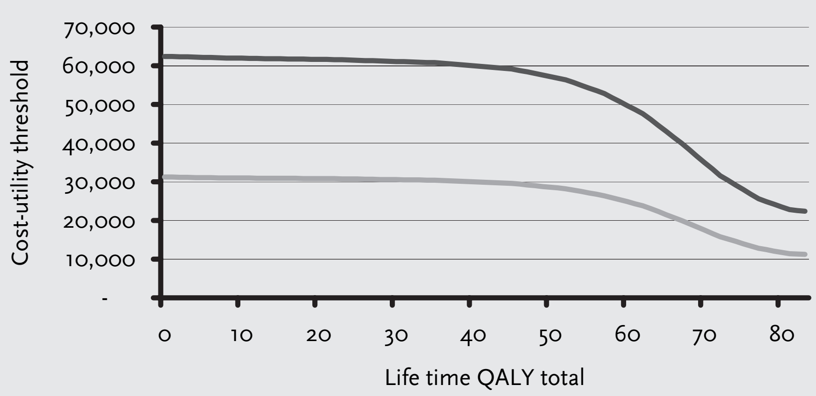


Table 7.1 Comparison of the Dutch proposal for categorization of medicines to the Belgium regulation

Drug	ATC	Condition	Reimbursement	
			Dutch proposal*	Belgium (%)
Acetylsalicylic acid	B01AC06	Thrombosis	2	75
Amlodipine	C08CA01	Angina pectoris, hypertension	2	75
Atorvastatine	C10AA05	Hypercholesterolemia	1 and 2	100–75
Budesonide	R03BA02	Obstructive airway diseases	2	75
Diclofenac	M01AB05	Rheumatic arthritis	2	75
Enalapril	C09AA02	Hypertension, coronary failure	2	75
Furosemide	C03CA01	Asthma cardiale, or edema as a result of coronary failure	2	75
Insulin	A10AD01	Diabetes mellitus	1	100
Levonorgestrel, estrogen	G03AA07	Hormonal contraceptives for systemic use (birth control)	2 or 3	20
Metoprolol	C07AB02	Angina pectoris/cardiac arrhythmia	2	75
Omeprazol	A02BC01	Ulcus duodeni, ulcus ventriculli	2 or 3	75
Oxazepam	N05BA04	Anxiety, tension, sleeping disorder	3	Not included
Paracetamol	N02BE01	Fever and pain	4	0
Paroxetine	N06AB05	Depression	1 or 2	75
Pravastatine	C10AA03	Hypercholesterolemia	1 and 2	100–75
Salmeterol,	R03AK06	Obstructive airway disease	2	75
Simvastatine	C10AA01	Hypercholesterolemia	1 and 2	100–75
Temazepam	N05CD07	Anxiety, tension, sleeping disorder	3	Not included

*Assumed levels of reimbursement: **1**=100%, **2** and **3** intermediate levels, **4**=0%. Some drugs can be classified into 2 categories depending on diagnosis (e.g., statins are classified into categories 1 and 2 for primary and secondary prevention). For other drugs unambiguous classification was difficult because a disease label covers a heterogeneous patient group (e.g. depression), and therefore classification may depend on the individual patient profile

The Netherlands because currently noncovered services make up only 3% of total health care expenditure. However, implementation of such a system will not be easy because it requires structural changes to the system of financing.

In The Netherlands user charges have been avoided with an appeal to solidarity. However, ideological currents seem to change, under the recognition that user charges not always are detrimental to solidarity. First, in a system without user charges countering moral hazard is difficult. The result is that the premium rates will have to be set higher than strictly necessary, which an efficiency loss is and may lead to declining levels of solidarity. Furthermore, increasing budget pressure may lead to the situation that no longer universal coverage is possible for all interventions that benevolent decision makers want to provide to the sick. Then the dichotomous decisions of whether to reimburse do not tie in with the observation that not all the sick live in equally bad health and not all who are denied treatment are in perfect health. Consequently the reimbursement scheme may be deemed unfair from the viewpoints of horizontal (persons in equal need should be treated the same) and vertical equity (persons with greater need should be treated more favorably than those with lesser needs). A way out of this can be offered by the proposed system of user charges because needs are satisfied in a proportional way. Moreover, user charges act as a form of cross-subsidy, ensuring that a larger benefit package can be maintained.

The suggestion to connect necessity of care to user charges is delivered in a time that much debate is going about the best way to introduce more copayments to Dutch health care. The government recently decided to introduce a no-claim discount. Everybody will have to pay higher premiums for health insurance, but those persons who consume little care (less than €255) will receive back the difference between the €255 amount and the care consumed. Both no-claim and direct copayments aim to decrease moral hazard and increase own responsibility, but the no-claim system was preferred because it was felt that it would not constitute as much a financial barrier to care as a direct payment would. To ensure this even further, visits to the general practitioner will not fall under the no-claim. However, the no-claim system fails to address the problem of increasing budget pressures because its revenues are limited. Moreover the approach still offends equity concerns, because the sick are less likely to be entitled to a no-claim refund. User charges related to relative need offer a more flexible approach because regardless of available resources, it integrates an idea about fairness in its aim to reduce the use of services that offer the poorest balance between severity of the disease, cost of treatment and its effects.

Several countries already apply systems in which medicines are classified into different categories that define the level of copayment. The French, for example, receive 100% reimbursement for 'essential' drugs, 65% for 'important' drugs, and 35% for 'comfort' drugs (Pelen, 2000). A similar system exists in Belgium, where drugs are classified into six categories (Annemans et al., 1997). To explore the potential returns of setting up a similar system in The Netherlands the Healthcare Insurance Board compared a Dutch proposal for categorization of drugs on the basis of medical need to the existing system in Belgium using the ten most frequently prescribed and ten most costly drugs as examples (◘ **table 7.1**). Although it was sometimes difficult to classify an intervention in a particular category, the table suggests that much could be gained from introducing such a refinement of the reimbursement system.

7.6 Mild diseases

Irrespective of the way in which medical necessity is balanced against financial arguments, the measurable definition of necessary care may improve priority setting. In the past, few services were excluded from funding due to the intractable definition of necessary care. With the introduction of a measurable definition it becomes easier to identify which services could be excluded from reimbursement, namely treatments for mild conditions. Two recent studies have attempted to identify conditions to which this argument would pertain (Bonsel et al., 2003; Wieringa et al., 2003; Poleij et al., 2002). Depending on what measurement method and what cutoff point are considered appropriate, several services that target conditions listed in ◘ **table 7.2** could be excluded from funding.

savings

The yearly cost savings could vary from €93,000 (no coverage if the quality of life impairments are less than 5% according to both measures) to €180,000 (no coverage if the quality of life impairments are less than 15%), representing approximately 2.5–5.0% of the medicine budget (Wieringa et al., 2003). Noteworthy is that many of the conditions listed can be treated with over-the-counter drugs which are already excluded from reimbursement in The Netherlands. However, additional cost savings could still be considerable.

7.7 Conclusion

The imminent health care reforms have revitalized the debate about criteria that can be used in priority setting. The criteria that were formulated by the Dunning Committee in 1991 still seem relevant in decision making. However, the performance especially of one of these criteria—necessary care—should be improved. We have therefore pleaded to give this criterion a measurable definition, and to be explicit about the way in which it is balanced against financial constraints. For the latter, this contribution discussed the possibilities of differentiating the cost-effectiveness thresholds or the level of copayments. The purpose is to introduce a subsidy from one class of sick to the other, which allows for expansion of the package and improves the underlying solidarity basis.

The proposed changes in the criteria for priority setting reflect that we are clearer now on the objectives of priority setting than we were in the past, and that we better understand the means available to reach those goals. However, further research is warranted to make these ideas applicable to health policy. In our view, it is especially relevant to further explore the proposal of differentiating reimbursement levels according to medical need. For a long time various countries have explored different ways to determine core services to be included in the benefit package. Although approaches may have differed, the resulting benefit packages are largely the same in different countries. Perhaps the debate on priority setting has focused too much on criteria for coverage decisions, while

Table 7.2 List of mild disease for which treatment may not be considered necessary (Bonsel et al., 2003; Poleij et al., 2002)

Condition	Qol weight EQ-5D+	Qol weight MIDAS
Common cold	1.00[a]	1.00[a]
Pharyngitis	1.00[a]	0.99[a]
Otitis externa	1.00[a]	0.98[a]
Cystitis	1.00[a]	1.00[a]
Acute sinusitis	1.00[a]	0.99[a]
Acute otitis media, myringitis	1.00[a]	0.99[a]
Acute tonsillitis	1.00[a]	0.99[a]
Acute bronchitis, bronchiolitis	0.99[a]	0.99[a]
Seborrheic dermatitis	0.99[a]	0.96[a]
Allergic rhinitis	0.99[a]	0.99[a]
Peritonsillar abscess	0.98[a]	0.95[a]
Exacerbation of contact dermatitis	0.98[a]	0.98[a]
Exacerbation of constitutional eczema	0.97[a]	0.95–0.89
Pityriasis versicolor	0.97[a]	0.98[a]
Folliculitis	0.96[a]	1.00[a]
Onychomycosis	0.96[a]	1.00[a]
Tinea pedis	0.96[a]	1.00[a]
Infectious conjunctivitis	0.94	1.00[a]
Chronic sinusitis	0.93	0.98[a]
Irritable bowel syndrome (spastic colon)	0.93	0.96[a]
Alopecia androgenetica	0.93	–
Erythrasma	0.93	1.00[a]
Urogenital candidiasis	0.93	1.00[a]
Gastritis, duodenitis	0.92	0.98–0.93
Acne vulgaris	0.90	0.98[a]
Hemorrhoids	0.89	0.95[a]
Osteoporosis	0.85	0.97–0.96[a]
Climacteric symptoms (menopause)	0.85	0.94

[a] 'Mild diseases' using the criterion Qol > 0.95

too little attention has been directed to decisions about restrictions to the patients' entitlements to covered services. It is likely that larger differences between international health care systems can be found in terms of specific entitlements, and that decision making at this level can be improved (Kooijman, 2004).

Previous attempts to define criteria for reimbursement have made some persons skeptical about the desirability of explicitness because explicit judgments are often surrounded by uncertainty, liable to manipulation, and not sufficiently responsive to change. These authors advocate implicit rationing, being left to physicians and patients at the microlevel (e.g. Mechanic, 1997). This contribution may actually add to their skepticism, because one of the proposed ideas (relating user charges to relative need) is already applied in several countries. It may therefore seem that explicit approaches for priority setting are more susceptible to failure than implicit approaches. However, no country is completely satisfied with existing methods of priority setting, finance, and delivery of care, and everywhere there is a search for new policy instruments. It seems to be a general problem that normative beliefs underlying certain policy decisions are difficult to discern, or, the other way around, that it is difficult to translate normative beliefs into policy. This contribution may improve understanding of the relationship between policy measures and values. This could be helpful to policy makers in different countries, no matter what means of priority setting they currently apply. After all, the characteristics of defensible decision making apply regardless of differences in the funding and provision of health care (Ham et al., 2003).

difficult

Part of Damien Hirst's Hello love

Chapter 8
Discussion: from opposition to integration of ethics and economics

The central question in this thesis has been whether the impact of economic evaluations in priority setting could be improved by the integration of equity concerns in economic models. In addressing this main question three sub-questions were examined:

- *How is cost-effectiveness evidence used in health care priority setting, and how are equity concerns taken into account? (8.1)*
- *Is it possible to determine what equity concepts are the most relevant ones and measure the weights for the development of an equity adjustment procedure? (8.2, 8.3 and 8.4)*
- *What is the potential of equity adjustment to regulate costs in health care expenditure? (8.5)*

In the light of these questions the last chapter discusses the findings, reflects on their implications, and provides suggestions for further research.

Chapter 8

Discussion: from opposition to integration of ethics and economics

8.1 Economic evaluations in priority setting

Chapter 1 of this thesis pleads for the integration of normative and economic concerns in economic evaluations of health care. Before addressing these issues, chapters 2 and 3 collected evidence from recent policy decisions that provided support for this approach. These chapters showed that the interpretation of economic evaluations is far from simple. A straightforward interpretation of economic evaluations would be to give priority to patients who derive health benefits at the lowest costs. Cost-effectiveness evidence, however, does not seem to influence decisions in such a simple way. The decision not to fund the cost-effective medicine sildenafil (chapter 2) suggested that reimbursement decisions depended on a wider set of objectives than just maximizing health gain from the available budget. One of these other concerns related to the severity of a condition, reflecting that equity concerns were also relevant in the allocation of health care resources (chapter 3). Unfortunately, evidence about the contribution of different health care technologies towards an equitable distribution of health is not systematically collected. In health care decisions, these equity concerns therefore usually remain implicit. A consequence is a reduction in the explicitness and transparency of the decision-making process. In order to promote the view that both equity concerns and economic concerns should be taken into account in an appropriate way, it seems worthwhile to pursue integration of equity concerns into economic models (chapter 3).

This conclusion ties in with observations in the literature about the influence of economic evaluations on reimbursement decisions. In Australia and the UK, a large grey area has been observed where no clear relationship exists between the cost-effectiveness of a health care intervention and decisions about reimbursement. A more favourable cost-effectiveness ratio increases the chance of funding, but is not decisive (George et al., 2001; Devlin and Parkin, 2004). This seems also true for the Netherlands. Economic evaluations have been performed in the Netherlands since the 1980s. Initially these evaluations were performed on a small scale. The Dutch Healthcare Insurance Board initiated and funded economic evaluations as a response to high-tech, high-cost health technologies such as heart and lung transplants. These initiatives evolved into the establishment of the Fund for Investigative Medicine that continued to fund economic evaluations of selected emerging technologies until the last decade (Berg et al., 2004). In this period, some funding decisions were contrary to the outcomes of the economic evaluation (e.g. lung transplant). In recent years it became clear that deviations from economic recommendations would not be limited to isolated cases. In 2005, a policy was introduced that required an economic dossier for new drugs when a premium price was requested. Some pharmaceutical companies had included economic dossiers on a voluntary basis since 2002, when they filed for reimbursement. An analysis of the reimbursement decisions in relation to these new drugs showed that sometimes interventions with unfavourable cost-effectiveness received funding, and vice versa, i.e. interventions with favourable cost-effectiveness were denied funding. Analysis of the underlying concerns in the decision-making process suggested that severity of a disease was a major, but implicit factor in reimbursement decisions (Pronk and Bonsel, 2004).In an ethical-empirical investigation of reimbursement decisions, Hoedemakers (2003) reached the same conclusion. He suggested that reimbursement decisions were affected by feelings of solidarity that went beyond income and risk solidarity specified in policy models to guarantee equal access. Feelings of solidarity were also implicit in reimbursement decisions that gave priority to those patients who were in the worst health state.

8.2 Towards equity adjustment

The first part of this thesis justified the pursuit of an equity adjustment procedure in economic evaluations. Contributing to the development of an equity-adjustment procedure was the main goal of the second part of this thesis. For this purpose we explored what equity concepts (chapter 5) were the most relevant and we aimed to measure the accompanying weights (chapter 6). First, however, chapter 4 illustrated how equity weights –based on the equity concept of proportional shortfall- can be utilized to adjust cost-effectiveness ratios. Real life cases were used to develop the equity-adjustment procedure. This was carried out to confirm the feasibility of the procedure, e.g. in terms of data collection. Moreover, we considered it relevant to enable people to scrutinize the desirability of the equity adjustment procedure in economic evaluations. We also used real life cases because we expected that this would make it easier for people to foresee the implications and to compare these with their own intuitions on how priorities ought to be set. In both respects the attractiveness of the procedure was confirmed: the equity adjustment procedure appeared to be feasible and the outcomes in priority setting seemed closer to intuition than priorities that were based solely on cost-effectiveness data. What remains is the challenge to collect empirical data to refine this policy model, so that it represents a good approximation of societal preferences for priority setting. Two aspects of the model need specific consideration: first it is unclear what notion of equity should be adopted, and second we need to establish an appropriate set of associated equity weights.

Chapter 4 offered a preliminary view on the notion of equity that should be adopted. In chapter 4 we decided to use proportional shortfall because it balances two other potentially relevant but conflicting arguments: fair innings and severity of illness. It is however uncertain whether proportional shortfall is in fact the most appropriate procedure. For instance, equity concepts that are grounded in ethical theory may be considered superior to ones that are merely based on social preferences. Proportional shortfall is not derived from a particular ethical theory, in contrast to, for example, fair innings. The proportional shortfall concept is a combination of two substantive principles proposed in the academic literature: fair innings and rule of rescue. In its approach towards reducing inequalities, proportional shortfall takes these conflicting interests into account to increase its acceptability among recipients of care. Although intuitively such a balance of conflicting principles makes sense, theory has yet to be developed to address the question of whether this combination of principles of justice has the same moral status as each principle by itself (Cookson and Dolan, 2000). A convincing argument for the use of proportional shortfall may however be made if proportional shortfall indeed reflects societal preferences. Chapter 5 explored whether this was the case. But even if proportional shortfall reflects societal preferences, people who think that society may be wrong can still dispute the concept's appropriateness. However, such a general disqualification of social preferences seems too extreme. Like Van Willigenburg (2003) we believe that it is possible to distinguish between appropriate and inappropriate feelings and that (under conditions) societal values hold moral status and are an appropriate guide to decision makers in resource allocation.

A second issue that was further explored was the trade-off between equity and efficiency. To illustrate the equity-efficiency trade-off, chapter 4 made some assumptions on how attractive a treatment ought to be in economic terms to be eligible for funding, given a certain burden of disease. The model became more persuasive if the level of differentiation in the accepted cost per QALY among treatments with different appeal to equity considerations was measured empirically. Measuring the trade-off between equity and efficiency concerns was the objective in chapter 6.

Chapters 5 and 6 responded to the need to refine the equity adjustment procedure using empirical data. For advocates of equity adjustment procedures, these chapters provide encouraging results concerning the ability to generate the required preference data. Nevertheless, the findings must be interpreted with caution. It appeared to be relatively easy to prove that the rule of health maximisation was not consistent with people's preferences. However, generating unambiguous evidence in support of an alternative allocation rule has proved to be a difficult task. The reason is that researchers who wish to carry out preference elicitation experiments have to deal with several feasibility and psychological issues at the same time. It would be wrong to assume that individual preferences are clearly formed and fixed so that the researcher merely needs to ask the respondent to describe his/her preference. Rather, preferences are (at least partially) constructed on the spot and individuals develop their preferences as a reaction to the context specific information they are provided with in the experimental context (Slovic, 1995; Tompkins, 2003). Since the research community is still in the dark about the mix of preferences and framing effects that have an impact on elicited responses in resource allocation tasks, it is inherently difficult to construct an experiment so that all relevant information is obtained, while no confounders or unanticipated interpretations of the questions are present. In the experiments that were described in chapters 5 and 6 we controlled for many potential biases, such as risk attitude, labelling, and ability to process mathematical information. Nevertheless it is in some respects difficult to get a clear view about what the expressed preferences in the experiments described in chapters 5 and 6 mean.

8.3 Defining equity

To establish what patient characteristics are considered relevant in the distribution of health, chapter 5 described how well an observed priority rank ordering of 10 diseases matched the theoretical rank orderings obtained from different health-related equity concepts, i.e. fair innings, severity of illness and proportional shortfall. This showed clearly that concerns about fair innings and proportional shortfall outweighed concerns about severity of illness. The data were inconclusive however, concerning the question of whether fair innings or proportional shortfall better reflected societal preferences. Preference heterogeneity may offer a partial explanation, but also limitations in the study design could have prevented conclusive results. In this respect the potential role of effect size as a confounder may have been relevant. The objective of this experiment was to simultaneously assess three independent relationships, i.e. between the observed rank ordering of the 10 conditions and the three theoretical rank orderings. It was difficult to design the experiment in such a way that the potential role of effect size as a confounder was eliminated in all three relationships at the same time and to the same extent (see chapter 5). To further clarify people's preferences, the experiment may be repeated with different hypotheses about effect size. Alternatively, different research designs may be used to avoid this problem altogether. For example, it may be possible to explore the preferences in a discrete choice experiment, so that a decomposed measure of equity can be obtained that shows how different assumptions about effect size or about a patient's no-treatment QALY-profile are valued.

Some brief comments about the operationalisation of the severity of illness concept are necessary to explain its poor performance. The reason is that equity weights derived from the severity of illness approach aim at valuing changes in health rather than valuing health states as such. Consequentially, the severity of illness approach tries to capture several aspects of a response to health care in one single set of numbers, i.e. initial quality of life, potential to benefit and the size of the actual health gain (Nord et al., 1999). In chapter 5, however, valuing health states was the objective, not valuing changes in health. Hence, severity of illness had to be defined differently. To apply the severity of illness approach into the experimental context of chapter 5, we decomposed this concept into its original factors. In this decomposition process we isolated the characteristic of severity of illness that distinguished it from other equity concepts: its origin in the rule of rescue. Dolan and Olsen (2001) previously used this conceptualization. The reason for isolating this characteristic was that the goal of the experiment was to detect what notions of equity reflect people's distributive preferences. This is easier when the experiment looks at each equity notion in isolation, rather than combining them. However, it also meant that our operationalisation of severity of illness was more extreme than the one promoted by Erik Nord (1999), who is the main protagonist of this concept. This raises the question of to what extent the poor performance of severity of illness approach in chapter 5 was due to the specifics of our operationalisation?

Our operationalisation of severity of illness assumed that the no-treatment QALY profile determined a patient's equity rank. In his 2005 paper, Nord took a different approach by relating the severity weights to the treatment effect. Equity weights apply to all the years in which the treatment yields different outcomes from those of the no-treatment profile, and no equity weights apply to the years that are unaffected by the treatment:

"*If a person's utilities in the next five years in case of non-intervention are expected to be 0.7, 0.7, 0.6, 0.5 and 0.0 (dead), and the utilities would be 0.8 in all 5 years (and then dead) in case of intervention, then the benefit from the intervention would be 0.1+0.1+0.2+0.3+0.8=1.5 QALYs. The severity approach implies the application of severity weights [...] to each of these annual utility gains. Similarly, if a person gets to live 5 years at a utility level of 0.8 instead of dying, then each of these five annual utility gains of 0.8 will be multiplied by the severity weight for the state 'dead'.* "(Nord, 2005)

Nord's operationalisation of severity of illness would have produced different equity rank orderings than our operationalisation in the experimental context of chapter 5, because it gave a markedly different role to two parameters of a patient's QALY profile. Since no health gain is attainable during healthy life years, the number of healthy life years does not impact positively or negatively on the severity weights as computed by Nord. In our interpretation they do matter: the more healthy life years remaining, the lower the equity weight. The two interpretations thus would have evaluated treatment for health risks differently, because usually a number of healthy life years can be expected in the no-treatment QALY profile (e.g. treatment of high blood pressure). Equity rankings would also be different for conditions that cause a significant loss of years of life (e.g. pulmonary hypertension). In our operationalisation the number of healthy life years lost was considered irrelevant. The opposite is true in Nord's operationalisation, because each lost life year counts when it can be saved with treatment.

In fact, under the assumptions that were used in chapter 5 (i.e. treatment would resolve all health problems, and no health losses occurred in the past), Nord's definition of severity of illness would have produced almost the same equity rank ordering as the dynamic fair innings approach. The main difference between Nord's severity of illness concept and dynamic fair innings is that in the fair innings approach all health loss counts, while in the severity of illness approach weight is only attributed to preventable health loss. Since the experiment described in chapter 5 assumed that no health losses occurred in the past and that all future health loss was preventable, the potential differences between the two concepts are negated in regard to these assumptions. The only difference originates from the fact that fair innings was expressed as the percentage of expected total health achieved. This also makes the number of QALYs that a person has enjoyed relevant, not simply the number of lost QALYs. Different assumptions would have to be used in studies that aim to explore the support for these two concepts.

There are therefore, some important lessons to be learned from the experiment described in chapter 5. First, in reviewing the differences between the conceptualization of severity of illness between Nord and us it seems better to name our approach the 'prospective health' approach, and to use the label 'severity of illness' for Nord's 2005 conceptualization. Second, the above discussion of the outcomes of chapter 5 shows that outcomes for priority setting are influenced by the assumptions and details in the operationalisation of the equity concepts. So far, debates about the appropriateness of different equity concepts have primarily concentrated on the underlying principles of justice. It has only been tentatively described how measures of fair innings should be produced (Williams, 2001). Similarly, there are some loose ends in Nord's description (1999) of the computation of severity weights, when severity of illness is used to evaluate health states rather than health changes. A detailed operationalisation like the one that we have provided for proportional shortfall is needed for all concepts, if we want to explore their social support.

8.4 Measuring equity weights

Assuming -for the time being- that fair innings offers a reasonable description of people's equity concerns, chapter 6 described an experiment that aimed to establish the relative weights for QALY gains of different patients that can be used to recalculate the value of QALY gains for those patients. In this experiment, weights were established using trade-off techniques. This technique makes use of the basic assumption that high levels of equality in health can compensate for low levels of efficiency (or vice versa). For this purpose, the trade-off experiment presented a series of choices between two populations that offered a different distribution of QALYs over two sections of the population.

The study design differed from other approaches presented in the literature in that it analyzed preferences for equity using a general model. This model assumed that a preference for a more equal distribution of QALYs might be the product of two conceptually different factors: a preference for equality per se and diminishing marginal utility for QALYs. A utility function over QALYs was analysed first to explore the impact of diminishing marginal utility. In this respect our study is consistent with other studies where equity weights have been explored (e.g. Lindholm, 1998; Dolan, 1998). This study, however, also controlled for a second factor that might impact on the value of a QALY gain: the distribution of health in society. The reason is that concerns about the distribution of health in society (i.e. the number of people in worse or better health states than the person for whom treatment is considered) reflect equity concerns per se, while concerns about the length of life or amount of QALYs may also reflect diminishing marginal utility which can be interpreted as an efficiency concern. To allow for separate assessment of both factors, our study assumed rank-dependence of equity weights. In this approach equity weights depended on how well off an individual is in terms of QALYs in comparison with other individuals in society. This means that equity weights depended on a person's rank, and not on the absolute differences in health outcomes between groups other than through their rank ordering.

In the application of the rank-dependent utility model our approach differed from previously published studies into the equity-efficiency trade-off. A discussion of the validity of this model is therefore appropriate. From a theoretical point of view it should be noted that the assumption of rank dependence is not necessarily a restrictive one. Rank-dependent utility models can be interpreted as a generalization of expected utility theory by not only transforming outcomes to utilities, but also probabilities to decision weights (Bleichrodt and Quiggin, 1997). Since most notions of equity can be presented in a rank-ordered form, rank-dependence imposes few restrictions on moral deliberations. Moreover, since the rank-dependent utility model can include a utility function over QALYs, it can encompass many of the proposed models in the literature. This model is thus quite general and -in comparison to previous approaches- less restrictive. In this sense the choice for this model does not limit applicability of the data. Nevertheless, the model may still be too restrictive. For example, it would be interesting to test the assumption that differences in QALY totals between groups only affects utility weights and not the equity weights. Since we did not record the answers to individual questions during the data elicitation procedure, it was not possible to investigate how well observed responses agreed with predicted ones to test the adequacy of the estimated model. We can therefore only test this out if new data are generated. A previous study by Johannesson and Gerdtham (1996) however, lends some support to the hypothesis that respondents focus on inequality as such rather than the size of the inequality. They found that the marginal trade-off between a group with more QALYs and a group with fewer QALYs was not affected by the size of the difference in QALYs between the two groups. Although the robustness of this result still needs to be explored (e.g. in larger samples), it seems worthwhile to further explore the potential of rank-dependent utility models.

The results of this experiment showed that people preferred an equal distribution and that in order to attain equality they are willing to give three times as much value to the health gains of the worst-off group relative to those of the best-off group. Intuitively this seems an acceptable rate of differentiation between groups. Moreover, uncertainty surrounding the validity of the findings was minimized. Considerable efforts were made to prevent the technical and psychological issues mentioned above from influencing the results of this experiment. For example, we carefully designed the questionnaire so that it would not direct people's responses in a particular direction. To achieve this, we applied the 'ping-pong approach' to focus on the preferences regarding the trade-off. A large number of random 'filler' questions were included to prevent subjects from recognizing this questionnaire structure and adapting their response mode to complete the task earlier. This strategy increased the number of questions and consequently there was an increased chance that cognitive limitations would impair the validity of the results (e.g. boredom might limit the reliability of the responses). This problem was anticipated however and could be dealt with in the experimental setting. The experiment could only reach a conclusion on the basis of a series of consistent responses, and not if people gave random responses. Therefore, the time people needed to finish different sections of the experiment helped us to identify respondents who provided many inconsistent responses during the interviews. These individuals were assisted more frequently during the remainder of the experiment. Although we cannot exclude the possibility that cognitive impairments have affected the results, there was no apparent evidence that cognitive limitations were a significant problem.

In spite of all our efforts to enhance the quality of the data, one of the findings was quite hard to interpret. In general people were inequality-averse as expected, but this preference was not found when the size of the better-off group was small. A possible explanation for this finding is that the observed equity weights were biased by what we defined as "insensitivity to group size". This is a well-documented psychological bias, described as the tendency for people to overestimate small proportions. Our data were corrected accordingly. However, we cannot say for sure that this assumption is correct and that the elicited preferences were not true preferences. Robustness of the results also needs further exploration with regard to the utility function over QALYs. Given the results of other explorations into the equity-efficiency trade-off (e.g. Lindholm, 1998; Busschbach, 1993), the finding of a linear utility function over QALYs was also somewhat surprising. It may be the case that exploration of people's preferences over a larger part of the QALY scale is needed to detect the impact of diminishing marginal utility. With regard to these uncertainties it seems too early to apply the generated equity weights in health policy.

Moreover, in policy models other characteristics of patients that may affect the trade-off between equity and efficiency may need to be considered.
Overall the conclusion is that the specification of an equity-weighting rule is an attainable objective in the long term For the present, however, caution is advocated in applying the results of equity-weighting research to policy.

8.5 Application to policy

The final question this thesis set out to answer was to what extent equity adjustment procedures may help to regulate costs in health care, in particular in the pharmaceutical sector where cost-effectiveness information is systematically used. For this purpose Chapter 7 explored different ways to employ equity adjustment procedures and reflected on their impact. In doing so, chapter 7 touched upon several issues. First it discussed the often-mentioned hypothesis that governments try to avoid making unpopular choices explicit and are therefore better off with implicit approaches. In other words: governments (or the voters who elect them) may not want to make decisions explicit and therefore there is no need for sophisticated equity adjustment procedures. Next, chapter 7 explored the potential impact of equity adjusted economic evaluations on cost containment and discussed the generalisability of the proposed model for a trade-off between equity and efficiency to different areas of health care policy. This is relevant since cost-effectiveness information is only systematically utilized in a few decision areas. Therefore this last chapter goes beyond the scope of economic models and explores how other strategies for priority setting in health care can be affected by the notion of equity adjustment.
If the hypothesis is true that implicit approaches to priority setting are favoured over explicit ones, it is not likely that even the most sophisticated models will be able to improve the rational basis underlying policy decisions. Chapter 7, however, suggested that implicit approaches like central budgeting and provision control cannot be sustainable in the long run, because they lead to problems with the quality of care. Moreover, the analyses showed that there are reasons to be optimistic about explicit approaches. Explicit priority setting has been a policy objective since publication of the well-known Dunning Report (Government Committee on Choices in Health Care, 1991; Health council, 1991). Of course, the direct effect of explicit priority setting has sometimes been disappointing. But indirectly the work that has been carried out in constructing explicit approaches seems to have paved the way for later funding decisions. Funding decisions in health care do not seem to be made at random. A pattern was found in the decisions, indicating that there was a rational and consistent basis, based on equity and efficiency concerns. The main problem is that this basis has not been transparent and therefore was allowed to frustrate decision-making. This finding suggests that further refinement of decisions support models is likely to improve their use.

Chapter 7 continued by showing that the theoretical framework that grounds the approach of equity weighting can be translated into different policy tools, and can affect different areas of health care policy. In doing so chapter 7 showed that characteristics of defensible decision-making apply regardless of differences in the funding and provision of health care. The improved understanding of people's equity concerns obtained from chapter 5 therefore is relevant in its own right, and not only for developing an equity adjustment procedure for the outcomes of economic evaluations comparable to the one that has been developed in chapter 6. Chapter 7 illustrated that the underlying notions of solidarity in terms of relative needs could also explain why some countries vary user charges according to medical need. The fact that society accepts costs to strive for equality in health is a relevant conclusion from this thesis, irrespective of what package of policy interventions is applied to tackle inequalities.

8.6 Conclusions

Improving the fairness of priority setting is possible through equity adjustment in economic evaluations. However, it is still unclear which equity concept gives the best approximation of social values. In this thesis several definitions of equity have been explored, as well as the relative weights attributed to health gains (QALY gains) of different people according to these concerns. Restrictions in the scope and design of the studies described in this thesis imply that a toolkit for policymakers with respect to equity weighting cannot yet be presented. Nevertheless, the general principles described in this thesis contribute to a better understanding and interpretation of cost-effectiveness, thus further facilitating the application of cost-effectiveness data in health policy. In answering certain research and policy questions, this thesis has also drawn attention to some matters that still need to be resolved. There are challenges in the years ahead with regard to measuring trade-offs between different equity objectives, as well as the prevention of biases and framing effects in the measurement of equity weights.

From Damien Hirst's The Pharmaceutical windows

Ref.

References

Adellaoui M, Barrios C, Wakker PP. Reconciling Introspective Utility with Revealed Preference: experimental Arguments Based on Prospect Theory. ENSAM/ESTP, Cachan, 2002

Abellan-Perpiñan J-M, Pinto-Prades JL. Health States after treatment; a reason for discrimination? Health Economics 1999;8:701-707

Al MJ, Feenstra T, Brouwer WBF. Decision makers' views on health care objectives and budget constraints: results from a pilot study. Health Policy 2004;70:33–48

Althof S, Turner L, Levine S, Risen C, Kursch E, Bodner D, Resnick M. Why do so many people drop out from auto-injection therapy for impotence? Journal of Sex and Marital Therapy 1989;15(2):121-9

Andersen R, Smedby B, Vagero D. Cost containment, solidarity and cautious experimentation: Swedish dilemmas. Social Science Medicine 2001;52(8):1195-204

Annemans L, Crott R, De Clerck H, Huybrechts M, Peys F, Robays H, Steens I, Vanschoubroek K, Winderickx P. Pricing and reimbursement of pharmaceuticals in Belgium. PharmacoEconomics 1997;11(3):203-9

Anonymous. Diagnose en behandeling van erectiele disfunctie (Diagnosis and treatment of Erectile Dysfunction: step by step [in Dutch]). Round table conference, Rotterdam, Medicom Excel, 1998

Anonymous. De Graaf: de pil en IVF horen niet in het ziekenfonds (De Graaf: Contraception and IVF do not belong in the health care benefit package [in Dutch]). Volkskrant, Amsterdam, 25-02-2005

Atkinson, AB. On the measurement of inequality. Journal of Economic Theory 1970;2:244-63

Bleichrodt, H. Health utility indices and equity considerations. Journal of Health Economics 1997;16(1):65-91

Bleichrodt H, Diecidue E, Quiggin J. Equity weights in the allocation of health care: the rank dependent QALY model. Journal of Health Economics 2004;23:157-171

Bleichrodt H, Doctor J, Stolk EA. A Nonparametric Elicitation of the Equity-Efficiency Tradeoff in Cost-Utility Analysis. Journal of Health Economics (in press)

Boer A. Het basispakket: inhoud en grenzen. Rapport naar aanleiding van de evaluatie van "10 jaar pakketdiscussie" (The social benefit package: coverage and boundaries. Report on the basis of 10 year reimbursement debates [in Dutch]). College voor Zorgverzekeringen, Amstelveen, 2001

Bonsel GJ, Jansen MF, Birnie E. Mild Diseases & Ailments Study. Ziektelast bepaling van 27 vermoedelijk 'lichte' ziekten ten behoeve van rationele beperking van het geneesmiddelenpakket: samenvatting (Mild Diseases & Ailments Study. Establishing proportional shortfall of 27 allegedly 'mild' diseases for rational decisions about inclusion of new medicines in the public health care package, in Dutch). Amsterdam, Afdeling Sociale Geneeskunde, AMC, 2003

Borgmann A. The moral assessment of technology. Democracy in a Technological Society. L. Winner. The Netherlands, Kluwer, Academic Publishers, 1992

Bostic R, Herrnstein RJ, Luce RD. The effect on the preference-reversal phenomenon of using choice indifferences. Journal of Economic Behavior and Organization 1990;13:193-212

Brazier J, Dixon S. The use of condition specific outcome measures in economical appraisal. Health Economics 1995;4(4):255-264

Brooks A. Viagra is licensed in Europe but rationed in Britain. British Medical Journal 1998;317:765

Brooks R. EuroQol: the current state of play. Health Policy 1996;37:53-72

Brouwer WBF, Van Hout BA. How should different life expectancies be valued? Diminishing marginal utility and discounting future effects have similar consequences. British Medical Journal 1998;317:1155

Busschbach JJV, Hessing DJ, De Charro FT. The utility of health at different stages in life; a quantitative approach. Social Science and Medicine 1993;37(2):153-158

Callahan D. Setting mental-health priorities - problems and possibilities. The Milbank Quarterly 1994;72:451-70

CBO. Behandeling en preventie van coronaire hartziekten door verlaging van de plasmacholesterolconcentratie. Consensus Cholesterol tweede herziening (Treatment and prevention of coronary heart disease using cholesterol lowering therapy. Consensus Cholesterol. Second revision [in Dutch]). Utrecht, Dutch Institute for Healthcare CBO, 1998

CBO. Osteoporose. Tweede Herziening Richtlijn (Osteoporosis. Second revision clinical guideline [in Dutch]). Utrecht, Dutch Institute for Healtcare CBO, 2002

CBS. Statline. 2003;

Clark J. Fairness in public good provision: an investigation of preferences for equality and proportionality. Canadian Journal of Economics 1998;31(3):708-29

Cohen CB. Quality of life and the analogy with the Nazis. Journal of Medicine and Philosophy 1983;8:113-35

Commission on the Future of Health Care in Canada. Medically Necessary: What is it, and who decides? Health Care Commission, Ottowa (Canada), 2002

Cookson R, Dolan P. Public views on health care rationing: a group discussion study. Health Policy 1999;49:63-74

Cookson R, Dolan P. Principles of justice in health care rationing. Journal of medical Ethics 2000;26:323-9

Cuadras-Morató X, Pinto-Prades J-L, Abellán-Perpiñan J-M. Equity considerations in health care: the relevance of claims. Health Economics 2001;10:187-205

Culyer AJ, Wagstaff A. Equity and equality in health and health care. Journal of Health Economics 1993;12:431-57

Devlin N, Parkin D. Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. Health Economics 2004;13:437-52

Dinsmore W, Evans C. ABC of Sexual Health: Erectile dysfunction. British Medical Journal 1999;318:387-390

Dolan P. The measurement of individual utility and social welfare. Journal of Health Economics 1998;17:39-52

Dolan P, Olsen JA. Equity in health: the importance of different health streams. Journal of Health Economics 2001;20:823-834

Dolan P, Olsen JA. Distributing health care. Oxford, Oxford University Press, 2002

Dolan P, Shaw R, Tsuchiya A, Williams A. QALY maximisation and people's preferences: a methodological review of the literature. Health Economics 2005;(14)1:197-208

Drummond MF, O'Brien B, Stoddart Gl, Torrance GW. Methods for the economic evaluation of health care programmes. Oxford, Oxford University Press, 1997

Ebert U. Measurement of inequality: an attempt at unification and generalization. Social choice and welfare 1988;5:147-69

Edwards AL. Techniques of attitude scale construction. New York, Appleton-century-crofts, inc, 1957

Foundation of Pharmaceutical Statistics (SFK). Farmacie in cijfers. De potentie van Viagra in Nederland (Pharmaceutical statistics. The potency of Viagra in The Netherlands [in Dutch]). Pharmaceutisch Weekblad 1999;134(6)

Furlong W, Feeny D, Torrance GW, Goldsmith CH, DePauw S, Zhu Z, Denton M, Boyle M. Multiplicative multiattribute utility function for the Health Utilities Index Mark 3 (HUI 3) system: A technical report. McMaster University Centre for Health Economics and Policy Analysis Working Paper 98-11, Hamilton, Ontario, 1998

George, B, Harris A, Mitchell A. Cost-effectiveness analysis and the consistency of decision making. PharmacoEconomics 2001;19(11):1103-9

Gheorghiu S, Godschaik M, Gentili A, Mulligan T. Quality of life in patients using self-administered intracavernous injections of prostaglandin E1 for erectile dysfunction. Journal of Urology 1996;156:80-81

Gilbert D, Walley T, New B. Lifestyle medicines. British Medical Journal 2000;321:1341-4

Gold MR, Patrick DL, Torrance GW, Fryback DG, Hadorn DC, Kamlet MS, Daniels N. Identifying and valuing outcomes. Cost-effectiveness in Health and Medicine. Gold M, Siegle J, Russel L, Weinstein M. New York, Oxford University Press, 1996a: Chapter 4, p 82-134

Gold MR, Siegel JE, Russell LB, Weinstein MC. Cost-effectiveness in health and medicine. New York, Oxford University Press, 1996b

Goldman L, Gordon DJ, Rifkind BM, Hulley SB, Detsky AS, Goodman DS, Kinosian B, Weinstein MC. Cost and health implications of cholesterol lowering. Circulation 1992;85(5):1960-1968

Goldstein I, Lue TF, Padma-Nathan H, Rosen RC, Steers WD, Wicker WA, for the Sildenafil Study Group. Oral Sildenafil in the Treatment of Erectile Dysfunction. New England Journal of Medicine 1998;338(20):1397-1404

Goldstein WM, Einhorn HJ. Expression Theory and the Preference Reversal Phenomena. Psychological Review 1987;94:236-54

Gonzalez R, Wu G. On the Shape of the Probability Weighting Function. Cognitive Psychology 1999;38:129-66

Government Committee on Choices in Health Care. Choices in Health Care. Ministry of Welfare, Health and Cultural Affairs, Rijswijk, 1992

Green C, Brazier J, Deverill M. Valuing health-related quality of life. A review of health state valuation techniques. PharmacoEconomics 2000;17(2):151-165

Hadorn DC. The role of public values in setting health care priorities. Social Science and Medicine 1991;32(7):773

Ham C, Robert G, Eds. Reasonable rationing. International experience of priority setting in health care. State of Health Series. Maidenhead, England, Open University Press, 2003

Harris J. QALYfying the value of life. Journal of medical Ethics 1987;13:117-123

Hayes S, Green A, Anonymous. Rationing of sildenafil. Letters to the editor. British Medical Journal 1999;318:1620

Hurley J, Ed. An overview of the normative economics of the health sector. Handbook of Health Economics. Amsterdam, Elsevier Science B.V., 2000

Jackson SE, Lue TF. Erectile dysfunction: Therapy health outcomes. Urology 1998;51(6):874-882

Jarow JP, Burnett AL, Geringer AM. Clinical efficacy of sildenafil citrate based on etiology and response to prior treatment. Journal of Urology 1999;162:722

Johannesson M. Should we aggregate relative or absolute changes in QALYs? Health Economics 2001;10:573–577

Jønler M, Moon T, Brannan W, Stone NN, Heisey D, Bruskewitz RC. The effect of age, ethnicity and geographical location on impotence and quality of life. British Journal of Urology 1995;75:651-655

Kind P. Applying paired comparisons models to EQ-5D valuations - deriving TTO utilities from ordinal preference data. EuroQol Plenary meeting, Oslo, 1996

Kooijman H. Zorg verzekered (Health care insured [in Dutch]). Medisch Contact 2004;59(3):63

Korenman SG. New insights into erectile dysfunction: a practical approach. American Journal of Medicine 1998;105:135-144

Kwok YS, Kim C. Valuing Viagra: what is restoring potency worth? Effective Clinical Practice 1999;2(4):171-5

Laupacis A, Feeny D, Detsky AS, Tugwell PX. How attractive does a new technology have to be to warrant adoption and utilization? Tentative guidelines for using clinical and economic evaluations. Canadian Medical Association Journal 1992;146(4):473-481

Lindholm L, Rosén M, Emmelin M. How many lives is equity worth? A proposal for equity adjusted years of life saved. Journal of Epidemiology and Community Health 1998;52:808-811

Litwin MS, Nied RJ, Dhanani N. Health-related quality of life in men with erectile dysfunction. Journal of General Internal Medicine 1998;13:159-166

Louviere JJ, Hensher DA, Swait JD. Stated choice methods. Analysis and application. Cambridge, cambridge University Press, 2000

Luce RD. Utility of gains and losses: Measurement-theoretical and experimental approaches. Mahwah, New Jersey, Lawrence Erlbaum Associates, Inc., 2000

Lycklama à Nijeholt AAB. De erectiepil (The erection pill [in Dutch]). Nederlands Tijdschrift voor Geneeskunde 1998;142(43):2332-3

Maynard AK. Developing the health care market. Economic Journal 1991;101:1277-1286

Mechanic D. Muddling Through Elegantly: Finding The Proper Balance In Rationing. Health Affairs 1997;16(5):83-92

Meuleman EJH, Van Berkel JTH, Rabsztyn P, Damen L. Sildenafil (Viagra) voor de behandeling van erectiestoornissen (Sildenafil (Viagra) in the treatment of erectile dysfunction [in Dutch]). Nederlands Tijdschrift voor Geneeskunde 1998;142(43):2337-41

Michel BC, Bonsel GJ, Stouthard MEA, McDonnell J, Habbema JDF. Kosten-effectiviteitsanalyse van levertransplantatie op lange termijn; het Groningse levertransplantatieprogramma 1979-1991 (Cost-effectiveness analysis of long-term liver transplantation; the liver transplantation program of Groningen 1979-1991 [in Dutch]). Nederlands Tijdschrift voor Geneeskunde 1993;137(19):963-9

Ministry of Health Welfare and Sports. Medicines policy in The Netherlands. Status as of 1 January 2004. In: International Publication Series Health, Welfare and Sports, no. 15. The Hague, 2002

Ministry of Health Welfare and Sports. Health Insurance in The Netherlands. Status as of 1 January 2004. In: International Publication Series Health, Welfare and Sports, no. 1E. The Hague, 2004

Mulhall JP, Jahoda AE, Cairney M, Goldstein B, Leitzes R, Woods J, Payton T, Krane RJ, Goldstein I. The causes of patient dropout from penile self-injection therapy for impotence. Journal of Urology 1999;162:1291-1294

NIH Consensus Conference. NIH Consensus Development Panel on Impotence. Journal of the American Medical Association 1993;270:83-90

Nord E. The trade-off between severity of illness and treatment effect in cost-value analysis of health care. Health Policy 1993;24:227-38

Nord E. The Person-trade-off approach to valuing health care programs. Medical Decision Making 1995;15:201-208

Nord E. Cost-value analysis in health care: making sense out of QALYs. Cambridge, Cambridge University Press, 1999

Nord E. Concerns for the worse off: fair innings versus severity. Social Science and Medicine 2005;60:257–263

Nord E, Pinto JL, Richardson J, Menzel P, Ubel P. Incorporating societal concerns for fairness in numerical valuations of health programmes. Health Economics 1999;8:25-39

Nord E, Street A, Richardson J, Kuhse H, Singer P. The significance of age and duration of effect in social evaluation of health care. Health Care Analysis 1996;4:103-11

Norheim OF. Healthcare rationing—are additional criteria needed for assessing evidence based clinical practice guidelines? British Medical Journal 1999;319:1426-9

Pelen F. Reimbursement and pricing of drugs in France. Health Economics in Prevention and Care (HEPAC; On Tuesday, January 01, 2002 this publication was renamed The European Journal of Health Economics) 2000;1(1):20-23

Poley MJ, Stolk EA, Brouwer WB, van Busschbach JJ. Ziektelast als uitwerking van het criterium 'noodzakelijkheid' bij het maken van keuzen in de zorg ('Necessity' determined on the basis of disease severity when prioritising health care interventions [in Dutch]). Nederlands Tijdschrift voor Geneeskunde 2002; 146: 2312-2315.

Pronk MH, Bonsel GJ. Out-patient drug policy by clinical assessment rather than financial constraints? Euorpean Journal of Health Economics 2004;5(3):274-277

Ramsey-Baggs P, Gaskell P. Rationing. British Medical Journal 1998;317:1527

Robinson R. Limits to rationality: economics, economists and priority setting. Health Policy 1999;49:13-26

Roemer JE. Theories of distributive justice. Cambridge, Maasachusetts, Harvard University Press, 1996

Rosen RC, Riley A, Wagner G, Osterloh IA, Kirkpatrick J, Mishra A. The International Index of erectile function (IIEF): a multidimensional scale for assessment of erectile dysfunction. Urology 1997;49:822-830

Rutten FFH. Interpretation and use of cost-effectiveness analysis. From cost to effects: A manual for evaluation studies in health care. Rutten-van Mölken MPMH, Busschbach JJV, Rutten FFH, 2000

Rutten FFH, Busschbach JJV. How to define a basis package of health services for a tax funded or social based health care system? European Journal of Health Economics 2001;2:45-46

Saha A. Expo-power utility: a ‘flexible’ form for absolute and relative risk aversion. American Journal of Agricultural Economics 1993;75:905-13

Sassi F, Archard L, Le Grand J. Equity and the economic evaluation of health care. Health Technology Assessment 2001;5(3):10

Scitovsky T. The joyless economy. An inquiry into human satisfaction and consumer dissatisfaction. New York, Oxford University Press, 1977

Sen AK. Inequality reexamined. Oxford, Oxford University Press, 1992

Sen AK. Why health equity? Health Economics 2002;11:659-666

Slovic P. The construction of preference. American Psychologist 1995; 50(5):364-371

Smith KJ, Roberts MS. The cost-effectiveness of sildenafil. Annals of Internal Medicine 2000;132(12):933-937

Stolk EA, Brouwer WBF, Busschbach JJV. Rationalising rationing: economic and other considerations in the debate about funding of Viagra. Health Policy 2002;59(1):53-63

Stolk EA, Busschbach JJV. Are patients and the general public like-minded about the effect of erectile dysfunction on quality of life? Urology 2003;61(4):810-815

Stolk EA, Busschbach JJV, Caffa M, Meuleman EJH, Rutten FFH. The cost-utility of Viagra® in the Netherlands. Institute for Medical Technology Assessment, Rotterdam, 1999

Stolk EA, Busschbach JJV, Caffa M, Meuleman EJH, Rutten FFH. Cost utility of sildenafil compared to papaverine-phentolamine injections. British Medical Journal 2000a;320:1165

Stolk E, Busschbach J. The obvious cost-effectiveness of sildenafil calls for a political discussion, not a technical one. British Medical Journal 2000b;321:510

Stolk EA, Pickee SJ, Ament AHJA, Busschbach JJV. Equity in health care prioritization; an empirical inquiry into social value. Health Policy (in press)

Stolk EA, Poley MJ, Brouwer WBF, Busschbach JJV. Proeftoetsing van het iMTA-model. Identificatie van aandoeningen met minimale ziektelast en proeftoetsing van de voor ziektelast gecorrigeerde doelmatigheidstoets (Pilot of the iMTA-model. Identification of the least severe conditions and pilot of a severity-adjustment procedure in economic evaluations). In: Toenders WGM. Vervolgonderzoek breedte geneesmiddelenpakket (Follow-up study into the definition of the medicines benefit package [in Dutch]). Amstelveen, CVZ, 2002

Stolk EA, Van Donselaar G, Brouwer WBF, Busschbach JJV. Reconciliation of economic concerns and health policy: illustration of an equity adjustment procedure using proportional shortfall. PharmacoEconomics 2004;22(17):1097-1107

Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use. Oxford, Oxford University Press, 1995

Toenders WGM. Breedte geneesmiddelenpakket (Definition of the medicines benefit package [in Dutch]). College voor Zorgverzekeringen, Amstelveen, 2001

Toenders WGM. Vervolgonderzoek Breedte Geneesmiddelenpakket (Follow-up study into the definition of the medicines benefit package [in Dutch]). College voor Zorgverzekeringen, Amstelveen, 2002

Tsuchiya A. Age-related preferences and age weighting health benefits. Social Science Medicine 1999;48:267-276

Tsuchiya A. QALYs and ageism: philosophical theories and age weighting. Health Economics 2000;9:57-68

Tversky A, Kahneman D. Advances in prospect theory: cumulative representation of uncertainty. Journal of Risk and Uncertainty 1992;5:297-323

Ubel PA, Baron J, Asch DA. Preference for equity as a framing effect. Medical Decision Making 2001;21:180-9

Van de Vathorst S. Justice, solidarity & responsibility in Dutch health care. Your money or my life. Department of Philosophy. Amsterdam, Vrije Universiteit.2001;

Van Willigenburg T. De status en rol van sterke gevoelens in morele oordeelsvorming (The status and role of gut feelings in moral judgement [in Dutch]). Algemeen Nederlands Tijdschrift voor Wijsbegeerte 2003;95(2):81-99

Volk RJ, Cantor SB, Spann SJ, Cass AR, Cardenas MP, Warren MM. Preferences of husbands and wives for prostate cancer screening. Archives of Family Medicine 1996;6:72-6

Wagstaff A. QALYs and the equity-efficiency trade-off. Journal of Health Economics 1991;10:21-41

Wagstaff A. Erratum. Journal of Health Economics 1993;12:237

Wakker PP, Deneffe D. Eliciting von Neumann-Morgenstern utilities when probabilities are distorted or unknown. Management Science 1996;42:1131-50

Waugh N, Scott D. How should different life expectancies be valued. British Medical Journal 1998;316:1316

Wetenschappelijke Raad voor het Regeringsbeleid. Volksgezondheidszorg (Public health care [in Dutch]). Den Haag, Sdu Uitgevers, 1997

Wieringa NF, Van Dieten HEM, Eijgelshoven MHJ. Ziektelast toegepast op het geneesmiddelenpakket. College voor Zorgverzekeringen, Diemen, 2003

Wilke RJ, Glick HA, McCarron TJ, Erdre MH, Althof SE, Linet AO. Quality of life effects of alprostadil therapy for erectile dysfunction. Journal of Urology 1997;157:2124-2128

Williams A. Economics, QALYs and Medical Ethics. A health Economist's Perspective. Centre for Health Economics, York health economics consortium, NHS for reviews and dissemination, York, 1994

Williams A. Intergenerational equity: an exploration of the 'fair innings' argument. Health Economics 1997;6(2):117-32

Williams A. The 'fair innings argument' deserves a fairer hearing! Comments by Alan Williams on Nord and Johannesson. Health Economics 2001;10(7):583-5

Williams A, Cookson C. Equity in health. Handbook of Health Economics. Culyer AJ, Newhouse JP. Amsterdam, Elsevier Science B.V., 2000. 1B: 1863-1910

World Health Organisation. How well do health systems perform? World Health Report 2000. Health Systems: Improving Performance, 2000

Sum.

Summary

Because of the increasing development of medical innovations and limited budgets to finance care, governments face a challenge to determine what services should be covered for their populations. Economic evaluation of health care interventions is viewed as a useful way of informing priority setting in health care. Increasingly, economic evaluations that can demonstrate favourable cost-effectiveness are required for new health care technologies, especially pharmaceutical products, and pharmaceutical companies must comply with these requirements if they wish their products to be included in basic benefits packages.

Despite the increasing number of economic evaluations in the field of health care, their impact on policy decisions has not been impressive. There is no clear relationship between the outcomes of a cost-effectiveness analysis and a reimbursement decision. To help improve consistency in priority setting, this thesis analysed the discrepancy between resource allocation decisions and economic appraisals. The central assumption is that the discrepancies between economic evaluation and health policy can be described and explained on the basis of concerns about fairness.

Economic evaluations are closely linked to the view that society desires to maximize health outcomes for the population within a given budget. In cost-effectiveness research, standardized methodology helps to compare the costs and effects of different interventions in order to identify those that offer the greatest value for money. In health care resource allocation, however, economic concerns may conflict with feelings of solidarity. Many people have expressed concern over some of the implications of economic evaluations in priority setting. Often people place greater importance on equity than is reflected by cost-effectiveness analysis. Accordingly, they may be inclined to fund interventions that are not very cost-effective, especially when they concern patients in severe conditions. A logical conclusion is that society considers it fair to sacrifice some health outcomes to reduce health inequalities in society. Basing health care priorities on cost effectiveness may therefore not be possible without incorporating explicit considerations of equity into cost-effectiveness analyses. Economic models can incorporate equity and efficiency concerns through the application of an equity weighting procedure. This approach involves estimating the sacrifices to efficiency that are considered acceptable in order to achieve a fair distribution of resources. Nevertheless, attempts to operationalise this idea of a trade-off between equity and efficiency are rare - the reason being that it is not known how equity can be defined and implemented. The purpose of this thesis was to contribute to the development of an equity weighting procedure, through an exploration of equity concerns and their relationship with economic evaluations.

The economic evaluation of sildenafil (Viagra) presented in Chapters 2 and 3 illustrated the relevance of research in the equity-efficiency trade-off. The economic evaluation demonstrated that sildenafil was a cost-effective treatment for patients with erectile dysfunction, yet the health authorities seemed not to be inclined to offer reimbursement, or only inclined to partially reimburse this medicine (i.e. by reimbursing only specific patient groups). According to some evaluators, the observed discrepancy between the outcomes of the economic evaluation and the reimbursement decision resulted from the limited validity of the utility measures used in the study. It is more likely however, that cost-effectiveness information was not the only decisive factor in the decision-making process. Chapter 3 analysed the funding debate surrounding sildenafil in order to identify the other factors involved. Many of the arguments that were put forward in the reimbursement debate to explain a negative reimbursement decision reflected concerns about fairness. For instance, arguments where cost-effectiveness was favourable were countered by arguments about people's 'individual responsibility' and the low 'burden of disease' or limited necessity of treatment. Unfortunately, a clear and practical operationalisation of these arguments is lacking in the debate. This chapter therefore ends by presenting a model that can be used to balance the two main concepts (efficiency and equity) in a systematic and transparent way through the operationalisation of an equity-efficiency trade-off. This trade-off implies that the strength of people's claim on health care resources is reflected in the requirements in regard to cost-effectiveness. The more the idea of withholding a treatment for a particular patient offends concepts of fairness, the less stringent the requirements may be regarding cost-effectiveness and vice versa.

Chapters 4, 5 and 6 presented empirical studies that aimed to contribute to the operationalisation of the equity-efficiency trade-off. Typically a trade-off between equity and efficiency involves the application of so-called equity weights, which can be used to recalculate the value of health gains for different patients. Chapter 4 considered the political and practical issues involved in defining equity and calculating equity weights, before it moved on to illustrate the potential impact of a trade-off between equity and efficiency on funding decisions. A large part of this chapter concentrated on the debate about the appropriateness of three different equity concepts to inform priority decisions: 'severity of illness', 'fair innings', and 'proportional shortfall'. Both fair innings and severity-of-illness seem to receive public support, in spite of inherent conflicts in their priority ranking. It seemed therefore sensible to opt for an equity concept that combines the main features of both: proportional shortfall. Proportional shortfall assumes that measurement of inequalities in health should concentrate on the number of quality-adjusted life years (QALYs) that people lose relative to the remaining QALY expectancy they would have in absence of the disease. In order to scrutinize the fit of this equity concept with social preferences, the remainder of this chapter described the consequences of an equity adjustment procedure on the basis of proportional shortfall. The chapter considered ten real-life interventions and provided tentative conclusions about which treatments would be more or less likely to receive funding if health policy decisions were to be guided by an equity-efficiency trade-off. Compared to the situation where the only criterion is efficiency, it seems that integration of equity concerns into economic models improved the potential of economic decision support models to predict actual policy decisions.

In spite of the face validity of proportional shortfall, it is too soon to claim superiority of proportional shortfall or in fact any other equity concept in the absence of empirical evidence about the social preferences for equity. In order to establish empirical evidence about the appropriateness of different equity concepts, Chapter 5 explored social support for three equity concepts: severity of illness, fair innings and proportional shortfall. The analysis consisted of a comparison between observed priority ranks of ten conditions that were obtained from 65 respondents with three theoretical rank orderings of the same ten conditions according to the three equity concepts. Fair innings best predicted the observed rank order of the ten conditions (r=0.95), followed by proportional shortfall (r=0.82). The weakest correlation was found with severity of illness (r=0.65). All correlations were significant at a 0.05 level. Fair innings was significantly higher correlated with the observed rank ordering than the other two equity concepts. However, a preference for QALY maximization could have boosted the correlation between the observed rank ordering and fair innings, because the effect size equalled the health gap in the no-treatment QALY profile. The data remain therefore inconclusive with regard to the question about what equity concept reflects societal preferences best. In spite of associated uncertainty the data may have implications for practice. The weak support for the severity of illness approach conflicts with actual decisions in health policy, which often reflects concerns about severity of illness. This raises the question of whether health care decision makers evaluate the claims of different patients for health care by the most appropriate criteria.

The previous chapters indicated that some categories of patients have stronger moral claims on scarce health care resources than others. Chapter 6 aimed to establish the relative weights for the health gains of different patients. These relative weights were investigated by analysing peoples' choices in a trade-off experiment. This experiment consisted of a series of choices between two health care programs that resulted in a different distribution of QALYs over two sections of the population. Respondents had to indicate which distribution they would prefer. New questions were defined on the basis of the answers for the previous ones, to focus on the value where people were indifferent between the two alternatives. QALY outcomes were varied across questions as well as the proportions of people in the two sections of society, to accommodate the assumption that people's preferences for different distributions were based on the outcomes for each group as well as on the probability of receiving that outcome or anything better or worse.

This meant that the study could isolate two factors that might affect the value of a QALY gain for a particular patient: diminishing marginal utility of QALYs, and the magnitude of existing inequalities in a society. The results suggest that preferences were unaffected by diminishing marginal utility. We did however find a clear inequality aversion: people gave about three times as much value to the health gains of the worst-off group relative to those from the best-off group. In regard to this result we should add that QALY totals were used to describe and compare the health outcomes of patients, which is consistent with a definition of equity in terms of fair innings. The results have little bearing in cases where other equity concerns are adopted. In the context of this experiment fair innings was preferred to proportional shortfall, because the phrasing of trade-off questions became more complex in cases where the latter equity concept was used. In turn this implied that the cognitive task for respondents became more complex and the experiment more vulnerable to biases, framing effects, and confounding by other explanatory variables than the ones that were explicitly considered.

Finally chapter 7 turned to the question of to what extent explicit incorporation of equity considerations into the process used to develop health care policies resolved problems in resource allocation and facilitated the control of health care expenditures. In the Netherlands, the criteria 'cost-effectiveness' (efficiency) and 'necessary care' (equity) were first put forward in 1991 to guide priority decisions, but attempts to apply these criteria in health care decision-making failed. Commonly this was attributed to disagreement over their meaning. Chapter 7, however, argued that the problem was not so much in the definition of the two criteria, but rather in the application to policy. The theory about the equity-efficiency trade-off may improve the application of both criteria because it indicates how the measurement issue could be resolved, but more importantly because it explicates how conflicting criteria can co-exist in a policy environment without hampering explicit and rational decision-making. To show that the rationales apply regardless of the finance and organisational structure of health care systems, this chapter went beyond the model of the trade-off between equity and efficiency and explored a system whereby user charges varied according to medical need.

With the discussion of the application of the equity-efficiency trade-off to policy, this thesis has come full circle. Chapter 8 discussed all findings. It seems safe to conclude that the equity-efficiency trade-off offers the potential to bridge the gap between those who advocate and those who oppose the use of economic evaluations in health care decision-making on the grounds of fairness. The development of such an equity adjustment procedure in economic evaluations can ensure that in priority decisions neither equity nor efficiency concerns are put aside but instead are treated in a systematic way. Yet we have some distance to go before we can achieve sound empirical operationalisation of the equity-efficiency trade-off. Uncertainty remains despite the contribution that this thesis has made to the development of a toolbox for evaluating the appropriateness of different equity concerns and balancing them against each other and against economic concerns. The conceptual and methodological challenges presented in this thesis suggest that there is ample scope for future research.

challenges

Sv.

Samenvatting

Door de spanning tussen de groeiende vraag naar zorgvoorzieningen en de schaarse middelen voor de financiering ervan, wordt het maken van keuzen in de zorg steeds belangrijker. Economische evaluaties kunnen worden ingezet om beleidsmakers te informeren over de consequenties van dergelijke keuzen en de besluitvorming te faciliteren. In steeds meer landen moet daarom bij een aanvraag voor vergoeding een economisch dossier worden aangeboden, waaruit blijkt of de betreffende interventie wel of niet doelmatig is. In Nederland wordt zulke informatie bijvoorbeeld gebruikt in de besluitvorming over de toelating van geneesmiddelen tot het ziekenfondspakket. Ondanks de geobserveerde toename in het gebruik van economische evaluaties blijkt de impact op besluitvorming evenwel beperkt. Vaak is er geen duidelijke relatie tussen de kosten en effecten van een behandeling en het besluit om een voorziening wel of niet uit de collectieve middelen te financieren. De doelstelling van dit proefschrift is om de oorzaak van deze discrepantie te verklaren en weg te nemen. Het uitgangspunt hierbij is dat de toepassing van economische evaluaties botst met opvattingen over rechtvaardigheid.

Economische evaluaties zijn een operationalisering van het streven naar doelmatigheid in de zorg. Door gebruik te maken van gestandaardiseerde uitkomstmaten kunnen de opbrengsten van verschillende behandelingen worden vergeleken en afgezet tegen de kosten, zodat de meest doelmatige interventies geïdentificeerd kunnen worden. Soms blijkt echter dat economische overwegingen botsen met bestaande opvattingen van solidariteit. Veel mensen hebben zich zorgen gemaakt over consequenties van economische evaluaties voor de verdeling van middelen in de zorg. Mensen hechten kennelijk meer waarde aan een rechtvaardige verdeling dan wordt voorgesteld in economische evaluaties. Dienovereenkomstig bestaat er soms bereidheid om interventies die niet kosteneffectief zijn toch te vergoeden, vooral wanneer de interventie gericht is op de behandeling van mensen met een ernstige aandoening. Om discrepanties tussen vergoedingsbesluiten en economische evaluaties op te lossen, zouden rechtvaardigheidsoverwegingen geïntegreerd kunnen worden in economische modellen. Dit is mogelijk door na te gaan hoeveel waarde mensen hechten aan een betere verdeling van gezondheid. Deze waarde kan worden vastgesteld door te meten in hoeverre mensen bereid zijn iets in te leveren van de totale volksgezondheid ten behoeve van een eerlijkere verdeling. Dit idee heeft echter nog niet geleid tot aanpassing van het instrumentarium om keuzen in de zorg te onderbouwen. De reden is dat het nog weinig inzichtelijk is wat mensen precies verstaan onder een eerlijke verdeling. Op dit punt spreken mensen elkaar vaak tegen. Bovendien is alleen maar vaag omschreven welk mechanisme ingezet kan worden om een verlies aan efficiëntie te compenseren met een winst in rechtvaardigheid. Het doel van dit proefschrift is bij te dragen aan de

ontwikkeling van instrumenten voor het onderbouwen van keuzen in de zorg, middels onderzoek naar maatschappelijke opvattingen over rechtvaardigheid en doelmatigheid.
De relevantie van onderzoek naar een afruil tussen doelmatigheid en rechtvaardigheid wordt geïllustreerd met behulp van de economische evaluatie van sildenafil (Viagra) in hoofdstuk 2. Deze economische evaluatie toont aan dat sildenafil een kosteneffectief product is, maar toch lijken beleidsmakers in de zorg niet bereid dit middel te vergoeden. Bovendien wordt in landen waar sildenafil wel vergoed wordt, de vergoeding gekoppeld aan bepaalde voorwaarden. De vergoeding wordt dan bijvoorbeeld beperkt tot bepaalde patiëntengroepen. Sommige beoordelaars verklaren deze terughoudendheid in de vergoeding van sildenafil uit gebrekkige validiteit van de gehanteerde onderzoeksmethoden. Het is echter meer waarschijnlijk dat in dit geval doelmatigheid niet het enige relevante criterium was in de besluitvorming. Om na te gaan welke andere overwegingen dan doelmatigheid een rol kunnen spelen bij beslissingen omtrent vergoeding, analyseert hoofdstuk 3 het vergoedingsdebat over sildenafil. Veel argumenten in het debat blijken opvattingen over rechtvaardigheid te reflecteren. De waardering voor sildenafil bleek bijvoorbeeld nauwelijks af te hangen van de doelmatigheid van dit geneesmiddel: er werd vooral gewezen op het feit dat behandeling niet 'noodzakelijk' is, en het feit dat mensen ook een 'eigen verantwoordelijkheid' hebben. Helaas ontbreekt in de debatten een heldere en praktisch toepasbare definitie van noodzakelijkheid en eigen verantwoordelijkheid, waardoor de overwegingen moeilijk te generaliseren zijn naar beslissingen omtrent de vergoeding van andere geneesmiddelen. Dit hoofdstuk sluit daarom af met een schets van de wijze waarop doelmatigheid en rechtvaardigheid van een behandeling op consistente en transparante wijze in balans gebracht kunnen worden. Het voorstel is de beslisregel omtrent doelmatigheid afhankelijk te maken van de mate waarin een behandeling noodzakelijk wordt geacht.
In de hoofdstukken 4, 5 en 6 wordt het idee van een afruil tussen doelmatigheid en noodzakelijkheid empirisch onderzocht. Hoofdstuk 4 beschrijft het idee om de doelmatigheidseis te variëren met de noodzakelijkheid van een behandeling en illustreert de mogelijke impact op beslissingen omtrent vergoeding. Daarbij wordt tevens in kaart gebracht welke normatieve overwegingen onderdeel kunnen uitmaken van de operationalisering van het rechtvaardigheidsconcept. Hiertoe geeft dit hoofdstuk een overzicht van de meest bekende opvattingen over de rechtvaardige verdeling van middelen in de zorg, de 'severity of illness', 'fair innings' en ziektelast ('proportional shortfall'). Een weging op basis van ziektelast wordt bepleit. De reden is dat 'severity of illness' en 'fair innings' beide maatschappelijke steun krijgen, ondanks soms conflicterende uitkomsten voor prioritering. 'Severity of illness' benadrukt namelijk de acuutheid van een aandoening, 'fair innings' het veroorzaakte gezondheidsverlies. Niet altijd is het echter zo dat de meest acute aandoeningen het grootste gezondheidsverlies veroorzaken. Ziektelast

combineert elementen van beide benaderingen en neemt daardoor een tussenpositie in, door na te gaan welk percentage van zijn resterende gezondheid een patiënt dreigt te verliezen als gevolg van een aandoening. Om na te gaan of de keus voor ziektelast de juiste zou kunnen zijn, wordt onderzocht hoe differentiatie van de doelmatigheidsdrempel voor ziektelast de besluitvorming over vergoeding zou beïnvloeden. Hiertoe is voor tien interventies nagegaan hoe de doelmatigheid van de betreffende interventie zich verhoudt tot de ziektelast van de betreffende indicatie. Voor elk van deze interventies was een economische evaluatie beschikbaar. Dit alternatieve model voor vergoedingsbesluiten lijkt beter te voorspellen welke interventies wel of niet vergoed worden dan het model waarin alleen met de kosteneffectiviteit van een behandeling rekening gehouden wordt.

Omdat er nog weinig bekend is over de maatschappelijke voorkeuren voor de verdeling van middelen in de zorg, kan bevestigd noch ontkend worden dat ziektelast de maatschappelijke voorkeuren adequaat weergeeft. Hoofdstuk 5 probeert inzicht te krijgen in de geschiktheid van de verschillende rechtvaardigheidsconcepten voor gebruik in beslissingen omtrent vergoeding. Daarvoor wordt onderzocht in welke mate severity of illness, fair innings en ziektelast consistent zijn met de maatschappelijke voorkeuren. Om dit te beoordelen hebben 65 respondenten een rangorde in prioriteit gemaakt van tien indicaties. Deze geobserveerde rangorde is vergeleken met de theoretische rangordeningen van de tien indicaties, zoals die verwacht worden op basis van de drie rechtvaardigheidsconcepten. De geobserveerde rangordening kwam het meest overeen met de rangordening volgens fair innings ($r=0.95$), gevolgd door ziektelast ($r=0.82$) en severity of illness ($r=0.65$). Alle correlaties waren significant op 0.05 niveau. De correlatie van de fair innings met de geobserveerde rangordening was echter significant sterker dan de andere twee correrlaties, hetgeen suggereert dat fair innings het best de maatschappelijke voorkeuren beschrijft. Het is evenwel mogelijk dat een voorkeur voor maximalisatie van de gezondheid deze correlatie versterkt heeft, omdat de effectgrootte gelijk werd gesteld aan het dreigende gezondheidsverlies. Vanwege deze ‘confounder' is het niet zeker hoe groot het verschil tussen de correlaties werkelijk is en blijft er onzekerheid bestaan omtrent de vraag welk rechtvaardigheidsconcept het best aansluit bij de maatschappelijke voorkeuren. Desalniettemin kunnen de resultaten implicaties hebben voor beleid. In vergoedingsbesluiten lijkt namelijk severity of illness een rol te spelen, terwijl dit concept de maatschappelijke voorkeuren voor verdeling minder goed beschrijft dan de andere twee concepten. Dit roept de vraag op of beleidsmakers in de gezondheidszorg de aanspraken van verschillende patiënten op schaarse middelen in de zorg met de meest adequate criteria evalueren.

De voorgaande hoofdstukken bevestigen dat bij de verdeling van middelen in de gezondheidszorg sommige groepen patiënten een sterkere morele claim hebben op schaarse middelen dan andere groepen. In hoofdstuk 6 is geprobeerd de onderlinge verschillen te kwantificeren door het relatieve gewicht te bepalen voor gezondheidswinsten die ten goede komen aan verschillende patiënten. Voor dit doel werd een experiment uitgevoerd, dat bestond uit een serie keuzen tussen twee gezondheidszorgprogramma's die zouden resulteren in een verschillende verdeling van gezondheid tussen twee segmenten van de populatie. Respondenten gaven telkens aan welke verdeling hun voorkeur zou hebben. Deze vergelijking werd uitgevoerd op basis van het totaal aantal voor kwaliteit-gecorrigeerde levensjaren (QALYs) voor individuen in elk segment van de populaties, wat consistent is met een definitie van ongelijkheid in termen van fair innings. In de verschillende vragen werd gevarieerd met het aantal QALYs en het percentage van de bevolking dat zich bevindt in elk deel van de denkbeeldige populatie. Zodoende werd rekening gehouden met het feit dat voorkeuren voor verdeling zowel gestuurd kunnen worden door de uitkomsten voor elke groep als de kans op die uitkomst. Deze factoren werden geïsoleerd in de analyses: enerzijds is onderzocht of QALYs een afnemend marginaal nut hebben voor de productie van gezondheid, anderzijds is de invloed bepaald van de mate van ongelijkheid in de verdeling van gezondheid over de populatie op de waardering van QALYs die toevallen aan verschillende individuen. De resultaten lieten geen afnemende meerwaarde van QALYs zien. Wel werd een duidelijke aversie tegen ongelijkheid gevonden: het gewicht dat de respondenten gaven aan gezondheidswinsten voor de groep in de slechtste gezondheid was ongeveer drie keer zo hoog als het gewicht dat gegeven werd aan gezondheidswinsten voor de groep mensen in de beste gezondheid.

Tot besluit analyseert hoofdstuk 7 in welke mate de theorie over een afruil tussen doelmatigheid en noodzakelijkheid in praktijk toepasbaar is voor het beheersbaar maken van de uitgaven in de gezondheidszorg. In Nederland werd al in 1991 voorgesteld om het doelmatigheidscriterium en het noodzakelijkheidscriterium een rol te laten spelen bij vergoedingsbesluiten, maar verschillende pogingen in de volgende jaren om deze criteria toe te passen zijn mislukt. Meestal werd dit toegeschreven aan een gebrek aan overeenstemming over de invulling van de criteria. Hoofdstuk 7 van dit proefschrift betoogt evenwel dat het probleem vermoedelijk niet zozeer lag in de definitie van criteria, maar in de toepassing ervan in beleid. Het model van de afruil tussen rechtvaardigheid en doelmatigheid kan gehanteerd worden om de toepasbaarheid van beide criteria verbeteren. De reden is dat het aanduidt welke meetbare invulling gegeven kan worden aan beide criteria, maar ook hoe deze eventueel conflicterende criteria gebruikt kunnen worden in beleid zonder afbreuk te doen aan het doel van rationele en expliciete besluitvorming. Verder laat hoofdstuk 7 zien dat de rationale die ten grondslag ligt aan de afruil tussen rechtvaardigheid en doelmatigheid geldig is, ongeacht de wijze waarop het zorgstelsel gefinancierd of georganiseerd wordt. Dit wordt geïllustreerd aan de hand van een systeem waarbij differentiatie plaats vindt van eigen bijdragen voor overwegingen omtrent ziektelast.

De discussie over de implicaties van een afruil tussen rechtvaardigheid en doelmatigheid voor beleid rondt dit proefschrift af. Op basis van het gepresenteerde onderzoek lijkt de conclusie gerechtvaardigd dat de discussie tussen voor- en tegenstanders van het gebruik van economische evaluaties bij keuzen in de zorg grotendeels op te lossen valt door integratie van distributieve overwegingen in het economisch model. Het ontwikkelen van een expliciet model dat de afruil tussen de twee doelen beschrijft, garandeert dat beide overwegingen op een systematische manier gebruikt worden. Aan de andere kant laat dit proefschrift echter ook zien dat het lastig is om rechtvaardigheid zo precies te definiëren dat resultaten over de afruil tussen rechtvaardigheid en doelmatigheid eenduidig te interpreteren zijn. Kortom, het is betrekkelijk eenvoudig om het principe van QALY maximalisatie te falsificeren, maar nieuwe principes voor de verdeling van middelen zijn moeilijk te formuleren door onzekerheid in de interpretatie van de empirische resultaten. Ondanks de bijdragen in dit proefschrift blijft er onzekerheid bestaan over de vraag welke verdeling het meest adequaat is en hoe een streven naar een eerlijke verdeling afgewogen moet worden tegen een streven naar doelmatige besteding van middelen in de zorg. De conceptuele en methodologische problemen die ik in dit proefschrift ben tegen gekomen, tonen aan dat er volop mogelijkheden zijn voor aanvullend onderzoek op dit gebied.

Bedankt

Dankwoord

De woorden '*Hora est*' hebben een dubbele betekenis. Ze sluiten een mooie periode af, maar ze klinken ook als een startschot voor een -hopelijk minstens zo mooie- nieuwe periode. Dat maakt promoveren speciaal. De gebruiken rondom de verdediging van het proefschrift benadrukken dat nog eens extra. Ik geniet er dan ook van met volle teugen, en wil graag de vele mensen bedanken die op allerlei manieren hieraan een bijdrage hebben geleverd.

Jan, jou wil ik als eerste noemen. Sinds jij mij in 1998 een aanstelling gaf op het iMTA, hebben we samengewerkt. Samen hebben we de kosten-effectiviteit beoordeeld van Viagra. De media hype zorgde er voor dat dit leuk werk was, maar ook het wetenschappelijke resultaat was bijzonder. Voor passende interpretatie van de gegevens hebben we onderzoek naar de rol van ziektelast en doelmatigheid geïnitieerd. Dit onderzoek gaf veel voldoening door methodologische uitdagingen, theoretische discussies, en alle politieke belangstelling. Dat het nu ook geresulteerd heeft in dit proefschrift is net zo goed een bekroning voor jouw inspanningen als voor de mijne. Ik had me geen betere co-promotor kunnen wensen.

Frans, als promotor was je voortdurend op de achtergrond aanwezig. Als je inbreng nodig was, kwam deze altijd snel. Je was altijd bereid tot discussie en je commentaren gaven reden tot nadenken. Ik dank je voor al je kritische opmerkingen, het vertrouwen, en de vrijheid die je me gegund hebt om economische evaluaties van een ander gezichtspunt te benaderen.

Sanne, Margot en Eveline: onze vriendschap dateert van de eerste dag van onze studie gezondheidswetenschappen. Ik ben trots dat jullie mijn paranimfen wilden zijn. Ook al mogen maar twee mensen die rol officieel vervullen, voor mij zijn jullie het alledrie.

Han en Werner, ik heb het geluk gehad in mijn promotie tijd voortdurend te kunnen rekenen op veel inhoudelijke steun, waaronder die van jullie. Jullie hebben wezenlijke adviezen gegeven over de interpretatie en het gebruik van economisch theorie en onderzoeksmethoden. Ook dank ik jullie voor de tijd die je hebt vrijgemaakt om concept artikelen te lezen en van commentaar te voorzien. Theo en Gijs, jullie bedank ik voor het vervullen van dezelfde rol, maar dan vanuit de discipline ethiek en wijsbegeerte. De samenwerking met de Faculteit der Wijsbegeerte is de kwaliteit van dit onderzoek ten goede gekomen. Ik ben blij dat er voorstellen liggen om de samenwerking voort te zetten. I am also grateful for everything I learned from Alan Williams, whose comments motivated me to deepen my knowledge and look at things from different angles. His passing is a great loss to the friends he leaves behind and to the scientific community. He will be missed by anyone who was fortunate enough to work with him, but his work will continue to inspire us all.

Wil Toenders, dankzij jou en je collega's van het college voor zorgverzekeringen stond ik tijdens het onderzoek steeds in nauw contact met beleidsmakers. Het onderzoek heeft geprofiteerd van alle aandacht voor de vertaling van theorie naar praktijk. In de eerste jaren was jij bovendien telkens de eerste die de resultaten zag en kritisch becommentarieerde. Ik heb ook veel gehad aan de verhelderende discussies met je collega's Johan van Luijn, Bert Boer, Marcel Eijgelshoven en de collega's van de doelmatigheidsafdeling, waarvoor dank.

De co-auteurs, collega's en oud-collega's dank ik eveneens voor hun bijdrage. Jan O. en Isabelle, jullie waren belangrijk voor me doordat jullie je met dezelfde ambitie en hetzelfde enthousiasme als ik op promotie hebben gestort. Marten, Nicole, Els, Ivon, en Floortje, het delen van een kamer creëert een band. Jullie zijn altijd positief en bereid tot het geven van een peptalk. Dankzij jullie was de promotie een gezellige tijd.

Dames van DD en vrienden van Koko, de studietijd in Maastricht gaf de aanzet voor een wetenschappelijke carrière. Dankzij jullie heb ik me nooit verveeld, waren er veel gezellige momenten, en werd elke inspanning beloond. Dames van SDV, jullie weten wat promoveren is. Met jullie ben ik twee keer gepromoveerd, en even vaak gedegradeerd. Hoe dan ook, voetballen met jullie is heerlijke ontspanning na hard werken. Ik kijk uit naar het volgende seizoen. Pa, ma, ondanks het feit dat wetenschap de Stolken in het bloed zit, had ik niet gedacht dat ook ik zou promoveren. Als jullie dat wel vanzelfsprekend hadden gevonden, was het vast anders gelopen. Dank voor alle steun. Wilma, straks is het jouw beurt!

Maar vooral ben ik jou, lieve Pier, veel dank verschuldigd. De geestdriftige manier waarmee ik me op hobby's en werk stort, contrasteert met jouw onverstoorbare karakter. Lekker dwars op zijn tijd ben jij altijd daar om mijn ambities te relativeren en om hoofd van bijzaken te scheiden. Noodzakelijk tegenwicht. Het afronden van dit proefschrift was zonder jou veel moeilijker geweest. Al de kleine en grote dingen (zoals het verzorgen van de vormgeving van dit proefschrift) neem ik te gemakkelijk voor lief. Pier, je bent onmisbaar.

Cv

Curriculum vitae

Elly Stolk was born January 10, 1975 in Bunnik, The Netherlands. After she graduated from the Christelijk Gymnasium in Utrecht in 1993, she studied health sciences in Maastricht from 1993 to 1998. Since 1998 she has been working as a researcher at the institute for Medical Technology Assessment of the Erasmus University in Rotterdam. After conducting applied cost-effectiveness research for several years, she developed a particular interest in the measurement of individual and social preferences for informing health care decision-making. She employs empirical methods ranging from questionnaires and trade-off exercises to focus groups and interviews, and draws on theory from economics as well as other social sciences. Topics of her recent studies include equity in health, quality of life measurement, pharmaceutical reimbursement, and the consequences of Alzheimer disease.

researcher