

Linköping University Medical Dissertation No. 1215

The Art of Saying No

The Economics and Ethics of Healthcare Rationing

Gustav Tinghög

Department of Medical and Health Sciences
Linköping University, Sweden



Linköping University
FACULTY OF HEALTH SCIENCES

Linköping 2011

Edition 1:1
ISBN 978-91-7393-282-0
ISSN 0345-0082

© Gustav Tinghög, 2011
Published articles have been reprinted with the permission of the copyright holder.

Cover artwork and design: Marit Furn
Webpage: www.maritfurn.se

Printed in Sweden by LiU-Tryck, Linköping, Sweden, 2011

To the Tinghögs

I'm gonna run to you – Bryan Adams

CONTENTS

ABSTRACT	4
LIST OF PAPERS	5
ABBREVIATIONS.....	7
INTRODUCTION.....	9
Thesis Aim.....	10
A Short Note on Disposition	11
METHODOLOGICAL CONSIDERATIONS AND SPECIFICATION OF THE AIM	13
A Population-Level View on Efficiency and Fairness.....	13
Economics and Ethics	14
Normative Economics and Positive Economics	14
Normative Ethics and Meta-Ethics	16
Reasonable Disagreement.....	16
Moral Intuition and Moral Reasoning	18
BACKGROUND.....	21
What is Healthcare Rationing?	21
What is Healthcare Need?	24
Explicit Healthcare Rationing in Practice	27
The Case of Oregon	27
The Case of Sweden.....	31
THEORETICAL CONTEXT.....	37
Economics	37
Welfare Economics	37
Utilitarianism.....	39
Cost-Effectiveness Analysis and Quality Adjusted Life Years	42

Ethics.....	47
Distributive Fairness and Healthcare Rationing	48
Fair Equality-of-Opportunity and Healthcare Rationing	51
Procedural Fairness and Healthcare Rationing.....	52
 MAIN FINDINGS AND DISCUSSION – A WELFARE ECONOMIC PERSPECTIVE ON HEALTHCARE RATIONING	 55
Paper I: “Individual Responsibility for What? – A Conceptual Framework for in a Publicly Funded Healthcare System”	55
Paper II: “Discounting, Preferences, and Paternalism in Cost- Effectiveness Analysis”	57
Paper III: “Incentivizing Deceased Organ Donation: A Swedish Priority-Setting Perspective”	60
Paper IV: “Horizontal Inequality in Rationing By Waiting Lists”	62
 EXTENDED DISCUSSION – A POPULATION-LEVEL ETHICAL PERSPECTIVE ON HEALTHCARE RATIONING	 67
Individual Responsibility – A Prospective Approach.....	67
Paternalism – Why Health is not Always Good	70
Incentives – Persuasion or Coercion?	73
Inequality – When are Inequalities Unfair?	75
 FINAL REMARKS	 79
 ACKNOWLEDGEMENTS.....	 83
 REFERENCES	 85
 PAPER I - IV.....	 93

ABSTRACT

It follows from resource scarcity that some form of healthcare rationing is unavoidable. This implies that potentially beneficial medical treatments must be denied to patients to avoid unacceptable sacrifices in other areas of society. By focusing on four, core, conceptual themes – individual responsibility, paternalism, incentives, and inequality – this thesis explores the matter of finding justifiable grounds for saying no in the context of health care.

By combining the perspectives of welfare economics and population-level ethics, the author explicate and discusses conflicting moral values involved in healthcare rationing. Four papers form the foundation for this thesis. Paper I articulates the potential role of individual responsibility as a welfare-promoting, rationing tool by exploring when healthcare services exhibit characteristics that facilitate individual responsibility for private financing. Paper II explores the normative relevance of individuals' time preferences in healthcare rationing and when paternalism can be justified in the context of individuals' intertemporal health choices. Paper III examines the compatibility between incentive-based organ donation and the ethical platform for setting priorities in Sweden. Paper IV empirically investigates the existence of horizontal inequalities in using waiting lists to ration care.

From the discussion it is suggested, *inter alia*, that: I) Prospective responsibility as opposed to retrospective responsibility is a more productive notion of responsibility when discussing actual policies. However, potential positive effects need to be weighed against the increased economic inequality that it is likely to invoke. II) Although cost-effectiveness analysis provides valuable input when making rationing decisions it should not be viewed as a decision rule, since it is based on utilitarian values that constantly need to be balanced against other nonutilitarian values. III) Potentially, increased health could negatively affect individuals' well-being if it creates opportunities that they are unable to take advantage of. This needs to be taken into account before embarking on paternalistic policies to improve health – policies that often target the lower socioeconomic segment.

The author concludes that decisions on rationing cannot be computed through a simple formula. Moreover, given that rationing is bound to be associated

Abstract

with reasonable disagreements we are unlikely to ever fully resolve these disagreements. However, by explicitly stating conflicting moral values we are more likely to narrow the disagreements and achieve a healthcare system that is both fairer and more efficient.

LIST OF PAPERS

- I. Gustav Tinghög, Carl Hampus Lyttkens and Per Carlsson.
Individual Responsibility for What? – A Conceptual Framework for Exploring the Suitability of Private Financing in a Publicly Funded Health-Care System. *Health Economics Policy and Law* (2010) vol. 5: 201-223
- II. Gustav Tinghög.
Discounting, Preferences, and Paternalism in Cost-Effectiveness Analysis. *Submitted*
- III. Faisal Omar, Gustav Tinghög and Stellan Wellin.
Incentivizing Deceased Organ Donation: a Swedish Priority Setting Perspective. *Scandinavian Journal of Public Health (in press)*
- IV. Gustav Tinghög, David Andersson, Petter Tinghög, Carl Hampus Lyttkens.
Horizontal Inequality when Rationing by Waiting Lists. *Submitted*

List of papers

ABBREVIATIONS

A4R	Accountability for Reasonableness
CEA	Cost-Effectiveness Analysis
DALY	Disability Adjusted Life Years
DU	Discounted Utility
EQ-5D	EuroQol 5-Dimension
OECD	The Organisation for Economic Co-operation and Development
PTO	Person Trade Off
QALY	Quality Adjusted Life Years
SG	Standard Gamble
SF-36	The Short Form (36) Health Survey
TTO	Time Trade Off
VAS	Visual Analogue Scale

Abbreviations

INTRODUCTION

It is an unfortunate, but unavoidable, truth about the world that not all theoretically feasible enhancements to health and wellbeing can, or even should, be attempted. All healthcare systems must decide how to set limits, explicitly or implicitly, efficiently or inefficiently, fairly or unfairly. The idea of rationing health care might seem like cold-hearted policymaking to many. However, resources are not limitless. Moreover, health care is not the only worthy societal goal, but needs to be weighed against other important goals, such as education and personal security. Hence, from a societal perspective, healthcare rationing is both inevitable and highly desirable. As the subtitle suggests, this thesis approaches the inevitable but delicate matter of healthcare rationing from two nonmutually exclusive perspectives – that of the economist and that of the ethicist.

Rationing is not an uncommon phenomenon. Quite the opposite, it is an activity that most of us engage in on a daily basis. We might, for instance, prefer to have filet mignon instead of sausage for dinner, but choose the latter since our available financial resources are insufficient to pay for our preferred choice. When setting limits in health care this everyday activity suddenly becomes very uncomfortable and morally intricate. Denying care can result in severe consequences and even carry life and death implications. Consequently, saying no is something that goes against common moral intuitions of always trying to help identifiable individuals in need. How to make rationing decisions on justifiable grounds is therefore both an ethically and politically sensitive matter. Hence, rationing has become an area that most decision-makers prefer to shy away from, and they let rationing happen by default. However, ignoring the issue of rationing, or making decisions implicitly, will lead to decisions which are more likely to be dubious and unfair.

Traditionally it has been difficult to get decision-makers to explicitly acknowledge the inevitable need to ration health care. In the United States, attempts to discuss rationing have been met by forceful resistance. The

creation of infamous concepts like “death panels”¹ has made rationing a topic that cannot be discussed openly – arguably, contributing to the most inefficient and unequal healthcare system in the developed world. In other parts of the developed world, the idea of explicit rationing has also met resistance (although not as forcefully as in the US). In Sweden, an explicit “not-to-do list” that was introduced in 2003 was promptly labeled the “blacklist” and invoked a public outcry, which forced politicians to retreat and once again rely on implicit rationing (Bäckman, Lindroth et al. 2005; Bäckman, Karlsson et al. 2006). Thus, saying no is associated with much less praise than saying yes. Barrack Obama’s presidential campaign had good reasons to choose “yes we can!” as its slogan for the 2008 presidential election in the United States, although given the financial situation a more appropriate slogan might have been “no we can’t!” Society can’t give what it doesn’t have (at least not in the long run). This thesis explores the delicate matter of finding justifiable grounds for saying no in the context of health care, focusing on four core conceptual themes; *individual responsibility*, *paternalism*, *incentives*, and *inequality*.

From a more abstract perspective, rationing involves two general moral aims; to allocate resources in a way that yields the greatest possible contribution to health, and to distribute these resources as fairly as possible. These aims often come into conflict, calling for careful consideration of both the economic and ethical aspects at hand. Hence, rationing always involves a value judgment when balancing these core objectives. As with most value judgments, rationing is bound to be associated with reasonable disagreements concerning what solution is the best. The four conceptual themes upon which this thesis is built are intended to highlight some areas where reasonable disagreements are likely to occur.

Thesis Aim

The general aim of this thesis is to combine the perspectives of economics and ethics for exploring how to balance fairness and efficiency in healthcare rationing. More specifically, the thesis will apply these two perspectives to four conceptual themes – *individual responsibility*, *paternalism*, *incentives*, and *inequality* – in the context of healthcare rationing. Each of the conceptual

¹ Former Alaska Governor and US vice president candidate Sarah Palin used this concept in referring to healthcare legislation aimed at reducing healthcare spending.

themes corresponds to the main focus of a specific paper included in this thesis:

Paper I articulates the potential role of individual responsibility as a welfare-promoting, rationing tool by exploring when healthcare services exhibit characteristics that facilitate individual responsibility for private financing.

Paper II explores the normative relevance of individuals' time preferences in healthcare rationing and when paternalism can be justified in the context of individuals' intertemporal health choices.

Paper III examines the compatibility between incentive-based organ donation and the ethical platform for setting priorities in Sweden.

Paper IV empirically investigates the existence of horizontal inequalities in using waiting lists to ration care.

A Short Note on Disposition

It should be noted that the composition of this thesis varies somewhat from that of a traditional thesis. "Traditional" in this sense would imply supplementing the papers with a comprehensive summary (a so-called "frame story") that streamlines the findings of the consecutive papers. The present frame story is not a comprehensive summary of the consecutive papers. Instead, it should be viewed as a comprehensive summary of the contextual and theoretical landscape to which the consecutive papers relate.

In considering this thesis, a fundamental factor that needs to be recognized at the outset is my background as a trained economist. The four papers included in the thesis have been primarily written from and evaluated against the welfare-economic assumption that the most desirable rationing policy is the one that best promotes overall welfare in society. Although issues related to fairness and rationing are present in each of the papers, they are not as thoroughly explored as issues related to efficiency and rationing. Hence, the discussion of this frame story is divided into two chapters; one focusing on the main findings of the papers in relation to the conceptual themes, and another which extends this discussion by applying a broader ethical perspective focusing on fairness issues related to the conceptual themes.

This frame story proceeds as follows. Chapter 2 is devoted to a methodological discussion, which specifies the scope and point of departure of this thesis.

Chapter 3 is divided into two subsections. The first addresses some of the conceptual ambiguities that surround the central concepts of healthcare rationing and healthcare need. The second provides the contextual background by outlining two real-world cases of explicit healthcare rationing and the experiences related to health policy gained from these.

Chapter 4 sets out the theoretical context of welfare economics and population-level ethics. These nonmutually exclusive perspectives allow explicit exploration of issues related to efficiency and fairness, which will be discussed in the remaining chapters.

Chapters 5 and 6 focus on the four conceptual themes – *individual responsibility*, *paternalism*, *incentives* and *inequality* – which are likely to be associated with reasonable disagreements. Chapter 5 focuses on the main findings of the papers and discusses them primarily from a traditional perspective of welfare economics. Chapter 6 extends the discussion raised in the papers by applying a broader ethical perspective.

Chapter 7 presents some concluding remarks concerning health policy and the interdisciplinary approach applied in this thesis.

METHODOLOGICAL CONSIDERATIONS AND SPECIFICATION OF THE AIM

This thesis approaches the matter of healthcare rationing from two nonmutually exclusive perspectives – that of the economist and that of the ethicist. Before embarking on a methodological discussion of economics, ethics, and the morally perplexing issue of finding justifiable grounds for healthcare rationing, it is necessary to clarify the scope of this thesis.

A Population-Level View on Efficiency and Fairness

This thesis addresses the issues of fairness and efficiency in relation to healthcare rationing at the population level, i.e. at the institutional level. In health economics this is the natural scope since welfare economics stipulates that all effects, regardless of where and to whom in society they appear, should be taken into account when assessing the desirability of various options. In bioethics, however, it is less common to focus on the normative issues that arise at the population level. Instead, bioethics has a long tradition of focusing on issues that arise at the individual level, particularly involving the patient-doctor relationship, the boundaries of life (e.g. abortion and euthanasia), and how to apply medical knowledge and technologies in practice. Obviously, we should not diminish the importance of issues that revolve around the morality of individual conduct and character. The moral issues that arise at the population level are, however, equally vexing and often involve higher stakes. Daniels notes (2006:23) “the focus on exotic technologies may blind bioethics to the broader determinants of health and thus to factors that have more bearing on a larger good both domestically and globally.”

Bioethics at the population level deals with how to achieve a fair distribution of health and healthcare resources, thus relying on theories of justice and political philosophy. Where bioethics at the individual level focuses primarily

on the rights and responsibilities that arise in the interaction between individual patients and their doctors, bioethics at the population level focuses on the obligations of societies toward their citizens in general, and vice versa.²

Economics and Ethics

The point of departure for this thesis is founded on three fundamental observations about the world, i.e.:

- i. **Resources are scarce in relation to human wants.** We will always want more than we can afford.
- ii. **Resources have alternative uses.** Using resources for one thing is always done at the expense of using those resources for something else.
- iii. **Individuals have different wants and preferences.** We value commodities and the state of affairs differently.

Based on these observations, the economic perspective applied in this thesis is concerned with how to allocate scarce resources as *efficiently* as possible to best satisfy human wants. In contrast, the ethical perspective is concerned with how to allocate scarce resources as *fairly* as possible to best satisfy human wants. To reconcile these two perspectives in a practical sense, we need to explore the normative basis for fairness and efficiency in the context of allocating scarce healthcare resources.

Normative Economics and Positive Economics

When referring to an economic perspective it is useful to distinguish between positive and normative economics. Positive economics is an observational science that focuses on logistic or technical consequences stemming from different allocations without making any claim regarding the desirability of the different consequences. Normative economics, on the other hand, is concerned with analyzing the desirability of consequences that arise from competing resource allocations. Hence, positive economics approaches the question "What is?" while normative economics approaches the question "What ought to be?" (McCloskey 1998).

² For further readings concerning the scope of population-level bioethics, see e.g. Wikler and Brock (2008)

Sen (1987) has argued that modern economics can be divided into two veins with rather different origins. One he labels the *engineering approach* and the other he labels the *ethics-related tradition*. Both, he argues, relate to politics, but in different ways. The engineering approach is, as Sen characterizes it, not always concerned with real people, since the economic models typically applied assume that individuals are motivated purely by selfish “non-ethical” concerns and are “not messed up by things as goodwill or moral sentiments” (1988:1). Sen (1987:4) continues by characterizing the engineering approach as:

...being concerned with primarily logistic issues rather than with ultimate ends and such questions as what may foster ‘the good of man’ or ‘how should one live’. The ends are taken as fairly straightforwardly given, and the object of the exercise is to find the appropriate means to serve them. Human behavior is typically seen as being based on simple and easily characterizable motives.

The much broader ethics-related tradition can be traced all the way back to Aristotle and his concerns for making judgments of social achievements and “the good for man”. Sen writes (1987:4):

This ethics-related view of social ‘achievement’ cannot stop the evaluation short at some arbitrary point like ‘efficiency’. These assessments have to be more fully ethical and take a broader view of ‘the good’.

Modern economics is often associated with positive economics, or what Sen calls the engineering approach. The fact that this thesis is called the economics *and* ethics of healthcare rationing is a typical example that ethical reasoning is something not typically associated with economics. Simply calling this thesis the economics or normative economics of healthcare (which was the initial plan) would be misleading to many of the potential readers of this work. In fact, over the years I have learned that more than a few even consider the idea of applying an economic perspective on health care rationing repugnant.

The primary ambition of this thesis is to contribute toward the ethics-related tradition of economics. However, it is important to acknowledge that the two approaches are not pure, but overlapping. Cost-effectiveness analysis (CEA)

highlights this by applying an engineering approach toward allocation decisions, based albeit (as we shall see) on a series of normative assumptions. Moreover, it is often necessary to spell out “what is” before approaching the question of “what ought to be”. The four papers included in this thesis illustrate this by applying various approaches, from strongly positive (Paper IV), to strongly normative (Paper II, Paper III), to somewhere in-between (Paper I).

Normative Ethics and Meta-Ethics

When referring to an ethical perspective on healthcare rationing it is useful to distinguish between *normative ethics* and *meta-ethics*. The ambition of normative ethics is to elaborate sound argumentation on moral questions. Moral questions are, roughly, questions that deal with the right and wrong, good and evil, associated with certain types of actions. For instance, normative ethics could be trying to deliver sound and well-articulated arguments for when, if ever, it is justifiable to withhold potentially beneficial care to patients, or what constitutes morally relevant reasons for treating individuals unequally. Meta-ethics deals with philosophical issues on a higher level of abstraction. It focuses on where ethical principles come from by trying to understand questions like: what *is* goodness? Is goodness a matter of taste or truth? Hence, meta-ethics focuses on trying to understand the nature of ethical thinking.

From the expressed aim, it should be clear that the type of ethical approach used in this thesis is normative ethics. Moreover, since the ethical discussions in this thesis address concrete moral issues, rather than trying to construct a comprehensive moral theory, this thesis could also be described as a work of *applied ethics*.

Reasonable Disagreement

An intricate but central concept in this thesis is *reasonable disagreement*. This thesis applies four conceptual themes – *individual responsibility*, *paternalism*, *incentives*, and *inequality* – in the context of healthcare rationing. Within these conceptual themes, reasonable disagreements are likely to arise on how to balance fairness and efficiency. It is not the ambition of this thesis to resolve

what is right and wrong in these cases of moral conflict. Rather the ambition is to highlight the conflicting moral principles that surround the four conceptual themes and present my own normative reflections.

Use of the “reasonable disagreement” concept implies that not all disagreements are reasonable. The idea is most notably present in the later work of John Rawls and is what leads to his idea of “overlapping consensus for the right reasons” (e.g. Rawls 1989, 1993), which was further developed and labeled “reasonable pluralism” by Josh Cohen (1994). Also, Norman Daniels (2008) uses the idea of reasonable disagreement in aiming to identify a fair deliberative process that could narrow or (optimally) dissolve reasonable disagreement. My conception of reasonable disagreement is constituted by: moral disputes where conflicting parties base their positions on arguments that are sound and logically consistent in a way that does not violate fundamental rights. In addition, the fundamental disagreement persists regardless of how open, well informed, and free from personal stakes the deliberation process has been.

Reasonable disagreements exist in many contexts (e.g. political, artistic, and philosophical), where individuals tend to come to different conclusions even though they share the same basis of knowledge and deliberation is conducted in good faith. The disagreements focused on in this thesis stem from substantive differences concerning the value base for conducting healthcare rationing.

In practice, many disagreements are rooted in personal stakeholding and inadequate knowledge. For example, many believe that resources are not scarce and, thus, they see no need to ration care. Secondly, various stakeholders involved in the healthcare context are often driven by agendas other than trying to achieve what is best for society at large. Thirdly, individuals may have a poor understanding of the practical implications of different rationing alternatives. Ultimately, the ambition of this thesis is to move past disagreements of this sort, reach the point of reasonable disagreement, and present my personal view on the matters. But reaching this point requires achieving adequate knowledge about the practical consequences that arise from different approaches toward healthcare rationing. One of the papers included in this thesis (Paper IV) primarily contributes to the matter of reasonable disagreement by increasing knowledge about the potential consequences a certain type of rationing might have on inequality.

Here, I also wish to clarify that when referring to justifiable grounds for healthcare rationing I intend a principal argument that, if challenged on reasonable grounds, becomes a part of reasonable disagreement. Consequently, the matter of finding justifiable grounds for saying no in the context of health care implies that any conclusion will inevitably be coupled with some form of reasonable disagreement.

Moral Intuition and Moral Reasoning

Healthcare rationing is a topic that commonly triggers our moral intuitions. For instance, most individuals have a strong moral intuition that it is wrong to deny medical assistance to someone in need, and that it is more important to save the lives of children compared to elderly patients. Given that moral intuition is such a prevalent feature in rationing, it is necessary to mention the role of moral intuition in relation to moral reasoning when discussing reasonable disagreements. This is not an easily resolved issue, but I will attempt to clarify the distinction and explain how I have chosen to handle the matter.

Moral intuition typically refers to the moral judgments (or responses) that occur quickly and carry a strong automatic belief about the moral appropriateness of an act, without having gone through a conscious reasoning process that produces this judgment. Moral intuition often manifests itself in strong revulsion or disgust for certain types of actions. Kass (1997:20) calls this form of moral intuition “the wisdom of repugnance” and exemplifies:

In crucial cases /.../ repugnance is the emotional expression of deep wisdom, beyond reason’s power fully to articulate it. Can anyone really give an argument fully adequate to the horror which is father-daughter incest (even with consent), or having sex with animals, or mutilating a corpse, or eating human flesh, or even just raping and murdering another human being? Would anybody’s failure to give full rational justification for his or her revulsion at these practices make that revulsion ethically suspect? Not at all.

Advocates for this type of emotional- or disgust-based “reasoning” asserts that intuitive repulsion often provides a good starting point for making moral

judgments, or at least that it should be considered enough to put the burden of proof on those who oppose a judgment based on moral intuition.

Moral reasoning is the attempt to support moral judgments by logically consistent reasoning derived from fundamental moral principles. Eyal writes (2008:114-115):

It seeks *general principles* [emphasis added] for the regulation of behavior that are such that they can command the agreement of other, informed, rational, and free agents. The test of a proposed principle or action is that it would be found acceptable not just to the person who proposes it but to all those whom the action affects. Justification is *to others*; moral motivation, the connection of morality with our will, is supplied not by sympathy alone but by our need to act in such a fashion that our actions are both understandable to ourselves and acceptable to others.

The ethical analysis in this thesis is written in the tradition of normative ethics and hence focuses on moral reasoning as a basis for finding justifiable grounds for healthcare rationing. I do not think that we can accept moral intuition as a source for overriding moral reasons when making judgments about healthcare rationing. Judgments based on moral intuitions are likely to offer poor moral guidance if based on an emotional foundation (e.g. Singer 2005). This is not to say that moral intuitions are usually misleading, or that they are easily distinguished from moral reasoning. Quite the opposite, I believe that it is of utmost importance for policy makers to have a solid understanding of prevailing moral intuitions and avoid policies that run counter to common moral intuitions, unless these counterintuitive policies are supported by powerful moral arguments or general principles. If not, the long-term legitimacy of the decision maker is at risk of being undermined. However, as Malmqvist (2008:31) rhetorically asks:

...if intuitive, theoretically unfounded judgments are a legitimate part of ethical reasoning, what room is there for reasoned argument? Can we ever do better than articulating and advocating conclusions that we already settled for on intuitive grounds? Can we ever be rationally persuaded to give

up those judgments that might simply be reflections of cultural bias, prejudice or evolutionary heritage?

Moreover, I see it as the task of moral philosophers to continuously challenge and question prevailing moral intuitions and see if they survive extensive moral scrutiny.

BACKGROUND

This background chapter is divided into two sections. The first section addresses several basic ambiguities concerning the concepts of *healthcare rationing* and *healthcare need*. The second section describes experiences gained from two real-world cases of attempting to engage in explicit healthcare rationing. The cases chosen are the initiatives undertaken in the state of Oregon during the 1990s and the ongoing priority-setting initiative taking place in Sweden. These cases are chosen to highlight some of the difficulties that different policy approaches toward healthcare rationing may encounter.

What is Healthcare Rationing?

As indicated in the introduction, rationing is a concept that carries a bad reputation. However, in its widest sense, rationing simply means the controlled distribution of scarce resources, and as such it occurs in good as well as in bad economic times. Moreover, rationing does not only involve negative consequences since efficient and fair rationing ultimately seeks to create opportunities to meet more healthcare needs than would otherwise be possible. This thesis uses the following definition of healthcare rationing: The withholding of potentially beneficial health care to individuals under conditions of scarcity.

The above definition is in accordance with how others have used the concept of healthcare rationing (e.g. Liss 1993; Ubel and Goold 1997; Norheim 1999). Following this definition, rationing aims to bridge the gap between need and available resources by limiting the possibilities to optimally satisfy healthcare needs. An in-depth discussion of rationing requires distinguishing between the different types of healthcare rationing. Building on the typology presented by Klein (1996), I will differentiate between the following three types of healthcare rationing:

- (i) Rationing by *denial* involves the general exclusion of particular types of healthcare services, e.g. tattoo removal, hearing aids, or laser eye surgery. It could also entail exclusion through changes in indications,

i.e. patients' healthcare needs must be associated with a more severe deterioration in health before treatment can be offered.

- (ii) Rationing by *dilution* implies that a particular healthcare need is only partially met. For example, rehabilitation to achieve full recovery after an injury may require 20 visits to a physiotherapist. However, in rationing by dilution, the patient might be granted only 10 visits to a physiotherapist, implying that the patient's health status is not fully restored. Finally, rationing by dilution could involve reductions in the quality of healthcare services, e.g. prescribing cheaper but less-effective drugs.
- (iii) Rationing by *delay* (often referred to in the literature as time rationing) means that patients must wait longer than necessary before receiving adequate treatment.

Moreover, as rationing takes place throughout the entire healthcare system it is also useful to differentiate between what I will refer to as *bedside* and *desktop* rationing. *Bedside rationing* occurs at the patient level when medical professionals through denial, dilution, or delay withhold services that could potentially benefit the patient. Bedside rationing is often not recognized as rationing by either patients or healthcare professionals. Instead, it is often viewed as a case of malpractice in situations where it can be established that a patient had to forgo what would have been a medically beneficial treatment. However, healthcare professionals regularly make judgments, for example, concerning whether a particular diagnostic test or treatment warrants the cost associated with it – not every patient that comes in with a tummy ache gets a full body scan to rule out the presence of a tumor. *Desktop rationing* is a more abstract type of rationing that occurs outside of individual physician-patient encounters, where policy-makers withhold services that could potentially benefit patients. What further distinguishes desktop rationing from bedside rationing is that the former typically affects statistical patients temporally distant from when the actual decision was made. Bedside rationing, on the other hand, typically involves identifiable patients in a context of personal decision making.

Although rationing is more common than one might think, not all forms of limit setting can be considered acts of rationing. Ubel and Goold (1997) have established three criteria separating limit setting from rationing.³ They assert

³ Ubel and Goold (1997) use the criteria to assess bedside rationing. However, I find them equally applicable to desktop rationing.

that a rationing decision must involve: 1) withholding, withdrawing, or failing to recommend a service that, according to best clinical judgment, is in the patient's best medical interests; 2) the decision to promote primarily the interests of someone other than the patient, and 3) the decision-maker must have some extent of control over the use of the beneficial service.

Given these criteria, denial of a treatment because it is considered futile is not a case of rationing. For example, not ordinating transfusion of donkey blood to patients suffering from allergy is not rationing, since scientific evidence and clinical judgment do not show any potential medical benefits from its use. A less obvious case would be decisions resulting from concerns of overusing treatments, which could impair their future effectiveness. Denying patients potentially beneficial antibiotics is such a borderline case of rationing. At the bedside level, this decision corresponds to the three criteria. At the desktop level, however, this decision is not a case of rationing since a decision to limit the prescription of antibiotics is intended to promote a medically beneficial effect among patients in general. Prescribing antibiotic treatment in every case where it might be medically beneficial could invoke resistance against antibiotics in society, which would be a more serious consequence. Hence, conflicts between bedside and desktop rationing may arise since what is in the best interest of a specific patient might not be in the best interest of patients in general.

It is also important to acknowledge that resource scarcity related to rationing does not necessarily have to be monetary. As has already been mentioned, time is a scarce resource, which constantly forces medical staff to ration care. Moreover, resource constraints may also be physical as in the case of organ transplantations, where limited availability of organs forces healthcare professionals to engage in rationing decisions regarding who should and who should not receive an organ.

Finally, it is important to clarify the distinction between rationing and priority setting. These are two closely related concepts and are often used synonymously. Although this is understandable – since rationing cannot take place without an initial stage of priority setting – the concepts are nevertheless clearly and importantly distinct from one another. Priority setting involves ranking different services for defined groups of patients and putting some service ahead of others, i.e. choosing what *to do*, while rationing focuses on what *not to do*. This thesis centers mainly on issues surrounding healthcare

rationing, but is nonetheless highly relevant to issues concerning priority setting.

What is Healthcare Need?

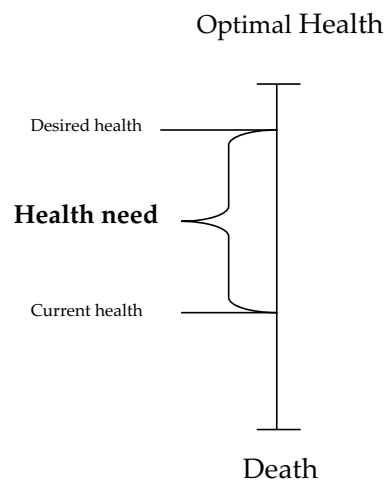
Decision-makers frequently use some notion of need to justify healthcare rationing. Despite widespread agreement that need is an essential component in conducting fair and efficient rationing, there is no consensus on what rationing *according to need* actually means; should it be interpreted to mean the severity of the health state, or the individual's capacity to benefit? To what extent should nonmedical factors such as social circumstances and individuals' past be incorporated?

The complexity inherent in questions like these gives many different meanings to the notion of rationing according to need, leading to quite different substantive rationing principles in practice. So, to avoid an incomprehensible discussion, it is important to expose some of the relevant dimensions for assessing healthcare need. I will not, however, make a fully comprehensive analysis of the concept of healthcare need. Instead I will outline only the distinction that is most relevant in the context of this thesis, namely the distinction between need for *health* and need for *health care*. Many of the thoughts presented below are heavily influenced by the work of Liss (e.g. Liss 1993 and National center for priority setting in health care 2007).

The need for health can be understood as the gap between current health and desired health (Figure 1). A person's desired health does not necessarily coincide with optimal health, since individuals typically desire less than optimal health; especially as we get older and view some deterioration in functionality as a natural course of aging. Further, the gap does not necessarily only involve one's current health; it could also involve the risk for future ill health. For example, when we take vaccine as a preventive measure we are trying to reduce the gap between the current risk and the desired risk⁴ of becoming sick.

⁴ Desired risk often corresponds to optimal risk, which would be no risk of becoming sick.

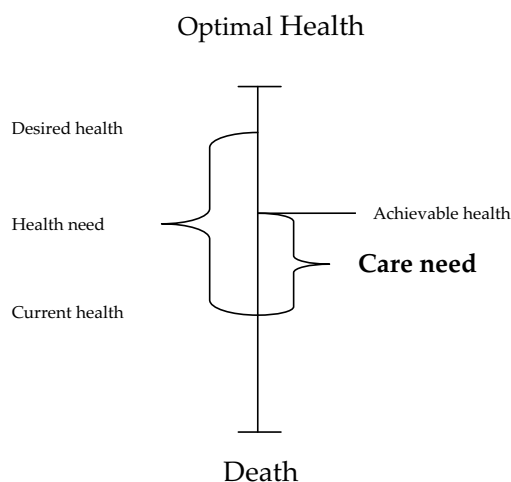
Figure 1. Graphical illustration of health need.



The defined concept of health need as a gap between current health and desired health does not, however, yield sufficient information for us to make rationing decisions on the basis of need. We must also assess what the individual has a need *for*. What kind of care is needed to reduce the gap between current health and desired health? A patient suffering from end stage renal disease has a health gap that can be reduced by kidney transplantation. Hence, two prerequisites must be present for a healthcare need to arise:

- i) A health need must exist, i.e. a gap between current health and desired health.
- ii) A care need must exist, i.e. an intervention that potentially can reduce the health gap must exist.

Figure 2. Graphical illustration of healthcare need.



Given the above definition, it is possible for an individual to have a health need without this being accompanied by a care need. If no treatments are available to reduce the gap between current health and desired health, then we cannot say that a healthcare need exists. It is important to make this distinction because the major determinant of individuals' health is not health care, but other social determinants (Marmot and Wilkinson 2005). These determinants include the economic and social conditions under which people live, and are much more influential as risk factors for many common diseases, e.g. cardiovascular disease and type II diabetes (Marmot and Wilkinson 2006). Although social determinants have an extremely important effect on health, I will touch only briefly on the issue since the main focus of this thesis is health care and how to find justifiable grounds to ration health care, not health (although this could be a consequence of rationing).

Degree of urgency is another dimension commonly used to define healthcare need. For example, it can be argued that a patient who faces an immediate threat to life and/or health has a higher degree of healthcare need compared to patients who face equal, but less immediate, threats to life and/or health. This view of healthcare need, which focuses more on the temporal aspect, is sometimes labeled the *rule of rescue*. The rationale behind this rule is that society has a obligation to do everything possible to help save those individuals facing an immediate threat to life and/or health (McKie and Richardson 2003).

Explicit Healthcare Rationing in Practice

Abstract models and theoretical concepts are of little use to us in the absence of a real-world context where they can be applied. This section presents a brief description of the practical experiences gained from initiatives undertaken in the State of Oregon in United States and in Sweden. These real-world experiences also serve as practical examples of approaches that have, from the outset, emphasized two opposing views on distributive fairness; maximizing health benefit within the population versus giving priority to those with the greatest need.

The Case of Oregon

Internationally, the most well-documented initiative to engage in explicit rationing was undertaken during the early 1990s by the State of Oregon. The intent of the initiative was twofold; to expand eligibility to high-priority services among Oregon's population to cover everyone below the federal poverty level, and to use limited resources to provide the most cost-effective services as a means to maximize health benefits among the population (Ham 1998). To achieve these objectives, a ranking list of condition-treatment pairs based on cost-effectiveness ratios were developed. Depending on the state's Medicaid budget constraint, every second year the state officials would literally draw a line across the list, ensuring public coverage for all items above the line, but leaving items below the line for individuals to fund out-of-pocket (Oberlander, Marmor et al. 2001).

To achieve any substantial savings and create resources to expand coverage, Oregon was required to exclude hundreds of serious condition-treatment pairs from coverage. The initial list revealed several counterintuitive results. For example, tooth capping was ranked above appendectomies for appendicitis (see Table 1), despite the fact that the latter is typically a lifesaving intervention. While problems in some of the data led to such results, this is nevertheless an expectable result from cost-effectiveness prioritization – the problem arose because life saving treatments like appendectomies are typically much more expensive than tooth capping (Oregon estimated that it could cap a tooth in over 100 patients for the cost of a single appendectomy).

Hence, a small benefit for a large number of individuals was ranked above a great benefit for one.

Table 1. Examples from the initial priority list in Oregon based on cost-effectiveness analysis (Hadorn 1991).

Treatment	Exp. net benefit	Exp. duration of benefit	Costs (US \$)	Priority ranking
Tooth capping	.08	4	38	371
Surgery for ectopic pregnancy	.71	48	4015	372
Splints for temporomandibular joint disorder	.16	5	99	376
Appendectomy	.97	48	5744	377

The initial rankings, which were based cost-effectiveness ratios, invoked strong negative public reactions. Following massive criticism, public consultations and medical experts were used to provide input on the initial list, which was then adjusted extensively (Hadorn 1991). Individual items were moved up and down the ranking list “by hand”, informally guided by factors such as the number of affected patients, societal value placed on the item (e.g. high value was placed on palliative and child care) (Hadorn 1996).

Four years were spent revising the initial list, and the final list was made available in 1994. At that time, list included 696 items, with the cut-off point drawn at 565 (Ham 1998). In the final list, the costs associated with treatments had a negligible influence. Instead the final list was based primarily on the expected benefit associated with treatments. Hence, the final adjusted list ranked surgery for ectopic pregnancy and appendectomy among the top items, while splints for temporomandibular joint disorder and tooth capping were dropped altogether. Most of the items that ended up below the cut-off point were services where individuals generally were considered to possess the ability to be responsible for their own care, or were conditions for which no effective treatments were available (Ham 1998).

To some extent, the initiative undertaken in Oregon was a success. It succeeded in decreasing the percentage of uninsured from 19.9% of the working age population to 7.6%. This was achieved while the percentage of uninsured in the US as a whole was on the rise between 1990 and 1996 (Alakeson 2008). Further, the state’s ambition to engage in explicit rationing received strong support among the public, which created a necessary platform

for constructive dialogue (Rutledge 1997). It is important to acknowledge, however, that the Oregon initiative affected only the Medicaid system, i.e. the poor population. As Daniels (2008:152) points out, “the plan could not avoid the appearance of the *haves* setting priorities for the *have nots*”. This is a distinguishing characteristic compared to similar efforts undertaken in systems with universal health insurance coverage.

Lessons from Oregon

Perhaps the most important lesson from the Oregon experience was that it illuminated some of the implicit value judgments imbedded in cost-effectiveness analysis (CEA), which are likely to conflict with deeply held values among the public. This led to abandoning the idea of using CEA as the sole principle for rationing, since as Hadorn (1991:2219) insightfully points out:

...any plan to distribute healthcare services must take human nature into account if the plan is to be acceptable to society. In this regard there is a fact about the human psyche that will inevitably trump the utilitarian rationality that is implicit in cost-effectiveness analysis: people cannot stand idly by when an identified person's life is visibly threatened if rescue measures are available.

The moral conflicts that arose in Oregon regarding the initial CEA ranking can be characterized in what Daniels (1994) has called four unsolved rationing problems; *the aggregation problem, the priorities problem, the fair chances versus best outcome problem, and the democracy problem*. These are all practical problems in healthcare rationing, but very moral in character. Hence, people are likely to disagree on what constitutes the correct course of action.

The aggregation problem can be framed by the question: When should society allow an aggregation of modest benefits to larger numbers of people to outweigh more significant benefits to fewer people? The approach taken in Oregon was based on the economic rationale of aggregation, i.e. health maximization. This led to some non-intuitive results, e.g. that certain lifesaving treatments were ranked below some more trivial condition-treatment pairs. Aggregation clearly goes against the rule of rescue, which prescribes that rationing should be done through one-to-one comparisons,

giving priority to the patient with the most urgent need. The strong reactions that the initial list evoked showed that most people do not want society to have 'maximize health benefits across population' as its sole objective for distributing resources. However, this does not mean that individuals are likely find all forms of aggregation impermissible.

The priorities problem can be framed by the question: How much priority should society give to treating the sickest or most disabled individuals (i.e. those with the largest *health* need)? If two individuals are competing to receive priority for a treatment that will give them an equal amount of health benefits, most people share the moral intuition that priority should go to helping the worst-off individual. Ranking based on cost-effectiveness analysis (CEA) ignores this common intuition and is neutral between individuals in cases where the benefits are of equal size. The fact that the revised Oregon list ranked palliative and child care⁵ at the top of the list indicates that the public did not accept this neutral standpoint, but wanted to give some level of priority to the worst off.

The fair chances versus best outcome problem can be framed by the question: To what extent should society strive towards producing what is considered to be the best outcome, instead of ensuring that all individuals are given equal or proportional chances of receiving treatment? CEA focuses solely on the outcome in terms of maximizing health in the population. It could, however, be argued that it is more important that everyone has a fair chance to receive treatment. For example, imagine Larry and Jeb who are competing for the same treatment, but only one of them can get it. Larry will survive 3 additional years if he receives the treatment, while Jeb will only survive 2 additional years. The example could analogously be framed as patient groups competing for scarce resources. Following the CEA rationale, Larry should receive the treatment. However, Jeb might insist that it is unfair that he has to stand back only because Larry will live longer. Instead, he might argue that it would be more fair to have a weighted lottery where Larry has a 60% chance of getting the treatment and Jeb has a 40% chance of getting the treatment. It is unclear if this type of rationale played any significant role in abandoning the initial CEA methodology.

⁵ Child and palliative care offer an example of two different interpretations of who is the worst off. Children are worst off in the sense that they have not had their fair share of a full life, and palliative patients are worst off in the sense of severity of illness. Neither example are however the same as *health-* or *health care* need as defined in the earlier section, which further illustrate the complexity with regards to defining need in the health sector.

Finally *the democracy problem* can be framed by the question: When should society rely on a fair deliberate process as the only way to determine what constitutes a fair rationing outcome? The normative relevance of public preferences is a complicated matter. Should *public* preferences be based on the general population, or involve only the preferences of those who have experienced the specific health condition or have other forms of expert knowledge? What is evident from the Oregon experience is that one cannot ignore the views of the general public. Nevertheless, it is unclear how much weight one should give to moral reasons compared to expressed public preferences in cases when these differ. A related somewhat disturbing question about human behavior which the experience from Oregon raises is: Could it be that people are only *reasonable* or *fair minded* as long as they are not directly affected by the outcome themselves?

The Case of Sweden

Swedish experience with explicit rationing represents a different approach compared to that of Oregon. In contrast to Oregon's cost-effectiveness approach, the Swedish approach has appealed to individuals' equal value and rationing according to need. Values related to cost-effectiveness and maximizing aggregate health in society have been relegated to a secondary role. Moreover, rather than explicitly listing services that should not be publicly funded, Sweden opted for an approach based on explicit principles to which any rationing decision should adhere.

The Swedish initiative to engage in a more open and systematic approach toward healthcare rationing started in 1992. An economic downturn led to recognition, at the national level, of the unavoidable need to ration care. The idea was that openness and transparency would create legitimacy for such politically difficult decisions. This led to the formation of a parliamentary priorities commission assigned to:

“consider the responsibilities of health and medical services, their demarcation and role in the welfare state; highlight fundamental ethical principles which can furnish guidance and form a basis of open discussions and of prioritization in health and medical services.”(Ministry of health and social affairs 1993:29)

The commission's work resulted in an ethical platform for making priority-setting decisions in health care (Government bill 1996/97:60). This platform consisted of three principles intended to guide decision makers at all levels in the healthcare system when making rationing and priority-setting decisions. The principles were (and still are):

- I. *The principle of human dignity.* Meaning that all individuals have equal value and rights regardless of personal characteristics or position in society.
- II. *The principle of need and solidarity.* Meaning that resources should be used in domains (or patients) where needs are considered to be greatest.
- III. *The cost-effectiveness principle.* Meaning that resources should be used in the most effective way without neglecting fundamental duties to improve health and quality of life.

The principles are ordered lexically in the sense that the human dignity principle has superiority over the need and solidarity principle, which in turn has superiority over the cost-effectiveness principle. Following the commission's work, the legislated goals of the Swedish healthcare system were amended. Prior to the commission's report the Swedish Health and Medical Service Act (1982:763 §2) stated that: "the goal of all health care services is good health and health care on equal terms for the entire population". The amendment which was added to the formerly stated goal was: "Provision of health care services must respect the equal value of all human beings, and the dignity of the single human being. The person with the greatest need for health care services should be given priority." (Swedish Health and Medical Service Act 1997:142 2§). This means that the cost-effectiveness principle is not explicitly mentioned in the preamble of Swedish healthcare legislation, while the principles of 'human dignity' and 'need and solidarity' are. This further underlines the superiority of these two principles in comparison to the cost-effectiveness principle.

In practice, since it gives little or no real guidance on how to set priorities, the principle of 'human dignity' has played a minor role in actual rationing decisions. It gives some indication on how not to set priorities, i.e. personal characteristics such as age, gender, ethnicity, and socioeconomic status should

not influence how priorities are set, unless particular medical relevance is associated with these personal characteristics.^{6,7} Instead, the principles of ‘need and solidarity’ have to become the leading principles when making rationing decisions.

Over the years, several national initiatives have aimed at applying the priority-setting principles in explicit rationing decisions. Since 2004, the National Board of Health and Welfare has used the priority-setting principles to rank pairs of health condition-intervention in process for producing national treatment guidelines on various disease categories (Carlsson 2010). Based on the ethical principles, the Board has used a specially designed model for priority setting. This model is described as a “pragmatic” interpretation of the ethical platform, where severity of illness is balanced against cost-effectiveness considerations (Carlsson, Kärvinge et al. 2007). In similar fashion, the Dental and Pharmaceutical Benefits Agency has applied the priority-setting principles when making reimbursement decisions on public funding for pharmaceuticals. In their work, cost-effectiveness has become a key determinate when making rationing decisions with regards to pharmaceuticals (Erntoft 2010).

At the autonomous regional level, several attempts have also been made to apply the priority-setting principles in practice. The first ambitious attempt to apply such principles was undertaken by the County Council of Östergötland in 2003, which developed a set of explicitly defined limitations on providing health care. Examples of rationing were that treatments at pediatric clinics for some minor childhood conditions (e.g. head lice, obesity) and the second hearing aid would no longer be funded publicly (Bäckman, Karlsson et al. 2006). The process leading to explicit rationing in Östergötland resulted in a heated debate across the entire country and was heavily criticized from leading politicians. Most likely this discouraged other county councils from following in Östergötland’s footsteps. During the past few years, however, a new wave of initiatives to engage in explicit rationing at the regional level has arisen. A handful county councils have created lists including hundreds of services and have decided to exclude some of the services with the lowest ranking to create funding space for more important services. These more

⁶ For example, it makes little sense to screen men for breast cancer. Hence, in that case, gender becomes a medically relevant personal characteristic.

⁷ However, when allocating vaccine against swine flu, the rule of human dignity was abandoned and priority was given to individuals “important to the functioning of society as a whole”. This further illustrates that the principle of human dignity tends to be more of a symbolic gesture than anything else.

recent initiatives have met surprisingly little objection and appear to have gained public support (Carlsson 2010; Waldau, Lindholm et al. 2010).

To some extent the initiative undertaken in Sweden has been a success. The principles defined by the commission have been applied in actual rationing decisions, which prove that they are applicable at least as a framework for departure in real practice. This has spurred several national and regional initiatives to openly discuss the necessity to set limits on what the public can offer. The National Centre for Priority Setting in Health Care has emerged as a venue for interdisciplinary research that will hopefully produce further valuable insights on how priorities are set in practice. The activities described above signify a long-term commitment from the national level of the health care system to openly discuss and explore the matter of explicit rationing and to develop a systematic approach for conducting rationing.

Lessons from Sweden

In 2005, the Swedish government assigned the National Board of Health and Welfare to follow up the national guidelines for priority setting in health care and their implementation. Based on this extensive report (National center for priority setting in health care 2008) three general problems in the Swedish approach, underpinned with ethical values, can be identified. I will label these: *the vagueness problem, the balancing problem, and the leveling problem*

The vagueness problem can be framed by the question: Should ethical principles be symbolic or guiding? The inherent vagueness of the priority-setting principles has arguably rendered them of little guiding value when making distributive decisions. The Swedish 'human dignity principle' has a strong position legally, but how it should be applied in practice, when it comes to distributing scarce resources, is unclear. For example, the principle gives no indication about when age becomes a medically relevant characteristic for rationing. Moreover, the principle of human dignity may appear as less vague if it had been framed as an all-embracing procedural principle, focusing on creating equal or proportional opportunities to health care. The 'human dignity principle' does, however, have strong symbolic value that could potentially strengthen legitimacy for the healthcare system, both internally and externally. The 'need and solidarity principle' also holds strong symbolic value. However, the meaning of need has never been sufficiently specified, limiting its applicability as a guiding principle. For instance, to what extent

should individuals' capacity to benefit be taken into account when assessing need?

The balancing problem can be framed by the question: Should one ethical principle have superiority, or is it necessary to balance different principles against each other? It could be argued that the strict lexical order of Swedish principles is both inapplicable in practice and highly unlikely to be in accordance with the true preferences of the public. It seems unreasonable that rationing decisions should be based solely on individuals' level of need, thereby disregarding other aspects such as costs and capacity to benefit. For example, strict adherence to the lexical order would imply that healthcare resources would disappear into a bottomless pit⁸ when it comes to treating patients with little or no capacity to benefit. Since this is not what is happening in practice, it indicates that the lexuality of the Swedish principles should be regarded as an ambition in trying to steer the process into giving more weight to individuals' health needs compared to their capacity to benefit. However, the strict lexical order stated is, at the very least, utterly confusing when trying to understand the role of the cost-effectiveness principle.

The final stylized lesson from Sweden is *the leveling problem*. This can be framed by the question: Should ethical principles apply equally when making rationing decisions at the individual level and at the population level? The Swedish principles are formulated as if they are supposed to guide rationing decisions made by all actors in health care, i.e. the national, regional, and clinical levels. However, ambiguity exists concerning the application of cost-effectiveness, which should be applied only at the population level. One could argue that it is questionable to apply a different value basis depending on whether or not rationing decisions concern statistical or identified patients. On the contrary, adherence to the rule of rescue is a strong moral intuition among many, and one could argue that it is an important aspect when trying to foster a compassionate society.

As a concluding remark concerning the Swedish experience it seems like the ethical principles to some extent have promoted an open dialogue regarding inevitable rationing decision. Although one could argue that openness helped to increase awareness of the value base for rationing decisions, transparency and actual understanding, regarding the "true" values which underpin

⁸ If it is assumed that the needs principle does not incorporate *capacity to benefit*.

Background

rationing decisions might have decreased as a result of the *vagueness, balancing, and leveling* problems.

THEORETICAL CONTEXT

This chapter outlines the theoretical context to which paper I-IV and the extended discussion included in this thesis relate. The theoretical landscape of healthcare rationing will be approached from two nonmutually exclusive perspectives; labeled the *economic* and the *ethics* perspectives. These perspectives allow explicit exploration of issues related to efficiency and fairness, which inevitably arise when discussing healthcare rationing.

Economics

The theory of economics does not furnish a body of settled conclusions immediately applicable to policy. It is a method rather than a doctrine, an apparatus of the mind, a technique of thinking which helps its possessors to draw correct conclusions.

-John Maynard Keynes

The economic perspective on rationing is concerned with how to allocate scarce resources as *efficiently* as possible to best satisfy human wants. However, a normative basis of efficiency is needed for economics to generate a satisfactory indication of how to best allocate resources. Welfare economics has traditionally provided such a normative basis.

Welfare Economics

Given that welfare economics is a normative theory, it is based on value assumptions. The first fundamental value assumption relates to the concept of economic efficiency⁹ traditionally defined through the *Pareto principle*. The Pareto principle stipulates that for a change to be efficient it must leave at least one person better off at the same time as no one else is left worse off. Or as stated by Pareto (1906:261) himself:

⁹ This should not be confused with technical efficiency, which is the ability to do more with the same resources, or alternatively to achieve the same outcome with fewer resources.

We will say that the members of a collectivity enjoy maximum ophelimity in a certain position when it is impossible to find a way of moving from that position very slightly in such a manner that the ophelimity enjoyed by each of the individuals of that collectivity increases or decreases. That is to say, any small displacement in departing from that position necessarily has the effect of increasing the ophelimity which certain individuals enjoy, and decreasing that which others enjoy, of being agreeable to some, and disagreeable to others.

The second fundamental value assumption in welfare economics is that the assessment of efficiency should be based on individuals' preferences as regards their own situation. Consequently, Pareto efficiency is coupled with the notion of *consumer sovereignty*, assuming that individuals are the best judges of their own preferences. According to the welfare economic theory a market with perfect competition, provided that certain assumptions regarding individual preferences hold,¹⁰ automatically leads to a Pareto-efficient and socially desirable distribution since all exchanges are assumed to be voluntary.¹¹ This implies that welfare economics gives little or no room for public interventions (except enforcing property rights) as long as the market is fully functioning and there is no wastage.

The healthcare market is, however, not a fully functioning market. Quite the opposite; the healthcare market is associated with extensive market failures,¹² which prevent Pareto-efficient allocations to arise. Thus, left to its own economic devices the healthcare market is likely to be highly inefficient. This creates a strong argument for public intervention to avoid inefficient and undesirable outcomes. Public interventions will, however, almost without exception, create both gainers and losers. Consequently, the Pareto principle is of little practical use when distributing scarce healthcare resources since applying it as guiding principle would block most public attempts of redistribution. In practice, economists have therefore relied on the less-strict, *potential Pareto principle* as a more practical definition of efficiency. The potential Pareto principle¹³ stipulates that redistribution is efficient and desirable if those made better off can hypothetically compensate those who are

¹⁰ The assumptions are *completeness*, *reflexivity* and *transitivity*. (see e.g. Shiell 2000)

¹¹ However, the market must fulfil certain demands to ensure that there is no waste: *efficient exchange*, *efficient allocation of production factors*, and *efficient choice of what to produce* (see e.g. Lipsey 2007).

¹² For example, *asymmetric information* and *externalities*. For more details on this topic see Paper I (Tinghög, Carlsson et al 2009) or Donaldson and Gerard (2005).

¹³ The potential Pareto principle is sometimes called the compensation principle or the Kaldor-Hicks criteria.

made worse off, while still making the “gainer” better off. Unlike the Pareto principle, there is no actual requirement on compensation from the “gainer” to the “loser”. The better off person could, if he or she wanted, compensate those who are made worse off, but this does not mean that he or she will actually do it. What efficiency means, according to the potential Pareto principle, is simply that benefits exceed burdens.

A general criticism of both Pareto principles is that they are efficiency oriented but equitably neutral, and as such give no consideration to the overall distribution of benefits and burdens among the population. Hence, welfare economics can justify policies that few people support, e.g. providing additional benefits to those already well-off. Sen (1979:22) comments:

An economy can be optimal in this sense [i.e. in the sense of both Pareto principles] even when some people are rolling in luxury and others are near starvation as long as the starving cannot be made better off without cutting into the pleasures of the rich.

Still, welfare economics requires an additional normative judgment about what it means for individuals to become “better off” for it to make practical sense. Traditional welfare economics assumes that individuals’ welfare is determined solely by the distribution of material resources. This assumption was questioned and refined during the 20th century when several economists proposed the idea that welfare involved both material and nonmaterial aspects that needed to be incorporated into the welfare concept (e.g. Pigou 1912). This conception of welfare as something which incorporates both material and nonmaterial aspects was denoted utility. Welfare economics has therefore become closely allied with utilitarian ethical theory.

Utilitarianism

The classical utilitarian normative claim underpinning welfare economics is that an act is justified if, and only if, it maximizes the overall good in society. Sen (1987) dismantles the utilitarian ethical theory into a combination of three elementary requirements.

- i. *Welfarism*, which requires that goodness of a state of affairs be a function only of the utility information regarding that state.
- ii. *Maximization*, which requires that utility information regarding any state be assessed by looking only at the sum of all the utilities in that state.
- iii. *Consequentialism*, which requires that judgment regarding choices and actions be determined solely on goodness of the consequent state of affairs.

It is not uncommon to use all of these requirements interchangeably to mean *utilitarianism*. This, however, is misleading since other moral principles incorporate one or two of these requirements, but are still not equivalent to the utilitarian moral principle. For the sake of understanding, it is therefore important to distinguish these requirements from utilitarianism, which necessitates that all three requirements be fulfilled.

The first requirement, *welfarism*, deals with suitable units for valuing social welfare. According to utilitarian moral theory, social welfare can be conceived as a function of individual utilities that incorporates both material and nonmaterial aspects of welfare through preference satisfaction. The view on what part of individuals' preference satisfaction should be considered to constitute the utility that utilitarians strive to maximize is, however, open to a wide array of interpretations. Modern welfare economics identifies two main conceptions of utility that give rise to welfare. As Kahneman and Sugden (2005:162) comment:

Two different interpretations of the term 'utility' have been used in the literature of economics. In its original interpretation, which derives from Bentham, utility is interpreted in hedonistic terms, as a measure of pleasure and pain / . . . / Of course, this is not the way that most economists have used the term 'utility' over the last 100 years. During this period, utility has usually been understood as decision utility. Decision utility is a representation of preferences, and the concept of preference is understood in terms of choice: a person's preferences are the mental entities that explain his choices, and are revealed in those choices.

The original hedonistic interpretation of utility as proposed by Bentham (1789) was simply that utility is equivalent to happiness (or pleasure). Thus, the right

course of action would be the one that creates “the greatest happiness for the greatest number.” Later work by Mill (1861) introduced the idea that not all forms of happiness should be regarded as having equal value when estimating total utility. In a much-cited quote Mill states: “It is better to be a human being dissatisfied than a pig satisfied; better to be Socrates dissatisfied than a fool satisfied.” The second main conception of utility rests on the idea that it is not happiness that should be singled out for maximization; instead, utility should be interpreted as representation of individuals’ preferences. The problem with utility as representation of individuals’ preferences satisfaction is that individuals do not always tend to prefer what is best for them. Individuals make mistakes and often have shortsighted preferences.¹⁴

The second requirement, *maximization*, stipulates that utilities should be aggregated across individuals according to an unweighted sum-ranking rule. This rule implies that only the sum of individual utilities is of importance when making allocation decisions. However, for a utility function to be measurable it needs to be represented by a valid cardinal utility function. Cardinal utility captures the strength of preference differences as opposed to ordinal utility, which only captures the ranking of preferences. For example, suppose a beer derives a utility of 12 utils, a cup of coffee 8 utils, and a cup of tea 4 utils. If preferences are represented by a valid cardinal utility function we could not only say that a beer is preferred compared to a cup of coffee, we could also state that it is preferred by exactly the same amount by which a cup of coffee is preferred to a cup of tea.

The third requirement, *consequentialism*, stipulates that the rightness or wrongness of an action should be assessed on the basis of the outcome, or consequence it produces. Hence, consequentialism is at times expressed through the aphorism; “The ends justify the means.” Moreover, it is usually understood as being distinct from deontology, in that deontology assesses the moral rightness of an act from the character of the act itself, rather than the consequences this act leads to. A deontologist could, for example, argue that the act of lying is always wrong regardless of the consequences that follow from telling the truth. While a consequentialist could argue that the foreseeable negative consequences from lying could make it permissible to lie. This means that no instrumental value is attached to processes that lead to consequences in utilitarianism. However, as Culyer (1998) points out, a

¹⁴ For more thorough exposition about conceptions of utility see, e.g. Hausman (2010) and Brock (1973).

process could, in principle, be conceptualized as a consequence if it affects the overall level of utility.

A general criticism often expressed against utilitarianism, which can be directed at a combination of the three requirements, is that utilitarianism ignores individual freedom (see e.g. Olsen 1997). When only consequences matter, it is easy to imagine actions that might increase total utility, but which do not correspond to a commonsense conception of justice, e.g. slavery and torture. Because of such counterintuitive implications, many utilitarians have abandoned the strict act-oriented version of utilitarianism for the rule-oriented version. Rule utilitarianism advocates the introduction of rules to guide the actions of individuals. This is because such rules are expected to maximize the total amount of utility in the longer run by creating a sense of security, reducing social tensions, and so on.

Cost-Effectiveness Analysis and Quality Adjusted Life Years

Cost-effectiveness analysis (CEA) is a standard economic tool for determining how to maximize health benefits when resources are limited. Thus CEA provides healthcare policy-makers with information about how to ration care in accordance with welfare economics and the utilitarian ethical theory. The outcome of CEA is usually expressed in terms of a cost-effectiveness ratio, where the health gain is the denominator and the corresponding cost is the numerator. Consequently, the cost-effectiveness ratio could be viewed as the "price" of the additional unit of health gain achieved by switching from current practice to the new strategy (e.g. €5000 per gained life-year or quality adjusted life year). If the decision-maker's willingness to pay is greater than the price, the new health program can be said to be cost effective. Hence, when a health program is said to be cost effective this simply indicates that it gives good value for money. CEA has nothing to do with saving money; actually a health program that saves money may very well be deemed cost ineffective. For the outcome of a CEA to result in an actual rationing decision it requires a final value judgment regarding what constitutes good value for money – an issue where CEA provides no guidance.

The quality adjusted life year (QALY) is the most widely used measure of health benefits in CEA. QALYs are often interpreted as utilities, where utility

is assumed to be dependent on health status (Q) and the number of life years. This can be described by the following formula:

$$QALYs = U(Q, T) = V(Q) \times T$$

Where V(Q) is the value function of quality measured on a scale between 1 (full health) and 0 (death) corresponding to the weight, which is multiplied by life years (T). For instance, an individual with a quality level of 0.5 who will live 10 additional life years will generate 5 (0.5*10) additional QALYs.

When applying QALYs in CEA it is important to distinguish between *measuring health* and *valuing health* through QALYs. When measuring health through QALYs, generic preference-based questionnaires such as EQ-5D or SF-36 are most commonly used. These questionnaires include generic dimensions for describing health, which permit comparisons between different medical conditions. The dimensions for EQ-5D are: mobility, self-care, usual activity, pain/discomfort, and anxiety/depression. Each dimension has three levels, and patients are asked to indicate which level is most in agreement with their own situation. By combining answers from the dimensions a health profile can be assigned to the respondent. In the case of EQ-5D, 243 health profiles are possible ($3^5=243$ health profiles). (see e.g. Brazier, Ratcliffe et al. 2007)

When valuing health through QALYs, the impact that different health profiles have on welfare is estimated by so-called *preference elicitation techniques*. The two main techniques, which are founded in expected utility theory, are time trade off (TTO) and standard gamble (SG) (Johannesson 1996). Usually representatives from the public have been asked to value hypothetical health states. This implies that individuals rarely value their own health state in CEA.¹⁵ Both TTO and SG involve presenting the respondent with a choice. In TTO, the respondent can choose to give up some life years to live for a shorter period in full health. If the respondent chooses not to trade off any life years to achieve higher quality of life it is assumed the he or she has full health (1 QALY). If the individual is indifferent between living 8 years in full health and 10 years in the presented health state, a value of 0.8 QALYs is assigned to the presented health state. In SG, the valuation of QALYs is generated in a similar fashion. Respondents are asked to choose between remaining in the presented

¹⁵ This is a violation against the second fundamental value assumption in welfare economics, i.e. that individuals are the best judges of their own welfare.

health state for a period of time, or choosing a treatment that has a chance of either restoring them to perfect health, or killing them. If the respondent is unwilling to accept any risk of death to achieve full health, a value of 1 QALY is assigned to the presented health state.

For QALYs to represent a cardinal utility function they need to reflect the intensity of individuals' preferences. This implies that individuals should prefer the option that maximizes the expected amount of QALYs for the QALY to be a representation of individual utility over health. If an option that does not maximize QALYs is preferred, then QALYs cannot be taken to reflect individual utility over health. Pliskin et al. (1980) provided an axiomatic framework for identifying when QALYs represent a valid cardinal utility function, which is derived from von Neumann-Morgenstern expected utility theory.¹⁶ According to the framework, three conditions must be satisfied for QALYs to represent individual utility over health: *mutual utility independence*, *constant proportional trade off*, and *risk neutrality over life years*.

Mutual utility independence implies that life-years (T) and quality (Q) must be mutually independent. If either T or Q is held fixed at some level, utility independence imposes that preferences with respect to gambles over the other attribute do not depend on the particular value chosen. For example, if an individual is indifferent between living 10 years in a health state with moderate pain and a gamble with a 50% chance of 10 years with severe pain and a 50% chance of 10 years in full health, the individual should also be indifferent between these alternatives if the time horizon is set to 20 years instead of 10. When valuing health through SG, utility of a health state is measured as a fraction of full health, and if mutual utility independence holds the quality weight (which is equivalent to the outcome from SG) should always be the same, irrespective of time horizon.

Constant proportional trade off implies that the proportion of remaining life that an individual is willing to trade off for a specific health/quality improvement is independent of the amount of time he or she is expected to remain alive. For instance, if an individual is indifferent between 10 years with severe back pain and 5 years of full health, then he or she should also be indifferent between 10 years with severe back pain and 20 years of full health. When valuing health through TTO, utility of a health state is measured as the number of years in

¹⁶ This theory will not be described further here. The reader is referred to the original work of von Neumann and Morgenstern (1947).

full health divided by the number of years in the assessed health state. If constant proportional trade off holds, this ratio should be the same, irrespective of time horizon.

Risk neutrality over life years implies that individuals should be indifferent between living 2 years in health state x and a lottery with a 50-50 chance of living in health state x for either 1 or 3 years. This implies that individuals should have a constant marginal utility from additional years in a specific health state, i.e. every additional year in full health is of equivalent value to the individual.

If these conditions are satisfied, individuals will rank different outcomes in accordance with their desirability. The outcome that maximizes QALYs will be the most the desirable. If the conditions are not satisfied, the ranking of outcomes based on the number of QALYs it yields gives no indication of desirability.

Later work by Johannesson (1994) and Bleichrodt et al. (1997) have graphically and mathematically demonstrated that the three conditions proposed by Pliskin et al. (1980) under certain assumptions can be reduced to two: risk neutrality over life years in all health states and *zero condition*. The latter implies that for a duration of zero life years, all health states derive the same utility, which in the context of health is entirely self-evident. These conditions are, however, sufficient only if we assume a chronic health state where quality of life is constant until death. If quality of life is not constant, an additional condition of *additive separability* needs to be fulfilled. This condition implies that utility of a health state in a specific period is unaffected by the health state that precedes or follows it (Broome 1993).

The QALY Approach Toward Rationing – Critique

The so-called “QALY approach” toward healthcare rationing is to allocate available resources in a way that maximizes the number of QALYs gained. This can be achieved by giving priority to medical interventions that have a relatively low cost per QALY gained. Criticism against the QALY approach to rationing health care can be categorized into the same three main groups that form the requirements for utilitarian moral principle, i.e. *welfarism*, *maximization*, and *consequentialism*.

Welfare economics singles out welfare or utility as the object of maximization. For CEA to fit into the welfare economic framework, QALYs need to represent a valid cardinal utility function, i.e. satisfy the conditions of *mutual utility independence*, *constant proportional trade off*, and *risk neutrality over life years*. However, individual preferences used to value health states are frequently found to violate these conditions (Tsuchiya and Dolan 2005). This means that QALY maximization does not reflect desirability since it does not capture non-health benefits appropriately. According to classical hedonistic utilitarianism, utilities are typically taken to stand for happiness, but there is little evidence that healthier individuals with more QALYs are also happier individuals. This has led to the emergence of a nonwelfarist approach,¹⁷ where QALYs are interpreted to be an indication of health, but the actual utility does not necessarily have to change proportionally to the size of QALY gains. If QALYs do not represent utilities, the outcome from a CEA cannot be applied as decision rule unless the main objective is to maximize health in the sense of QALYs. However, the outcome may still be a valuable input to decision making since maximizing health is likely to be an important (but not the only) objective.

A second major criticism relates directly to the desirability of maximization, and that it ignores concerns regarding distributional fairness. Considerable empirical evidence shows that people do not think resources should be allocated simply to maximize outcomes. Studies show that people have a general preference for directing health resources toward individuals who are in very poor health states, even when doing so means generating less overall health benefit and thereby sacrificing some QALYs (see e.g. Nord 1999).

A related criticism toward the QALY approach and the desirability of maximization concerns the assumption that certain groups are discriminated against. Most notably the QALY approach has been criticized for being unfair to elderly individuals with preexisting disabilities and conditions (Harris 1987). For example, a life-prolonging treatment aimed at otherwise healthy individuals who could be saved to lives in perfect health would *ceteris paribus* generate more QALYs compared to the equivalent program aimed at individuals with preexisting disabilities who are unable to achieve perfect health. It is also easy to imagine that aged individuals, due to natural causes, can be expected to have a shorter remaining lifespan than younger individuals. Hence, at times, the QALY approach is accused of being *ageist*

¹⁷ Which is, confusingly, often referred to in the literature as extra-welfarism.

since, *ceteris paribus*, older individuals with a shorter life expectancy will always receive lower priority compared to younger individuals (e.g. Harris 1987; Evans 1997; Tsuchiya 2000).

The QALY approach evaluates allocations by their consequences on aggregated QALYs. It attaches no instrumental value to the method upon which the outcome is based. A wide range of evidence suggests that the procedure on which decisions are based upon, influence the individuals' judgment of the consequence that follows from a decision. For example, in a study by Frey and Pommerehne (1993) shows that, raising prices to ration demand was considered less fair than allocation by a "first come, first served" rule, or allocation by an authority, although the consequences were the same.

A final criticism¹⁸ directed at the standard QALY approach is that future health benefits are assigned lower value compared to more immediate health benefits. This issue has received surprisingly little consideration within the framework of CEA. However, is it really a justifiable moral view that the outcome of a CEA recommends saving one life today rather than 131 lives in 100 years?¹⁹

Ethics

Justice obliges us to pursue fairness in the promotion of health, but policy needs the guidance of ethics in determining what this mean

-Norman Daniels

The ethical perspective on healthcare rationing is concerned with how to allocate scarce resources as *fairly* as possible to best satisfy human wants. Understanding the population-level ethics perspective is, thus, closely aligned with the distribution of scarce resources and applying different theories of fairness to healthcare need. For this to make any practical sense, however, we need to establish what is meant by *fairness*. In contrast to the economic

¹⁸ For a more extensive overview of the ethical issues related to CEA and QALYs see, e.g. Brock (2003)¹⁹ If a 5% annual discount rate is applied.

¹⁹ If a 5% annual discount rate is applied.

perspective, which has a rather coherent framework for interpreting efficiency, fairness is open to a wide array of interpretations. This section will present of five theories on distributive fairness with particular relevance in the context of healthcare rationing; *maximization*, *egalitarianism*, *maximin*, *sufficientarianism*, and *prioritarianism*. In addition some ideas related to *equality of opportunity* and *procedural fairness* will be presented.

Distributive Fairness and Healthcare Rationing

Five main views on substantive fairness are commonly presented in the philosophical literature to make sense of fairness in the context of health care (Olsen 1997; Cookson and Dolan 2000). Insight into these, and how they can be applied in ranking healthcare need, provides a clearer understanding of the normative basis of different ethical approaches to healthcare rationing.

Fair distribution, according to the *maximization principle*, is that which maximizes the total sum of the entity to be distributed. As regards healthcare need, this translates into a distribution of healthcare resources that maximizes the total sum of health in society. Consequently, the maximization view on fairness places a strong emphasis on individuals' capacity to benefit from treatment, while giving no extra weight to healthcare need of those who are most severely ill or worst off. Utilitarianism is a particular type of maximizing view, which focuses on utility as the relevant entity for distribution. Utilitarians would, however, claim that their view on fairness does give extra weight to the most severely ill. The rationale for this claim is that a particular health improvement for a severely ill individual is likely to create greater utility improvement compared to a healthier individual who experiences the same health improvement. For instance, a paraplegic individual who suddenly recovers the ability to move a finger would most certainly derive a higher increase in utility compared to an otherwise healthy individual who recovers the ability to move the same finger.

Fair distribution, according to the *egalitarian view*, is that which creates the most equal outcome of the entity to be distributed. As regards health, this implies that focus should be placed on those individuals with the lowest level of current health, and even slight improvement for those individuals should receive priority. The primary concern among strict egalitarians is, therefore, the health state which an individual is in before receiving treatment, while no

consideration is devoted to individuals' capacity to benefit from treatment. In contrast to the maximization view on distributive fairness, the egalitarian view is comparative, implying that it is concerned with how people fare relative to others. Everyone should receive an identical share even if adherence to equality reduces the overall level of health or utility in society. Because of this feature, many have rejected egalitarianism due to the so-called *leveling-down objection* (e.g. Nozick 1974 and Temkin 1993). The leveling-down objection states that there is nothing desirable about equality in itself if everyone becomes worse off. For example, Temkin (2003) suggests that it is unreasonable to consider that an outcome where everyone becomes blind is more desirable than an outcome where only one person becomes blind.

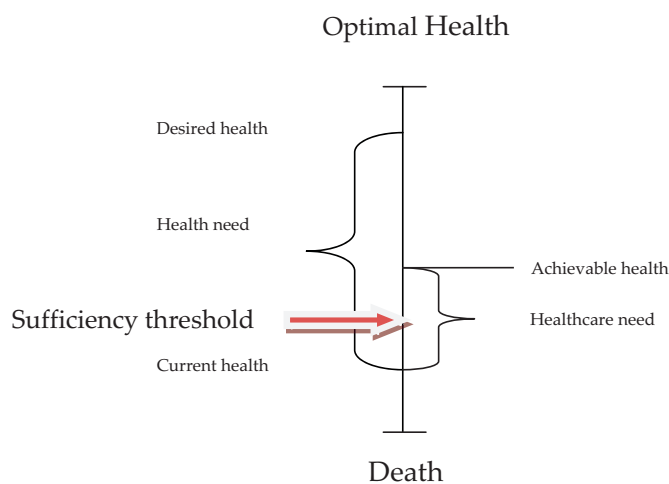
A less-strict version of the egalitarian view is the Rawls' type of egalitarianism of *maximin view* (Rawls calls it the *difference principle*). As opposed to the strict egalitarian view, the maximin view accepts inequalities, but only under the premise that it is not possible to further improve the situation for the worst off. Applied to health, this implies that the most severely ill should receive priority as long as the treatment reduces the health gap to some extent (no matter how small this reduction might be). However equality in itself is not the most important aspect for those in support of the maximin view.

The *sufficientarian view* on fairness is another, less-strict version of the egalitarian view. It proclaims that fairness requires that everyone gets enough of the entity that is to be distributed, but not that everyone gets the same. When individuals are "sufficiently" well off, it is no longer warranted that society should give priority to an individual merely because he or she is worse off. As Frankfurt (1987:34) suggests: "The fundamental error of egalitarianism lies in supposing that it is morally important whether one person has less than another regardless of how much either of them has". And again (1987:268): "We tend to be quite unmoved, after all, by inequalities between the well-to-do and the rich /.../ The fact that some people have much less than others is morally undisturbing when it is clear that they have plenty".

As regards health, the sufficientarian view implies that individuals below a certain level of current health should always receive priority regardless of how much a treatment reduces the existing health gap. Priority should be given until they have been elevated above this minimum, predefined, health state, which we can call the sufficiency threshold (illustrated in figure 3). Hence,

ranking on the basis of healthcare need implies that everyone gets *sufficient* health before applying any considerations regarding the capacity to benefit. The problem of inequalities in the sufficientarians' view is not that some people have a better current health state than others, but that some have an unacceptably low level of current health. Consequently, sufficientarianism – as opposed to egalitarianism – focuses on the alleviation of absolute detritition instead of relative health inequality when assessing healthcare need.

Figure 3. Graphical illustration of a health sufficiency threshold.



Many philosophers have argued about where the predefined threshold should be set. Sen (1985) and his basic capability approach represents a sufficientarian view proposing that the minimum level should be set so that everyone is equally assured of the basic capability to function in important and valuable ways.²⁰

The prioritarian view on fairness stipulates that benefiting an individual matters more the worse off that person is. This view was most prominently developed through the work of Temkin (1993) and Parfit (1995, 1997). As opposed to the egalitarian and sufficientarian views, the prioritarians allow considerations that involve individuals' capacity to benefit. As regards health, this implies that it is more important to reduce the health gap for the worse-off person than it is to achieve the equivalent reduction for a better-off person. The

²⁰ Martha Nussbaum (2000) has later attempted to define the capabilities that should be supported by all democracies.

prioritarian view is most accurately viewed as a family of views since different versions vary in terms of how much priority they give to the needs of those worst off. Prioritarians who wish to give only a little extra weight to the healthcare needs of those who are worse off are close to the utilitarian rule of ranking, which is based purely on individuals' capacity to benefit. Prioritarians who want to give substantial extra weight to the healthcare needs of those who are worst off are, on the other hand, close to the egalitarian rule of ranking, which gives no consideration to the capacity to benefit.

Fair Equality-of-Opportunity and Healthcare Rationing

In addition to the theories on distributive fairness, which focus on fairness related to the distribution of specific outcomes, the ethics perspective also encompasses theories related to the distribution of individuals' prospects or opportunities. Rawls (1971) initially introduced the concept of equality-of-opportunity in his idea of *justice as fairness*. According to this idea, social and economic inequalities are justifiable if they satisfy two conditions. The first is known as the *difference principle*, which asserts that inequalities are only justified if they are advantageous to the worst-off in society. The second condition is what has become known as the *fair-equality-of-opportunity principle*, which asserts that justice should not favor those with advantageous social contingencies.²¹ Over the years, many philosophers have elaborated on trying to further specify the meaning of fair-equality-of-opportunity. Roemer (1995) further explicates the general equality-of-opportunity view on fairness:

Equality-of-opportunity views are closely allied with a commitment to personal responsibility. Society under such views is not required to insure individuals against bad results when they are the consequence of individual choices made after opportunities have been equalized. On the equality-of-outcome view, in contrast, society's mandate is to render all lives equally

²¹ The first principle in Rawls idea of justice as fairness, which I have not mentioned, is what is known as the equal liberty principle. This principle asserts that everyone should have equal right to the most extensive basic liberties, e.g. freedom of speech and political liberty.

successful, at least in so far as this is feasible. Thus, persons are not effectively held responsible for their choices.

As regards health, the equality-of-opportunity view focuses on equalizing people's opportunity for lifetime health rather than achieved levels of health. According to this view, an unequal distribution of health and healthcare resources could be considered fair if people choose not to exercise their opportunity for achieving health. Dworkin (1981) formulated the influential distinction between *brute* and *option luck*. Brute luck refers to occurrences of luck that individuals could not have anticipated or avoided, while option luck denotes deliberate gambles or actions:

If I buy a stock on the exchange that rises, then my option luck is good. If I am hit by a falling meteorite whose course could not have been predicted, then my bad luck is brute (even though I could have moved just before it struck if I had had any reason to know where it would strike) (Dworkin 1981:73).

This distinction helps shape the general *luck egalitarian* idea that inequalities are fair as long as they follow from individuals' deliberate and fully informed choices (see e.g. Arneson 1989). However, outcomes that stem from choices or behaviors over which the individual has no control should be deemed unfair. In the context of health care this could imply that an individual who has deliberately engaged in an unhealthy lifestyle should be given lower priority compared to an individual who has acted "responsibly". However, individuals should not be held responsible for the characteristics of the group they belong to. Hence, according to the equality-of-opportunity view it may be justifiable to give a "responsible" person within a group with similar opportunities to live a healthy lifestyle priority. But since individuals rarely have similar opportunities to act responsible, because of the social determinants of health, a less-healthy lifestyle might be just as responsible as a healthy lifestyle since individuals have different degrees of power to exercise responsibility (Roemer 1995).

Procedural Fairness and Healthcare Rationing

Procedural fairness refers to the idea that disputes concerning resource allocations should be resolved through a just process. Hence, it is the characteristics of the procedure itself that guarantee achievement of a fair

outcome. Dolan et al (2007) identify three reasons why procedural characteristics matter: first the procedure could matter simply because it promotes better outcomes (the consequentialist view); second, it might matter because the procedure has an inherent value in itself (the proceduralist view); third, procedures might be important because they promote values that are not associated with outcomes, e.g. legitimacy for the decision-maker (the instrumentalist view).

The most renowned theory of procedural fairness related to health care is Daniels and Sabin's (2008) *accountability for reasonableness* (A4R). A4R extends Rawls' theory of justice to include health care as a primary good that should be distributed fairly.²² Due to the natural lottery, the need for health care is much more varied across the population in comparison to other primary goods, e.g. food, clothing, and basic liberties such as freedom of thought. This makes it less likely for people to reach agreement on a single moral principle for resolving disputes concerning how to distribute scarce healthcare resources compared to other primary goods.²³ Given that reasonable people will inevitably disagree on what is fair distribution of healthcare resources, disputes need to be resolved through a fair process. Daniels (2000:1300) explicates:

In pluralist societies we are likely to find reasonable disagreement about principles that should govern priority setting. For example, some will want to give more priority to the worst off, some less; some will be willing to aggregate benefits in ways that others are not. In the absence of consensus on principles, a fair process allows us to agree on what is legitimate and fair.

The key thought behind the A4R is that although individuals are likely to disagree about first-order reasons for resource allocations, they ought to be able to reach a consensus on second-order reasons concerning how to handle disagreements on first-order reasons. A4R offers four conditions that a decision-making process for allocation of healthcare resources must meet to be legitimate and fair:

²² Rawls' theory of justice assumes that everyone has full health.

²³ Rawls states that behind a veil of ignorance individuals will unanimously choose to maximize the primary goods among the worst-off individuals.

- (i) *Publicity condition*: Decisions regarding rationing and their rationales must be publicly accessible.
- (ii) *Relevance condition*: The rationales for rationing should aim to provide a reasonable explanation of how the organization seeks to provide 'value for money' in meeting the varied health needs of a defined population under reasonable resource constraints. Specifically, a rationale will be reasonable if it appeals to evidence, reasons, and principles that are accepted as relevant by fair-minded people who are disposed to finding mutually justifiable terms of cooperation.
- (iii) *Revision and appeals condition*: Rationing decisions must have some degree of reversibility. A mechanism must exist to enable unfavorable decisions to be disputed. Further, decisions must be revised in the light of new evidence or arguments.
- (iv) *Enforcement condition*: A means must be in place to regulate the process to ensure that conditions 1 to 3 are met.

This procedural account for healthcare rationing and priority setting has been endorsed in many countries, e.g. Mexico, UK, Canada, and New Zealand. Several authors, however, have raised objections to this procedural account (e.g. Rid 2009, Friedman 2008, Sabik and Lie 2008). First, and most notably, the approach has been criticized for not making a clear distinction between legitimacy and fairness. Even though a fair process may yield legitimate outcomes, these outcomes might still be unfair. Second, it has been argued that the crucial relevance condition is too vague, and therefore it offers little guidance in practical decision-making. Third, the appeal by Daniels and Sabin for a fair process has been criticized (by people who insist that the process alone cannot tell us what is right or fair) for understating the importance of substantive principles when making rationing decisions. However, this criticism is somewhat incorrect since A4R also contains the "relevance condition", which I understand to stipulate that "fair-minded" people have reached agreement on the relevant substantive principles.

MAIN FINDINGS AND DISCUSSION – A WELFARE ECONOMIC PERSPECTIVE ON HEALTHCARE RATIONING

In this chapter I will briefly describe the four papers included in this thesis and discusses their main findings from a welfare maximizing perspective. Each of the papers corresponds to one (or more) of the conceptual themes that will also serve as a framework for the extended discussion in the following chapter, when approaching the matter of finding justifiable grounds for healthcare rationing from a broader ethical perspective. The conceptual themes are: *individual responsibility, paternalism, incentives, and inequality*.

Paper I: “Individual Responsibility for What? – A Conceptual Framework for in a Publicly Funded Healthcare System”

Although it seems likely that some notion of individual responsibility is an essential feature in rationing decisions, it is rarely explicitly articulated or evaluated in the context of health policy. Paper I presents a tentative conceptual framework for exploring when healthcare services contain characteristics that facilitate individuals’ ability to take responsibility for their own health care through private financing.

The potential role of individual responsibility in healthcare policies is not a new topic. The debate surrounding individual responsibility for health has revolved mainly around three philosophical issues: When do causal factors constitute effort or circumstance? When should choices relating to health be regarded as genuinely deliberate and free as opposed to being socially conditioned? What constitutes socially acceptable risk taking with regard to one’s own health? These interrelated issues center on a retrospective notion of individual responsibility, i.e. the moral relevance of holding people responsible for past health-related behavior. However, so far, no satisfactory

answers on how to incorporate retrospective responsibility into actual policies have emerged.

This paper applies a prospective notion of individual responsibility for health care, focusing on responsibility in terms of the healthcare services that individuals are capable of financing and managing without public involvement or subsidies – a conception of individual responsibility that has been argued to be more policy-relevant in welfare-promoting societies (Le Grand 2003).

When exploring this forward-looking conception of responsibility the characteristics of a specific healthcare service and the market failures surrounding the specific service become crucial. In some instances, the individual may be fully capable of managing the care required by certain health states. Six attributes for exploring the suitability of private financing for specific healthcare commodities are identified and discussed:

- (i) it should enable individuals to value need and quality both before and after utilization
- (ii) it should be targeted toward individuals with a reasonable level of individual autonomy
- (iii) it should be associated with low levels of positive externalities
- (iv) it should be associated with a demand sufficient to generate a private market
- (v) it should be associated with payments affordable for most individuals
- (vi) it should be associated with ‘lifestyle enhancements’ rather than ‘medical necessities’.

These attributes enable exploration of individual responsibility connected to health care as a heterogeneous group of commodities. Attributes 1 to 3 in the framework (sufficient knowledge, individual autonomy, and positive externalities) are all connected to the specific characteristics of health care that violate crucial assumptions for a well-functioning market. These attributes allow identification of services that can be allocated through the working of market mechanisms without resulting in inefficient outcomes due to market failures.

Attribute 4, sufficient demand, is further connected to specific characteristics concerning the provision of health care, which may call for public funding to ensure that individuals are able to take individual responsibility. Healthcare services that (for reasons other than market failures) are not available through private alternatives are unsuited for individual responsibility. The lack of these services would consequently make them high-priority candidates for inclusion in the publicly funded healthcare system.

In addition, policymakers also have an obligation to protect people from unbearable financial costs associated with poor health. Rationales for collective responsibility begin with some concept of fairness, usually involving the belief that no one should be denied health care due to the inability to pay, or the idea that public funds should not be spent on 'pleasure seeking' instead of pain/symptom avoidance. Such beliefs about equality and fairness constitute an important basis for including the concepts of affordability and 'lifestyle enhancements' in the framework.

Paper II: "Discounting, Preferences, and Paternalism in Cost-Effectiveness Analysis"

When making rationing decisions based on CEA, discounting enables policy-makers to compare competing health programs which entail outcomes that appear at different times. Paper II assesses the moral reasonableness of the arguments traditionally motivating public authorities to discount health outcomes when making intertemporal rationing decisions through cost-effectiveness analysis. In addition, the paper addresses the issue of when it is justifiable for policy-makers to override individuals' fully informed time preferences in making rationing decisions over time.

Frank Ramsey (1928) presented one of the earliest formulas addressing the wide range of reasons expressing why resources available in the future should be valued less than the same amount available today:

Social discount rate = $\delta + \epsilon_x g_x$

Where δ denotes individual time preference, ϵ_x denotes the elasticity of the marginal utility of the good x , and g_x is the growth rate of good x . This formula

helps to disentangle three main components why resources available in the future should be valued less than the same amount given today. First, individual time preference (δ) is motivated by the assumption that individuals are impatient and, in general, prefer to consume a given amount of resources sooner rather than later. Second, the growth rate of good x (g_x) motivates discounting because if society can transform resources available today into greater amounts in the future through investments there is an opportunity cost that a social time preference should take into account. The third part of the formula, the elasticity of the marginal utility of the good x (ϵ_x), serves as a link between these two main sources. This term implies that if we have more of a good in the future, the utility that the individual derives from the same unit of that good in the future is of less magnitude.

The paper concludes that the behavioral effects related to time preferences advanced in the literature give little or no reason for why society at large should favor the present over the future when making *intergenerational* choices regarding health. From a fairness perspective it seems impermissible that society should favor current generations over future generations purely on the basis of position in time. Moreover, there is the argument that contemporary democracy may be insufficiently sensitive to the concerns of future generations, which makes a paternalistic approach towards the social discount rate preferable. The fact that behavioral effects give little support for discounting future health does not mean that discounting in CEA should be abandoned altogether. Factors such as growth and diminishing marginal utility, which are likely to change as functions of temporal distance, still justify assigning less value to future outcomes. However, this does not mean that temporal distance in itself gives a legitimate reason for discounting, and there is no *prima facie* reason to assume that these factors should apply equally to health and money.

Regarding *intrapersonal choices*, the paper categorizes preference failures, which justify paternalistic intervention, into two distinct groups – myopic and acratic.

Myopic preference failures concern the cognitive inability to compute information adequately to make choices that maximize welfare of the individual. For example, this could include cases where logical or causal connections between choice and outcome are difficult to make, or where the quantity of information is too great for the individual to assess (Goodin 1993; New 1999). The information itself might, however, be perfectly adequate.

Judgments about probabilities and likely future outcomes of actions are particularly common myopic preference failures. The work of Tversky and Kahneman (1979) has shown empirically how people with perfect information commonly display inadequate reasoning power to interpret and calculate the implications of certain statistical or probabilistic situations.

Acratic preference failures concern the existence of reduced willpower. This could include situations where individuals act against their own better judgment and fail to do what is in their long-term best interest. For example, individuals often lack the ability to emotionally assess the force of information – knowing the facts, but not feeling the force of an outcome psychologically. Hence, even though individuals may have good judgment as to what is necessary for long-term good, they frequently choose a different course of action. This line of argument is commonly used to motivate mandatory public programs, ensure that individuals save for retirement and avoid procrastination, or to deny the preferences of addicted persons once they have checked into a rehabilitation center.

Paper II concludes that since myopic and acratic preference failures are common in the context of health choices, elicited time preferences are of little normative relevance when making decisions on the social discount rate. Some time-dependent factors of individuals' time preferences, e.g. dread and anticipation, could potentially constitute a legitimate basis for discounting in the intrapersonal context once they have been separated from myopic and acratic elements. As with intergenerational discounting, the combined argument from growth and diminishing marginal utility supply the strongest arguments for discounting in the intrapersonal context. However, there is no *prima facie* reason to assume that this argument should apply equally to health and monetary values.

Paper II is written from a traditional welfare-economic perspective focusing on solutions that aim to maximize welfare. However, it is worthwhile to consider the option of stepping outside of the traditional framework of utilitarianism when considering the issue of discounting future health. If the societal objective of making intertemporal choices is solely to maximize net welfare over time, this could easily lead to an unfair distribution of healthcare resources across generations, forcing one generation to make excessive sacrifices for the sake of future generations. Intuitively, it seems unjust to

require the current generation to commit itself to large-scale medical experiments with severe immediate consequences for the purpose of benefiting future generations, although this assumingly, would increase the net sum of welfare over time. No generation can be morally required to make more than a certain sacrifice for the sake of future generations. This is an issue that the utilitarian framework for discounting within CEA does not consider. However, CEA in general does not consider how burdens or benefits are distributed, nor does it reflect every aspect of importance in the allocation of scarce healthcare resources. This could be considered a weakness if one chooses to view CEA as a decision rule rather than decision input.

Paper III: “Incentivizing Deceased Organ Donation: A Swedish Priority-Setting Perspective”

Healthcare rationing does not necessarily have to be associated with financial shortage. It can also be associated with nonfinancial factors. Organ transplantation is a typical example of rationing driven by nonfinancial factors. The limited number of available organs forces healthcare professionals to engage in rationing decisions to determine who should and who should not receive an organ. Paper III approaches the issue of healthcare rationing indirectly by first developing a proposal for compensated deceased organ donation that could potentially address the organ shortage and increase overall welfare in society. In addition, the paper examines whether the monetary incentives in the proposal are compatible with the ethical values of Sweden’s healthcare system.

Using incentives to increase the supply of organs is often associated with a mental image of the rich buying organs from the desperately poor. In reality, however, incentives are any factors (financial or nonfinancial) that motivate individuals to take a particular course of action. In other words, incentives are the expectations that encourage people to behave in a certain way and could be tailored to achieve organizational success. Financial incentives, e.g. salaries and bonuses, are the most obvious types of incentives, but in certain settings nonfinancial incentives could be far more powerful tools when trying to influence individual behavior. When people act against economic self-interest to promote wider societal interest, various types of nonfinancial incentives can usually explain why they choose to behave prosocially.

Incentives for individuals to behave prosocially can be divided into three categories: *extrinsic motivation*, *intrinsic motivation*, and *signaling motivation*. Extrinsic motivation refers to motivation by factors external to the individual. For example, individuals may be motivated by material rewards such as money. These rewards provide satisfaction and pleasure that the behavior or task itself may not provide. Intrinsic motivation refers to motivation from factors internal to an individual; such as the pleasure one derives from the task itself, or from the sense of satisfaction in completing or even working on a task. Pure altruism, meaning selfless concern for the welfare of others without regard to reward or the benefits of recognition, could be one such intrinsic motivational factor for behaving prosocially. Finally, signaling motivation refers to motivation by an individual's concern about how he or she is perceived by others, i.e. we are more inclined to feel good about ourselves when others hold us in high esteem. These categories of motivation are mutually interdependent and may interact and possibly crowd out one another, resulting in less prosocial activity in some instances (Mellström and Johannesson 2008). Consequently, to increase an individual's willingness to engage in prosocial activities, it is important for policy makers to have a sufficient understanding of how different types of incentives interact.

According to behavioral agency theory, prosocial activity can be partially explained by differences in individuals' preferences to signal altruism in attempting to increase social esteem, or strong preferences to avoid signaling greed (Bénabou and Tirole 2006; Ellingsen and Johannesson 2007). These models incorporate all three types of incentives and give reasonable explanations for why individuals in some specific settings tend to lower their prosocial activity when material rewards are introduced (Frey and Oberholzer-Gee 1997; Gneezy and Rustichini 2000; Mellström and Johannesson 2008). It is also widely acknowledged that individuals will act more prosocially if their behavior is observable. Several field studies have shown that donations to charity increase when associated with a high degree of publicity (Harbaugh 1998; Soetevent 2005).

Building on behavioral agency theory and empirical insights concerning how to increase prosocial behavior, the paper presents a proposal for compensated deceased organ donation. This proposal incorporates intrinsic, extrinsic, and signaling motivations and can be summarized as follows:

- i. €5000 contribution to the estate or family of the deceased toward funeral expenses.
- ii. Deceased (advanced directive) or family could decline the contribution in full.
- iii. All or part of the monetary contribution could be directed as a publicized donation to a reputable charity of the deceased's (if the will is known) or family's choice.
- iv. In all circumstances, the right to self-determination of the deceased must be respected, and the decision of the family cannot be in contradiction to the explicitly stated provisions of the will of the deceased.

The proposal is carefully discussed in relation to the three ethical priority setting principles; human dignity, needs and solidarity, and cost efficiency. The discussion shows that the proposal for compensated deceased donation would be fully compatible with the priority setting principles, and therefore it merits serious consideration.

In arriving at this conclusion, the paper outlines three potential moral pitfalls associated with incentivizing individuals' health choices through monetary compensation; *coercion*, *commodification*, and *loss of public trust*. These potential pitfalls are all aspects that could make excessive use of incentives in health care ethically objectionable. Hence, the fundamental question that needs to be asked when incentivizing individual choices concerning their own health is: When do these objections become strong enough to hinder society from maximizing overall welfare? In the case of our incentive-based proposal we do not view these objections to be strong enough to allow people to die. Consequently, we suggest that the potential benefits of introducing a regulated system of compensated deceased organ donation outweigh potential disadvantages in the context of Swedish health care.

Paper IV: "Horizontal Inequality in Rationing By Waiting Lists"

Various mechanisms can be used to ration health care, the most common of which is price. Demand is restrained by monetary price, which the patient faces at the point of demand. In public systems where care is often free or priced well below market-clearing level, demand and supply must be

reconciled through other rationing mechanisms. Hence, publicly funded systems commonly rely on waiting lists as a mechanism to limit availability, while still trying to maintain that persons in equal need are treated equally. Paper IV investigates whether socioeconomic inequalities exist in waiting times for six types of elective surgeries in the Swedish healthcare system. To achieve this aim, the paper sets out to test five hypotheses that more resourceful groups receive priority when care is rationed through waiting lists:

- i. Male patients receive priority over female patients.
- ii. Younger patients receive priority over older patients.
- iii. Patients with higher incomes receive priority over patients with lower incomes.
- iv. Patients of Swedish origin receive priority over patients of foreign origin.
- v. Patients currently active in the workforce receive priority over patients currently inactive.

Paper IV was based on a retrospective approach in the sense that the lengths of the waiting times were calculated for all patients who received an elective surgery in 2007. This implied that patients who remained on the list, or exited the list for reasons other than surgery, were not included in the data set. Data on waiting time for elective procedures were obtained from the administrative healthcare register in the county of Östergötland, Sweden. The elective procedures were based on the ICD-10 codes used to classify nonacute surgical procedures. In total, 24 nonacute surgical procedures were included. These 24 elective procedures were aggregated into six surgical categories; vascular, orthopedic, gynecologic, urologic, ophthalmologic, and general surgery. The study included 4636 completed waiting-time episodes. Multiple linear regression models were used to investigate the associations between waiting times and our selected predictors; age, gender, disposable income, ethnicity, and workforce activity. Data on these independent variables were retrieved from Statistics Sweden (SCB), and by using the civic person identification number we could link waiting-time data from an administrative database to socioeconomic registers.

The results presented in paper IV provide some support for the assertion that horizontal inequality occurs when rationing elective surgeries through waiting lists. This picture, however, is not clear-cut, as inequality was not present in all

surgical specialties. No evidence of horizontal inequality was found in vascular and urologic surgery. Concerning gynecology, the results contradicted the stipulated hypothesis, i.e. that more resourceful groups receive priority when care is rationed by waiting lists. In the specialties of orthopedics, ophthalmology, and general surgery the data supported the presence of horizontal inequalities. Particularly noteworthy in this context was the finding that lower household income predicted longer waits in both orthopedics and general surgery.

In connection with the finding that low income predicted longer waiting times in orthopedics and general surgery, it is worth noting that these were also the surgical specialties where patients in general experienced the longest waiting times. A hypothesis raised by the findings of the paper is that income becomes an increasingly influential factor as waiting time becomes longer. Alternatively, the findings could suggest the existence of a threshold effect, where disutility associated with waiting is negligible as long as the waiting time does not exceed a certain length. This may be explained by the existence of an initial adjustment period allowing individuals to mentally prepare and make necessary practical arrangements. But in cases where time on the waiting list exceeds this threshold, the disutility associated with continued discomfort and anxiety take over, making individuals increasingly active in trying to gain priority, thereby favoring more resourceful groups.

The paper did not confirm the hypotheses concerning ethnicity and gender. Hence, men and patients of Swedish origin did not appear to receive priority. A possible explanation for why these hypotheses were not confirmed could be that the study focused solely on differences among patients who have already been referred for a specific elective procedure. Also, it is possible that females and patients of foreign origin could be discriminated against before being placed on the waiting list by a referring physician, but not once they have been placed on the waiting list.

This study documents that access is unequal when waiting lists are used to ration care, but only to a limited extent. Whether this is good or bad news depends on your worldview, i.e. your expectations on the functioning of public health care and your normative views on how it should function. From a policy-maker's perspective it is important to recognize that horizontal inequalities arise even when care is not rationed through ability to pay. Consequently, horizontal equality in waiting times cannot be assumed in

publicly funded healthcare systems unless this is closely monitored and openly discussed. Although our findings are to some extent context-specific, they are relevant for decision-makers in other countries that rely heavily, like Sweden, on waiting lists as a mechanism to ration care.

Main findings and discussion

EXTENDED DISCUSSION – A POPULATION-LEVEL ETHICAL PERSPECTIVE ON HEALTHCARE RATIONING

In this chapter I will extend the discussion raised in the papers by applying a broader ethical perspective via the same four conceptual themes relating to healthcare rationing used in the preceding chapter, i.e. *individual responsibility*, *paternalism*, *incentives*, and *inequality*. Each of these conceptual themes represents areas where reasonable disagreements are likely to occur. My ambition with this chapter is not to resolve the reasonable disagreement associated with these conceptual themes, but rather to bring forward some of the complexities that surround them. In addition, I offer some personal reflections on what I believe to be the most appropriate way forward in dealing with these normative issues in practice.

Individual Responsibility – A Prospective Approach

Paper I, *Individual responsibility for what?*, deals primarily with the question of how individual responsibility for health care can be efficiently applied when rationing care. It focuses on the heterogeneity of healthcare commodities, the occurrence of market failure in the healthcare context, and the implications this could have for individuals' ability to take responsibility for their own health care. The question of how individual responsibility for health care can be efficiently applied when rationing care is important from both an economic and an ethical standpoint. Rationing in nonfunctioning markets will inevitably lead to excessive sacrifices in well-being – and this is not merely an economic concern. A question that the article only briefly touches on is: When is it fair to assign responsibility for past behavior? Hence, I will pursue this discussion in the context of healthcare rationing.

Poor health and the need for health care is not something that appears like a “donkey from nowhere”.²⁴ Individuals often have opportunities to minimize the risk for future healthcare consumption by refraining from unhealthy behaviors and avoiding hazardous situations. It is increasingly obvious that individuals’ choices – healthy lifestyles – are perhaps the most significant factors for being healthy. Yet, the extent to which past behavior should (or should not) influence healthcare rationing is a morally intricate matter. On one hand, we could argue that it would be excessively harsh for society to abandon individuals in need because they knowingly behaved irresponsibly. On the other hand, it seems equally impermissible to dismiss individual responsibility altogether when setting priorities. Especially since one man’s freedom to behave irresponsibly becomes another man’s financial burden in a publicly funded healthcare system. Hence, one could argue that some form of reactive measure to hinder irresponsible use of collective resources should be in place.

A common line of thought when discussing the role of individual responsibility is that an individual should be held accountable for the foreseeable consequences that are within that individual’s control. Likewise, individuals should not be held accountable for unforeseeable consequences due to circumstances outside that individual’s control. This thinking aligns with the *luck egalitarian* view on distributive fairness, which broadly asserts that fairness requires an distribution of resources that reflects individuals’ degree of responsibility (see e.g. Dworkin 1981; Arneson 1989). When discussing individual responsibility in the context of health it is difficult to ignore the fact that most unhealthy behaviors, e.g. smoking and unhealthy diets, are disproportionately common among individuals in the lower socioeconomic segments of society. Social circumstances present factors that influence certain individuals to engage in unhealthy living. Moreover, it seems unreasonable to penalize an individual for being born into a particular socioeconomic group since this circumstance is clearly beyond the individual’s control.

Even if we disregard the influence of social circumstances, other prerequisites need to be satisfied before it would seem reasonable to claim that someone is responsible for their ill health due to irresponsible past behavior. Specifically, four crucial components need to be present before any form of *retrospective responsibility* can be justified. First, the causal association between a person’s

²⁴ This is an old south Swedish expression that simply means appearing for no obvious reasons.

behavior and the state of ill health must be sufficiently clear. Second, the behavior must be conscious and voluntary. Third, the state of ill health must be the result of an unnecessary risk. Fourth, the individual needs to be fully aware of society's reactive attitudes toward certain types of conscious risk taking. Since all of these prerequisites are hard to observe in practice it seems impermissible that a retrospective notion of responsibility can be given more than a symbolic role in actual healthcare rationing.²⁵

However, while stating that I think retrospective responsibility should play a peripheral role in healthcare rationing, I still think it is important, for motivational purposes, that society to acknowledge that individuals have a responsibility for their own health. But for this to have any positive consequences it would be more appropriate to assign responsibility prospectively. This can be done through various means, e.g. taxing commodities associated with unhealthy activities or requiring compulsory insurance for certain types of risky activities. *Prospective responsibility* of this kind would be consistent with giving weight to individual responsibility when rationing care at the *desktop level*, but not at the *bedside level*. I would argue that avoiding assessment of individual responsibility at the bedside level is a desirable policy feature. If doctors were to assess not only patients' needs, but also their responsibility for ill health, it could undermine the important aspect of *trust* in the doctor-patient relationship.

A potential disadvantage of applying prospective responsibility, e.g. through taxation, is the fact that ill health is more common among the poor. Tax on cigarettes, for instance, is highly regressive in the sense that it shifts resources from the relatively poor smoker to the relatively affluent nonsmoker. Consequently, the positive effect of less unhealthy behavior needs to be weighed against the increased economic inequality that this type of prospective responsibility is likely to cause. In addition, policy-makers in public health, who design policies related to prospective responsibility, need to carefully consider other matters where reasonable disagreement is likely to arise: What constitutes justifiable grounds for overriding individuals' fully informed preferences when promoting healthy behavior? When do health policies become unjustifiable forms of paternalism?

²⁵ See the chapter by Liss in report published by National center for priority setting in health care (2008) for a more detailed discussion of the necessary prerequisites for applying retrospective responsibility for health as criteria in rationing health care.

Paternalism – Why Health is not Always Good

Paper II, *Discounting, preferences and paternalism in cost-effectiveness analysis*, deals with paternalism in the context of intertemporal rationing decisions. The article highlights two cases where it is justifiable to override individuals' fully informed time preferences relative to making rationing decisions over time. Preference failures (as opposed to market failures) that justify paternalistic intervention are categorized into two distinct groups; myopic and acratic. Myopic preference failures refer to a cognitive inability to compute information adequately in order to make choices that result in desirable outcomes. Acratic preference failures represent a weakness of will, which reduces the ability of individuals to act in accordance with their second-order preferences. Since the article focuses on paternalism related to time preferences, it touches only briefly on the crucial issue of whether improved health is always something that society should strive to achieve. I address the paternalism aspects of this discussion below.

Paternalism, in the present context, can be generally viewed as state interference with a person's liberty and is motivated by a claim that the affected person will be better off or protected from harm (Dworkin 2010). A fundamental assumption in welfare economics is that each individual is the best judge of how to maximize their own welfare. This is a questionable starting point when it comes to designing health policy, since extensive market failures are associated with the distribution of health care. To promote desirable outcomes that increase the overall good in society, at times it is necessary to override individuals' preferences by limiting or changing the array of choices available to them – paternalistic intervention.

In economics, preferences based on insufficient information or inadequate facts are generally viewed as market failures where paternalistic interventions are justified to avoid inefficient outcomes. However, it becomes more complicated to justify paternalistic interventions in cases where individuals have sufficient and adequate facts, but still make choices that negatively affect their health. Consequently, two issues have specific importance when discussing justifiable forms of paternalism in public health interventions. First, to what extent do individuals understand the potential health risk posed by unhealthy living? Second, if risks are adequately perceived, to what extent do individuals actually take them into account when making decisions?

Paternalistic interventions to promote health are motivated mostly by stating that individuals tend to have an incorrect perception about the risk associated with certain unhealthy behaviors. This motivation is dissatisfying. By now, nearly everyone in wealthy countries, e.g. Sweden and the US, should have a good understanding that it is bad for your health to smoke, drink, and follow an unhealthy diet. For example, cigarette packages carry messages in large capital letters informing the user that “SMOKING WILL KILL YOU” or “SMOKING SERIOUSLY DAMAGES YOUR HEALTH”. Further, Viscusi (1990,1992) has empirically shown that an incorrect perception of health risks is not a valid explanation for why individuals engage in more risky behavior. The health risks associated with smoking are not only well known, but are in fact generally overestimated, even among those who smoke. Thus, the motivation to incorrectly perceive health risks seems to be an (empirically) weak argument for why society should override individual preferences in its quest to promote health.

A stronger argument, I believe, for why society should act paternalistically and promote health is the claim that health is of special moral importance since it protects the range of *opportunities* open to individuals (e.g. Daniels 2008). Given that society has an obligation to protect opportunities, it also has an obligation to promote and protect individuals’ health even if this goes against individual preferences. However, I would argue that a related question has received insufficient attention in the literature on health and health care: What good are individual opportunities in the absence of corresponding *possibilities* to actually reap the benefits of having a wider range of opportunities?

Opportunities stem from individuals’ mental and physical abilities. Increased health and knowledge are two aspects that typically elevate an individual’s outset and opportunities in life. Possibilities stem from the potential to use these mental and physical abilities. Having a mentally challenging job, or the opportunity to travel and experience are typical examples of such possibilities. To exemplify, learning to float in water creates an opportunity to swim, while a lake or a swimming pool provides the possibility to swim. Or in the case of mental abilities, learning to speak Swahili creates an opportunity to communicate with people from East Africa, while meeting people from East Africa provides the possibility to speak Swahili.

Society offers a wide range of possibilities to people who are wealthy. They can ski, take a year off work, and live an active life after retirement, or they can keep on working just for fun if they like, and so on. But being able to make use of all these available *possibilities* often requires good health. Therefore, it is perfectly rational for wealthy people to promote healthy living for themselves. For the less wealthy, however, a much more limited range of possibilities is offered, since living an active life is often intensely resource consuming. Hence, for less-wealthy individuals it might very well be burdensome and irrational to increase the range of opportunities through healthy living if they have no corresponding possibilities. Instead, an increased level of health (opportunities) could possibly become a source of distress for less-wealthy individuals unable to live life to its fullest. For individuals to promote their own health when they are unable to take advantage of the opportunities that health creates could therefore decrease their general well-being.

Following this reasoning, health is of little value to individuals if it only creates opportunities they are unable to take advantage of. Both wealth and health create opportunities/possibilities for higher levels of well-being. However, I would argue that neither is sufficient for a higher level of well-being if they are isolated from each other. Instead, I believe that a steady-state relationship must exist between the two, and it needs to be maintained to increase well-being. Consequently, too much emphasis on either one could potentially have negative effects on well-being unless corresponding investments are made in the other. This needs to be taken into account before embarking on paternalistic policies to improve health – policies that often target the lower socioeconomic segment.

Incentives – Persuasion or Coercion?

Paper III, *Incentivizing deceased organ donation: A Swedish priority-setting perspective*, examines whether a diverse set of incentives (both financial and nonfinancial incentives) could combine considerations of efficiency and fairness when trying to increase the stock of organs available for transplantation. The article identifies three potential pitfalls where the use of incentives in health policy may become objectionable; *coercion*, *commodification*, and *loss of public trust*. The fundamental question that I pursue here is: When does coercion become so ethically objectionable that it hinders society from maximizing overall welfare?

Coercion occurs when individuals, due to some form of threat or external pressure, feel forced to act in a way that is not in line with their own best interests (see e.g. Anderson 2008). The common sense opposition to coercion is that it is much easier to justify, for example, extracting organs from individuals if this is done voluntarily, compared to confiscation of organs coercively (Eyal 2008). Nonetheless, few would argue to categorically forbid all forms of coercive incentives, since a certain level of coercion seems to be justifiable if the consequences are ‘good enough’.

From a utilitarian standpoint, coercive acts can be justified if overall utility in society increases. Intuitively this seems like a rather weak rule for justification – as Eyal (2008:109) points out, “It is inappropriate to calculate the likely impact on net utility before deciding not to rape someone”. Consequently, the utilitarian framework seems to yield intuitively unfair outcomes when applied as a framework for justifying coercive policies.²⁶

A common argument in deontology-inspired philosophy is that coercion harms dignity in the sense that it is intrusive to individual autonomy (e.g. Nozick 1974). Therefore, society has an obligation to protect and promote autonomy since this is usually considered an important component of a good life. However, this may come at the expense of a decreased level of overall good in society. It seems reasonable to argue that it is permissible for society to sacrifice some level of autonomy in order to increase the overall good in society, at least as long as the autonomy level does not fall below a certain minimum threshold for each individual. Autonomy is important, but it is not

²⁶ For a more detailed discussion of coercion and utilitarianism see, e.g. Eyal (2008).

the only important aspect of a good life. It needs to be balanced against other conflicting values, e.g. well-being, to achieve satisfactory outcomes.²⁷ Moreover, it could be argued that the main purpose of incentivizing desirable health behavior is to increase the net sum of autonomy in the long run, which could also help justify some short-term harm to dignity through coercion.

When discussing the use of incentives in health policy it is useful to distinguish between *positive incentives* (carrots) and *negative incentives* (sticks). Positive incentives focus on rewarding desirable behavior, while negative incentives focus on penalizing undesirable behaviors. In the preceding discussion on the role of individual responsibility in healthcare rationing I have already indirectly addressed the role of negative incentives in health policy. In addressing incentives, I will focus on the point where positive incentives in health policy shift from being persuasive to being coercive.

When offering positive incentives to individuals, e.g. monetary compensation for organ donation, the individual is faced with a persuasive or coercive offer. The general idea is that when individuals find themselves in a situation where even a very unfair and unattractive offer is the best available for the individual, it could be argued that the created incentive is coercive and hence morally impermissible (e.g. Feinberg 1986 and Van De Veer 1979). Moreover, positive incentives could also be regarded as a form of bribe conflicting with pro-social behavior and altruistic motivation. Monetary compensation for organs or blood is an example of the incentives that are typically viewed as being in conflict with altruistic motivation. Hence, this type of material incentive has a tendency to crowd out altruistic motives. Even if the compensation is high enough to compensate for this crowding-out of pro-social behavior we might end up with a less compassionate society. This could be considered an undesirable outcome associated with excessive use of positive incentives in health policy.

An ongoing trend in many healthcare systems is the use of positive incentives to encourage individuals to adopt desirable health behaviors, e.g. physical exercise and participation in screening programs. As Voigt (2010) points out, the various strategies to implement positive incentives in health promotion may also have a substantial impact on inequalities within the healthcare system. They may indeed have the potential to efficiently contribute to better

²⁷ For a more extensive discussion about how to balance autonomy and well-being when they are in conflict see, e.g. Brock (1988).

health among the population. Wrongly implemented, however, they could increase inequality if the individuals likely to benefit from the incentives are those who are most well off. A way to avoid this might be to focus on participation-incentive schemes tailored to the abilities and needs of the disproportionately unhealthy groups – avoiding positive incentives associated with attainment of specific health goals – since these are likely to primarily benefit healthy individuals (Voigt 2010).

When discussing the positive incentives of monetary compensation associated with health, I think that an unmentioned psychological aspect should also be addressed in this context, namely *repugnance*. This is an emotional state of extreme dislike, manifesting itself in feelings of disgust that cannot be easily explained through reasons. Roth (2007) notes that transactions that are not considered repugnant, such as in-kind exchanges, frequently become repugnant when money is added. Roth (2007:44) exemplifies: “...loans themselves and adoption and love are widely regarded as good things when given freely, even when their commercial counterparts are regarded in a negative way”. Some argue that morality is not a matter of reasoning, but rather having a gut feeling about what is right and wrong.²⁸ Whether or not this type of “disgust-based” morality has any real philosophical merits is a very complex question. On one side it seems questionable since repugnance has been used throughout history by individuals in governing roles to justify various forms of persecution, e.g. anti-Semitism and homophobia (see e.g. Nussbaum 2004). Nevertheless, for policy-makers who want to design efficient incentive schemes the existence of repugnance could be a very important aspect to consider. Public policies that run counter to prevailing moral intuitions are likely to undermine the legitimacy of the policymaker unless they are supported by powerful moral arguments or general principles.

Inequality – When are Inequalities Unfair?

Paper IV, *Horizontal inequality in rationing by waiting lists*, investigates whether the publicly funded Swedish healthcare system has horizontal inequalities in access to care. The article documents that access is unequal (but only to a limited extent) when waiting lists are used to ration care. However, the article says little about the normative question related to inequality and healthcare

²⁸ See, e.g. Burke (1894).

rationing: When are inequalities unjust, and when does inequality become inequity?

To avoid an incomprehensible discussion about inequality, it is important to distinguish between inequality as regards *health* and inequality as regards *health care*. The main focus of this thesis has been to discuss some of the reasonable disagreements concerning the rationing of health care. Although I will refrain from an in-depth discussion concerning the point at which health inequalities become unjust, it is impossible to fully avoid the issue. Inequality/inequity in health care is most often framed in terms of equal access for equal need. Consequently, *need* is often used in specifying unjust inequalities. As illustrated in previous sections, interpreting and applying the concept of need is perhaps the most recognized area of reasonable disagreement in healthcare rationing.

Somewhat less complicated, but equally important for understanding when healthcare inequalities become inequities, is the concept of equal access. Goddard et al (2001) present a valuable framework comprised of four key factors for assessing equal access to health care:

- i. Availability: differences may exist in the sense that certain healthcare services might not be available to some population groups, or that clinicians may have different propensities to offer treatment to patients with “identical” needs from different population groups;
- ii. Quality: differences in the level of quality of certain health services offered to “identical” patients may vary between population groups;
- iii. Costs: differences in costs that a health service imposes on identical patients may vary between population groups;
- iv. Information: differences in health services may fail to ensure that all population groups are informed with equal clarity about the availability of certain services.

In offering a broad – but in the field of public health, influential (see e.g. Braveman et al. 2001) answer to the question of when inequalities become inequities, Whitehead (1992) asserts that inequalities count as inequities if they are *avoidable*, *unnecessary*,²⁹ and *unfair*. Following this definition, not all avoidable inequalities are unfair, nor are all unfair inequalities avoidable. Applied to health care this could imply, e.g. that unequal availability to care

²⁹ I will however refrain from using this concept since I interpret it to be fully incorporated within the other two concepts when applied to inequalities.

between population groups, which is avoidable in practice, is not automatically unfair *per se*. Likewise, unfair inequalities in access to care are not always avoidable. For example, an unavoidable fact is that not everyone can live equally close to a university hospital with full medical facilities. According to Whitehead's account of inequity, neither of these types of inequality constitutes inequity. Consequently, it is not sufficient for inequalities to be deemed unfair or avoidable on their own; for inequities to arise both prerequisites need to be fulfilled.

If we assume that all existing inequalities are unavoidable, we still need to apply a notion of fairness to make sense of when inequality becomes inequity. Rawls' *Theory of Justice* (1971), which includes the *difference principle*, has been the most widely discussed theory for deciding which of the last 40 years' inequalities are unjust. The difference principle allows inequalities in the distribution of goods, but only in so far as these inequalities benefit the worst-off individuals in society. According to the difference principle, inequalities are permitted only as long as they do not compromise basic rights and liberties. Rawls' theory of justice is an example of the currently dominating approach toward fairness that Sen (2009) calls the *arrangement-focused approach*. This approach strives toward a perfectly just or fair society by identifying institutions and social arrangements that can achieve this. The idea of finding a flawless theory of fairness that can actually be implemented in practice is perhaps unrealistic. For instance, the concept of *reasonable people* plays a central role in most arrangement-focused approaches,³⁰ but who are these reasonable people? Where do they hide? Moreover, as Williams comments (1997:128):

There is a regrettable tendency for equity arguments to be conducted within a rhetorical framework in which it appears to be possible to "do good" at no opportunity cost whatever. It generates a great deal of righteous self-satisfaction for the romantic escapists, and it puts economists back in the role of the dismal scientists always stressing the sacrifices, but it does not help the hard-pressed decision-makers who grapple with the issues in real life every day.

From a health policy perspective it seems farfetched that decision-makers should turn to Rawls or other arrangement-focused theories of justice when assessing inequalities relating to health care. Moreover, it does not seem like a

³⁰ A concept that both Rawls (1971) and later Daniels (2008) frequently use in their theories.

reasonable objective for policy-makers to strive toward finding arrangements that will secure a perfectly fair healthcare system. A more reasonable objective would be to focus on what Sen (2009) calls the *realization-focused approach* toward justice. This approach implies that instead of trying to achieve a perfectly just system, decision-makers should concentrate on removing clear cases of injustice. Clear cases of injustice are unfortunately common in the real world, and new cases of injustice continuously arise. Although a perfectly just healthcare system is an admirable ambition, starting out by making things better is probably a more realistic way forward.

FINAL REMARKS

The interdisciplinary approach

Combining economics and ethics at the population level is not a new invention. Some level of cross-disciplinary discussion has always accompanied issues such as discounting, equity, and paternalism. Unfortunately, economics and ethics are commonly viewed as mutually exclusive disciplines, and only rarely are the two explicitly merged. This is remarkable since the defining task of both disciplines is to find satisfactory indications for the optimal distribution of scarce resources.

A crucial strength of economics as a normative science is that it recognizes that, ultimately, the conclusions it produces must be possible to implement in the “real world”. Consequently, economics tries to take the step from abstract theory to practical action, which is something that ethics at the population level seldom tries to accomplish. However, economics could become more useful in making policy decisions if greater attention was paid to the moral assumptions upon which it is based. A comparative strength of ethics at the population level is that it provides a more nuanced approach for making judgments on the distribution of scarce resources. Ethics could, on the other hand, become an even more useful tool for policy-makers if its conclusions were designed to be more in touch with the “real world”. Both disciplines have much to gain through greater cross-disciplinary dialogue, but for this to be productive, economists need to acknowledge the flaws in the normative basis of economics, and ethicists need to acknowledge that any conclusions that follow from their reasoning should be less detached from real-life situations.

In healthcare rationing it seems as if economists and ethicists alike lack the ability to strike a chord with public opinion. Arguably, this is because economists have traditionally targeted the rational, utility-maximizing individual, while ethicists typically target the reasonable, morally-thinking individual. Quite frankly, however, individuals tend to be neither rational nor reasonable when it comes to healthcare rationing. The emerging trend toward greater interest in behavioral economics is positive in this respect. A better understanding of norms and actual behavior will hopefully serve as a natural

link between ethics and economics and will help both disciplines stay in touch with reality. Another emerging, albeit troubling, trend is that health economics is rapidly becoming a discipline that mostly incorporates technically skilled statisticians and mathematicians with little background or interest in normative science.

The art of saying no

Healthcare rationing is not for the weak-minded. For most people, telling someone that they cannot have what they want is stressful even in the “ordinary” world. In the healthcare context, where wants become needs, this becomes even more stressful. Moreover, saying no in the context of health care is often frowned upon by the general public, adding to the emotional burden of saying no. The public’s dislike is often fueled by scare tactics employed by the media, stakeholders, and politicians competing for office. As a result, politicians who say no tend not to get reelected, and medical personnel usually gain nothing but trouble by saying no. So, who in their right mind would ever say no in the context of health care?

I have no answer to this question, but I desperately would like one. Neither economics nor ethics provide any satisfactory answers. Future research within the field of behavioral economics and neuroeconomics will hopefully shed further light on the matter. Even if this future research is promising, it does not mean that we can simply answer moral questions by appealing to what people generally think or do. We must continually pursue answers to understand why we regard certain actions to be right or wrong, and expose these reasons to public scrutiny.

The case of explicit rationing undertaken in Sweden and in the State of Oregon elucidates two different approaches toward explicit healthcare rationing. They highlight some of the available policy options for explicitly dealing with healthcare rationing. More importantly they reveal valuable lessons on how the public might perceive these policy options. What the Oregon experience shows is that simple decision rules are unlikely to be a viable way forward in rationing health services. The so-called QALY approach has merits, but it should not be portrayed as “the truth”. It is the outcome of a particular normative standpoint that needs to be balanced against non-utilitarian moral values. What the Swedish example shows is that an ambition to increase openness toward core ethical values might very well result in reduced

transparency. This is arguably what has happened in Sweden, where the ethical principles have a strong legal status, although their meaning and application has remained unclear and insufficiently specified. What the Swedish experience also shows is that it is possible to foster greater acceptance for the fact that rationing is an inevitable feature of any healthcare system. Decision-makers should look beyond the immediate reactions of stakeholders and the media when assessing the public's reaction to rationing decisions.

I see one simple solution to the reasonable disagreements brought forward in this thesis, and that is to acknowledge that there are no simple solutions. Given the complexity of the issues, it would be self-defeating and unrealistic to strive toward consensus agreements on all issues at hand. Instead, the most promising way forward is to be explicit about reasonable disagreements. The second step would be to create institutions and procedures where these reasonable disagreements could be meaningfully addressed. For such institutions and procedures to be productive it will be necessary to foster public willingness to look beyond conflict and accept reasonable disagreement.

Final remarks

ACKNOWLEDGEMENTS

I am thankful to many people who have supported and encouraged me over the years. I especially like to express my sincere gratitude to the following:

Per Carlsson, my supervisor, for always supporting me and being a genuinely nice person. I know that I have not always been easily managed but you have allowed me to continuously grow as a researcher and pursue new ideas.

Carl Hampus Lyttkens, my co-supervisor, for invaluable guidance and support. I have always felt secure having you on my "team".

Petter Tinghög, my brother and secret supervisor, for writing my job application to become a PhD student in Linköping. I wish everyone could have a big brother like you. Your input and support during the final hectic stages of writing this thesis have been invaluable.

Magnus Johannesson, my academic role model, for introducing me to neuroeconomics and providing me with the opportunity to conduct exciting research in the future.

Thor-Henry Brodtkorb, my dear friend and collaborator, for endless discussions which are always both fruitful and fun. We still have a lot of exciting project to pursue in the future. Your friendship and input during the years have been greatly appreciated.

Emelie Heintz, my dear friend and colleague, for introducing me to the word diabetic rephropathy but perhaps more importantly for being a truly awesome friend.

Almina Kalkan and **David Andersson**, my dear younger colleagues, for being the next generation.

Lena Hector, for being the nicest of all nice persons and for always being helpful and looking after the weird administrative rules of academia which I tend to neglect.

Faisal Omar, my collaborator and flatmate, for always helping out and making Linköping feel more like home. It is my intention to be equally supportive when you soon are to wrap up your thesis.

All my friends and colleagues at the **Division of Health Care Analysis** and the **National Center for Priority Setting in Health Care** for making my time as a PhD student pleasant and fun. A special thank you to: **Lars Bernfort**, **Thomas Davidson**, **Nathalie Eckard** and **Ann-Charlotte Nedlund** for giving me valuable input on earlier drafts of this thesis. Special thanks are owed to **Jenny Alwin** for giving me valuable pointers on how to obey the administrative rules of Linköping University

My one year stay at the **Program in Ethics and Health at Harvard University** was a truly enlightening experience. Therefore I wish to thank all of those who made my stay there such an eye opening experience and crucially influenced the scope of this thesis. I especially like to thank **Greg Bognar**, **Dan Brock**, **Alena Buyx**, **Norman Daniels**, **Nir Eyal**, **Elisabeth Fenton**, **Adriane Gelpi**, **Paul Kelleher**, **Keren Ladin**, **Andras Miklos**, **Kristin Voigt**, and **Dan Wikler**.

The county council of Östergötland, for making this thesis possible by their financial support.

The Brocher Foundation, for supplying me with superb research facilities and a creative surrounding when writing parts of this thesis.

IFK SPADEN, my floorhockey team, for being such an inspiring bunch of people. I especially like to thank **Johan Fehrman** and **Ingemar Lundin** for always remaining my friends.

Marit Furn who put much talent, creativity and energy into designing the cover of this book.

Gabriella Tinghög Olofsson, for always looking after her little brother.

Bryan Adams, **Veronica Kloo**, **Nisse Hellberg**, **Old 97's** and **Kula Shaker** for bringing the music that inspired me to write this thesis.

Academic input aside the most profound thank you goes out to my parents **Mats** and **Boel Tinghög**, for making me such an unjustifiably privileged person. You have always supported me in all sorts of things, large and small, and I am deeply grateful for this.

REFERENCES

- Alakeson, V. 2008. "Why Oregon went wrong." *BMJ* 337: 900-901.
- Anderson, S. 2008. Coercion. *The Stanford Encyclopedia of Philosophy* (Fall 2008 Edition) E. N. Zalta(ed.), Available at:
<http://plato.stanford.edu/archives/fall2008/entries/coercion/>
- Arneson, R. J. 1989. "Equality and Equal Opportunity for Welfare." *Philosophical Studies* 56: 77-93.
- Bénabou, R. and J. Tirole 2006. "Incentives and prosocial behaviour." *American Economic Review* 96: 1652-1678.
- Bentham, J. 1789. *An introduction to the principles of morals and legislation*. New York, Dover Publications.
- Bleichrodt, H., P. Wakker and M. Johannesson 1997. "Characterizing QALYs by Risk Neutrality." *Journal of Risk and Uncertainty* 15: 107-114.
- Braveman, P., B. Starfield and H. J. Geiger 2001. "World Health Report 2000: how it removes equity from the agenda for public health monitoring and policy." *BMJ* 323: 678-81.
- Brazier, J., J. Ratcliffe, J. A. Salomon and A. Tsuchiya 2007. *Measuring and Valuing Health Benefits for Economic Evaluation* New York, Oxford University Press.
- Brock, D. W. 1973. "Recent work in utilitarianism." *American Philosophical Quarterly* 10: 241-276.
- Brock, D. W. 1988. "Paternalism and autonomy." *Ethics* 98: 550-565.
- Brock, D. W. 2003. Ethical issues in the use of cost-effectiveness analysis for the prioritization of health care resources, In *Making choices in health: WHO guide to cost-effectiveness analysis*. T. Tan-Torres Edejer, R. Balthussen, (eds). Geneva, World Health Organization.
- Broome, J. 1993. "Qalys." *Journal of Public Economics* 50: 149-167.
- Burke, E. 1894. *The works of the right honourable Edmund Burke*. Boston, Little, Brown.
- Bäckman, K., E. Karlsson and P. Carlsson 2006. *Öppna prioriteringar i Östergötland. Del 3*. Linköping, PrioriteringsCentrum.
- Bäckman, K., K. Lindroth and P. Carlsson 2005. *Öppna prioriteringar i Östergötland. Del 2*. Linköping, PrioriteringsCentrum.
- Carlsson, P., C. Kärvinge et al 2007. *National model for transparent vertical prioritisation in Swedish health care*. Linköping, PrioriteringsCentrum.

- Carlsson, P. 2010. "Priority setting in health care: Swedish efforts and experiences." *Scandinavian Journal of Public Health* 38: 561-564.
- Carlsson, P. Kärvinge, C. et al 2007. *National model for transparent vertical prioritisation in Swedish health care*. Linköping, PrioriteringsCentrum.
- Cookson, R. and P. Dolan 2000. "Principles of justice in health care rationing." *Journal of Medical Ethics* 26: 323-9.
- Cohen, J. 1994. "A more democratic liberalism." *Michigan Law Review* 92: 1506-1543.
- Culyer, A. J. 1998. How ought health economists to treat value judgements in their analyses. In *Health, health care and health economics: perspectives on distribution*. M. L. Barer, T. E. Getzen and G. L. Stoddart. Chichester, John Wiley.
- Daniels, N. 1994. "Four unsolved rationing problems. A challenge." *Hastings Center Report* 24: 27-29.
- Daniels, N. 2000. "Accountability for reasonableness." *BMJ* 321: 1300-1301.
- Daniels, N. 2006. "Equity and population health: toward a broader bioethics agenda." *Hastings Center Report* 36: 22-35
- Daniels, N. 2008. *Just health: meeting health needs fairly*. Cambridge, Cambridge University Press.
- Daniels, N. and J. E. Sabin 2008. *Setting limits fairly: Learning to share resources for health*. Oxford, Oxford University Press.
- Donaldson, C. and K. Gerard 2005. *Economics of health care financing: the visible hand*. Basingstoke, Palgrave Macmillan.
- Dworkin, G. (2010). Paternalism. In *The Stanford Encyclopedia of Philosophy* (Summer 2010 Edition). E. N. Zalta (Eds). Available at: <http://plato.stanford.edu/archives/sum2010/entries/paternalism>
- Dworkin, R. 1981. "What is equality? Part 2: equality of resources." *Philosophy and Public Affairs* 10: 283-335.
- Ellingsen, T. and M. Johannesson 2007. "Pride and prejudice: the human side of incentive theory." *American Economic Review* 98: 990-1008.
- Erntoft, S. (2010). *The use of health economic evaluations in pharmaceutical priority setting: the case of Sweden*. Lund, Lund Business Press.
- Evans, J. G. 1997. "Rationing health care by age: the case against." *BMJ* 314: 822-825.
- Eyal, N. 2008. "Utilitarianism and coercion." *Politeia* 90: 108-123.
- Feinberg, J. 1986. *Harms to self*. New York, Oxford University Press.
- Frankfurt, H. G. 1987. "Equality as a moral ideal." *Ethics* 98: 21-43.

- Frey, B. S. and F. Oberholzer-Gee 1997. "The cost of price incentives: An empirical analysis of motivation crowding-out " *American Economic Review* 87(4): 746-755.
- Frey, B. S. and W. W. Pommerehne 1993. "On the fairness of pricing-an empirical survey among the general population." *Journal of Economic Behavior and Organization* 20: 295-307.
- Friedman, A. 2008. "Beyond accountability for reasonableness." *Bioethics* 22: 101-12.
- Gneezy, U. and A. Rustichini 2000. "Pay enough or don't pay at all." *Quarterly Journal of Economics* 115: 791-810.
- Goddard, M. and P. Smith 2001. "Equity of access to health care services: theory and evidence from the UK." *Social Science and Medicine* 53(9): 1149-1162.
- Goodin, R. 1993. "Democracy, preferences and paternalism." *Policy Sciences* 26: 229-247.
- Hadorn, D. 1996. "The Oregon priority-setting exercise: cost-effectiveness and the rule of rescue, revisited." *Medical Decision Making* 16: 117-119.
- Hadorn, D. 1991. "Setting health care priorities in Oregon. Cost-effectiveness meets the rule of rescue." *JAMA* 265: 2218-2225.
- Ham, C. 1998. "Retracing the Oregon trail: the experience of rationing and the Oregon health plan." *BMJ* 316: 1965-1969.
- Harbaugh, T. W. 1998. "What do donations buy? A model of philanthropy based on prestige and warm glow." *Journal of public economics* 67: 269-284.
- Harris, J. 1987. "QALYfying the value of life." *Journal of Medical Ethics* 13: 117-123.
- Hausman, D. M. 2010. "Hedonism and welfare economics." *Economics and Philosophy* 26: 321-344.
- Johannesson, M. 1994. "QALYs, HYE's and individual preferences--a graphical illustration." *Social Science and Medicine* 39(12): 1623-1632.
- Johannesson, M. 1996. *Theory and methods of economic evaluation of health care*. Dordrecht Kluwer.
- Kahneman, D. and R. Sugden 2005. "Experienced utility as a standard of policy evaluation." *Environmental and Resource Economics* 32: 161-181.
- Kahneman, D. and A. Tversky 1979. "Prospect theory: An analysis of decisions under risk." *Econometrica* 47: 313-327.
- Kass, L. R. 1997. "The wisdom of repugnance." *The New Republic* 2 June: 17-26.

- Klein, R., P. Day and S. Redmayne 1996. *Managing scarcity : priority setting and rationing in the National Health Service*. Buckingham ; Philadelphia, Open University Press.
- Le Grand, J. 2003. *Motivation, agency, and public policy: of knights and knaves, pawns and queens*. Oxford, Oxford University Press.
- Lipsey, R. G. and K. A. Chrystal 2007. *Economics*. Oxford ; New York, Oxford University Press.
- Liss, P.-E. 1993. *Health care need: meaning and measurement*. Aldershot, Avebury.
- Malmqvist, E. 2008. *Good parents, better babies: an argument about reproductive technologies, enhancement and ethics*. Linköping, Linköping University.
- Marmot, M. and R. G. Wilkinson 2005. *Social determinants of health*. Oxford ; New York, Oxford University Press.
- Marmot, M. and R. G. Wilkinson 2006. *Social determinants of health*. Oxford ; New York, Oxford University Press.
- McCloskey, D. N. 1998. *The rhetori of economics*. Madison, University of Wisconsin Press.
- McKie, J. and J. Richardson 2003. "The rule of rescue." *Social Science and Medicine* 56: 2407-2419.
- Mellström, C. and M. Johannesson 2008. "Crowding out in blood donation: was Titmuss right?" *Journal of the European Economic Association* 6: 845-863.
- Mill, J. S. 1861. *Utilitarianism*. Indianapolis, Bobbs-Merrill Educational Publishing.
- Ministry of health and social affairs 1993. *No easy choices: the difficult priorities of health care* Stockholm, Socialdepartementet.
- National center for priority setting in health care 2007. *Resolving health care's difficult choices : survey of priority setting in Sweden and an analysis of principles and guidelines on priorities in health care*. Linköping, PrioriteringsCentrum.
- National center for priority setting in health care 2008. *Resolving Health Care's Difficult Choices - Survey of Priority Setting in Sweden and an Analysis of Principles and Guidelines on Priorities in Health Care*. Linköping, Prioriteringscentrum.
- New, B. 1999. "Paternalism and public policy." *Economics and Philosophy* 15: 63-83.
- Nord, E. 1999. *Cost-value analysis in health care: making sense out of QALYs*. Cambridge, Cambridge University Press.

- Norheim, O. F. 1999. "Healthcare rationing-are additional criteria needed for assessing evidence based clinical practice guidelines?" *BMJ* 319: 1426-1429.
- Nozick, R. 1974. *Anarchy, state, and utopia*. Oxford, Blackwell.
- Nussbaum, M. C. 2000. *Women and human development: the capabilities approach*. Cambridge, Cambridge University Press.
- Nussbaum, M. C. 2004. "Danger to Human Dignity: The Revival of Disgust and Shame in the Law." *The Chronicle of Higher Education* Available at <http://chronicle.com/article/Danger-to-Human-Dignity-the/21047> (retrieved 2010-12-08).
- Oberlander, J., T. Marmor and L. Jacobs 2001. "Rationing medical care: rhetoric and reality in the Oregon Health Plan." *CMAJ* 164: 1583-1587.
- Olsen, J. A. 1997. "Theories of justice and their implications for priority setting in health care." *Journal of Health Economics* 16: 625-639.
- Pareto, V. 1906. *Manual of political economy*. London, Macmillan.
- Parfit, D. 1995. *Equality or Priority*. Laurence: University of Kansas, The Lindley Lectures.
- Parfit, D. 1997. "Equality and priority." *Ratio*, 10: 202-221.
- Pigou, A. C. 1912. *Wealth and Welfare*. London.
- Pliskin, J. S., D. S. Shepard and M. C. Weinstein 1980. "Utility functions for life years and health status." *Operations research* 28: 206-224.
- Ramsey, F. P. 1928. "A mathematical theory of saving." *The Economic Journal* 38: 543-549.
- Rawls, J. 1971. *A theory of Justice*. Cambridge, Harvard University Press.
- Rawls, J. 1989. "The domain of political and overlapping consensus." *New York University Law Review* 64: 233-255.
- Rawls, J. 1993. *Political liberalism*. New York, Columbia University Press.
- Rid, A. 2009. "Justice and procedure: how does accountability for reasonableness result in fair limit-setting decisions?" *Journal of Medical Ethics* 35: 12-16.
- Roemer, E. J. 1995. "Equality and responsibility." *Boston Review* 20: Available at <http://www.bostonreview.net/BR20.2/roemer.html>
- Roth, A. E. 2007. "Repugnance as a constraint on market." *The Journal of Economic Perspectives* 21: 37-58.
- Rutledge, K. M. 1997. "The Oregon Health Plan: lessons learned." *Healthcare Financial Management* 51: 48-51.
- Sabik, L. M. and R. K. Lie 2008. "Principles versus procedures in making health care coverage decisions: addressing inevitable conflicts." *Theoretical Medicine and Bioethics* 29: 73-85.

- Sen, A. 1987. *On ethics and economics*. Oxford, Basil Blackwell.
- Sen, A. K. 1979. *Collective choice and social welfare*. Amsterdam, North-Holland.
- Sen, A. K. 1985. *Commodities and capabilities*. Amsterdam, North-Holland.
- Sen, A. K. 2009. *The idea of justice*. Cambridge, Harvard University Press.
- Shiell, A., J. Seymour, P. Hawe and S. Cameron 2000. "Are preferences over health states complete?" *Health Economics* 9: 47-55.
- Singer, P. 2005. "Ethics and intuitions." *The Journal of Ethics* 9: 331-352.
- Soetevent, A. R. 2005. "Anonymity in giving in a natural context—a field experiment in 30 churches." *Journal of public economics* 11-12: 2301-2323.
- Temkin, L. S. 1993. *Inequality*. New York, Oxford University Press.
- Temkin, L. S. 2003. "Equality, priority or what?" *Economics and Philosophy* 19: 61-87.
- Tinghög, G., P. Carlsson and C. H. Lyttkens 2010. "Individual responsibility for what? - a conceptual framework for exploring the suitability of private financing in a publicly funded health-care system." *Health Economics Policy and Law* 5: 201-223.
- Tsuchiya, A. 2000. "QALYs and ageism: philosophical theories and age weighting." *Health Economics* 9: 57-68.
- Tsuchiya, A. and P. Dolan 2005. "The QALY model and individual preferences for health states and health profiles over time: a systematic review of the literature." *Medical Decision Making* 25: 460-467.
- Ubel, P. A. and S. Goold 1997. "Recognizing bedside rationing: clear cases and tough calls." *Annals of Internal Medicine* 126: 74-80.
- Waldau, S., L. Lindholm and A. H. Wiechel 2010. "Priority setting in practice: participants opinions on vertical and horizontal priority setting for reallocation." *Health Policy* 96: 245-254.
- Van De Veer, D. 1979. "Coercion, seduction, and rights." *The Personalist* 58: 374-381.
- Whitehead, M. 1992. "The concepts and principles of equity and health." *International Journal of Health Services Research* 22: 429-445.
- Wikler, D. and D. W. Brock 2008. Population level bioethics: mapping out a new agenda. *Global bioethics: issues of conscience for the twenty-first century*. R. M. Green, A. Donovan and S. A. Jauss. New York, Oxford University Press.
- Williams, A. 1997. "Intergenerational equity: an exploration of the 'fair innings' argument." *Health Economics* 6: 117-132.
- Viscusi, K. 1990. "Do smokers underestimate risks?" *Journal of Political Economy* 98: 1253-1270.

- Viscusi, W. K. 1992. *Smoking: making the risky decision*. New York, Oxford University Press.
- Voigt, K. 2010. "Incentives, health promotion and equality." *Health Economics Policy and Law*. Published online: 21 September 2010. DOI:10.1017/S174433110000277.
- Von Neumann, J. and O. Morgenstern 1947. *Theory of games and economic behavior*. Princeton, N.J., Princeton University Press.

References

PAPER I - IV

PAPER

I

Individual responsibility for what? – A conceptual framework for exploring the suitability of private financing in a publicly funded health-care system

GUSTAV TINGHÖG*

Center for Medical Technology Assessment, Department of Medical and Health Sciences, Linköping University, Linköping, Sweden

PER CARLSSON

The National Center for Priority Setting in Health Care, Linköping, Sweden

CARL H. LYTTKENS

Department of Economics, School of Economics and Management, Lund University, Lund, Sweden

Abstract: Policymakers in publicly funded health-care systems are frequently required to make intricate decisions on which health-care services to include or exclude from the basic health-care package. Although it seems likely that the concept of individual responsibility is an essential feature of such decisions, it is rarely explicitly articulated or evaluated in health policy. This paper presents a tentative conceptual framework for exploring when health-care services contain characteristics that facilitate individual responsibility through private financing. Six attributes for exploring the suitability of private financing for specific health-care commodities are identified: (i) it should enable individuals to value the need and quality both before and after utilization; (ii) it should be targeted toward individuals with a reasonable level of individual autonomy; (iii) it should be associated with low levels of positive externalities; (iv) it should be associated with a demand sufficient to generate a private market; (v) it should be associated with payments affordable for most individuals; and finally, (vi) it should be associated with ‘lifestyle enhancements’ rather than ‘medical necessities’. The tentative framework enables exploration of individual responsibility connected to health care as a heterogeneous group of commodities, and allows policymakers to make decisions on rationing by design rather than default.

Introduction

A fundamental objective of tax-funded health-care systems is to grant universal access to public health-care services and make it affordable at the point of use.

*Correspondence to: Gustav Tinghög, Department of Medical and Health Sciences, Linköping University, SE-581 83 Linköping, Sweden. Email: gustav.tinghog@ihs.liu.se

This implies not only a public obligation to improve health, but also to protect people from the financial costs associated with treating poor health. However, the scarcity of public funds inevitably makes it impossible to fund all services that are medically feasible. Consequently, policymakers within a system of essentially publicly funded health care are frequently faced with intricate decisions about which health-care services to exclude and thus require individuals to finance privately. Such health-care rationing can be done in two ways: either by excluding services that were previously reimbursed or by denying reimbursement for new medical services. Moreover, public spending in areas where individuals are capable of taking financial responsibility inescapably diverts resources from those areas where individuals are unable to take such responsibility. Thus, an important aspect to consider when trying to achieve a fair health-care distribution is which characteristics of services might actually facilitate individual financial responsibility for their own health care. Given the increasing public demand for transparency in health-care rationing, welfare societies face the challenge of explicitly articulating what role individual responsibility should play in a public health-care system – the key issue explored in this paper.

Within economics, the standard approach for determining what health care to include or exclude from public funding has been cost-effectiveness analysis (CEA), in which health-care services are ranked by the ratio of their costs to their health effects. Those services with a ratio below the threshold of what society is willing to pay for a unit of health gain, are recommended for exclusion from the publicly funded health-care package. However, even though CEA is a powerful tool for input of information into the priority-setting process, the approach is nevertheless subject to several limitations that constrain its practical use as a guiding principle for rationing. First, cost-effectiveness studies are seldom homogeneous in terms of methodology, which sometimes makes comparisons between different settings potentially misleading (Drummond *et al.*, 1993). Second, there is no uniformly accepted standard practice for incorporating future unrelated costs or benefits that may accrue in other parts of society (Garber and Phelps, 1997, 2008; Meltzer, 1997). Third, the cost-effectiveness ratio gives no information regarding the magnitude of the total budget effect an inclusion or exclusion will have. Even if the methodological shortcomings of CEA are disregarded, it still fails to take into account a wide range of factors and constraints that may be of importance in the priority-setting process. Most importantly, it does not explicitly consider distributional effects and fails to take into account the fact that some services may or may not be available for purchase out-of-pocket if not covered publicly, with obvious consequences for individuals' possibility of taking responsibility for their own health care (Hoel, 2007). Consequently, the standard economic approach towards health-care rationing fails to take contextual factors into account, important for the feasibility of individual responsibility. This makes CEA insufficient as an exclusive tool for rationing decisions.

The potential role of individual responsibility in health care policies is not a new topic. However, most of the earlier work on individual responsibility in health care has centered on the moral relevance of responsibility with regard to risk factors for poor health (Minkler, 1999; Wikler, 2002; Olsen *et al.*, 2003; Cappelen and Norheim, 2006; Buyx, 2008), and thus mainly discusses individual responsibility for own health-related behavior. This discussion includes the debate about whether those who take bigger risk with their health should pay higher taxes. However, an issue of equal importance is whether the characteristics of a specific health-care service and the applicability or shortcomings of market forces surrounding the specific service are such that the individual reasonably could – or could not – be given responsibility for financing the service in question. The health care required by certain health states may sometimes be fully manageable by the individual. In this paper, we will disregard responsibility in terms of underlying factors for health-care need (i.e. health risks). Instead, we will explore responsibility in terms of the health-care services that individuals are capable of financing and managing without public involvement or subsidies – a conception of individual responsibility more relevant in the policymaking context (Le Grand, 2003).¹

Although the concept of individual responsibility is a necessary consequence of any substantial amount of freedom and therefore present in practically all areas of society, it is rarely articulated or used as a rationing principle in the health-care context. Many have resisted the idea of assigning individual responsibility an explicit role in health-care rationing, because of the risk of intentional use to legitimate inequalities in health. At the macro policy-level, the Netherlands is the only country we are aware of that has explicitly stated individual responsibility as a possible principle for health-care rationing (Government Committee on Choices in Health Care, 1992). Its proposed system for priority-setting is constructed as a funnel with four criteria aimed to distinguish services that should receive funding, from those that should not. The four criteria are: Is the treatment necessary? Is it effective? Is it cost effective? And – as a final gate-keeping criterion – whether or not the service can be made the responsibility of the individual. Based on this system, services such as *in vitro* fertilization, homeopathic medicines and dental care for adults were excluded from publicly funded health care. Even though the system initially received a lot of attention in public debate on health-care rationing and individual responsibility, its actual effect on rationing in the long run has remained limited. Two plausible explanations for this limited influence are: poor operationalization of the criteria and the fact that all criteria are intended to lead to a yes or no answer to funding, while a decision maker weighs the criteria against each other (Stolk *et al.*, 2002).

In Sweden, explicit priority-setting principles based on ethical core values have also been defined at the macro policy-level. These principles are supposed

1 See, for example, Barry (2005).

to guide local and regional authorities when constructing rationing policies. The principles are: First, the principle of human dignity, meaning that all individuals have equal rights regardless of personal characteristics and position in society; Second, the principle of need and solidarity, meaning that resources should be used in domains (or patients) where needs are considered to be largest; Third, the principle of cost-effectiveness, meaning that resources should be used in the most effective way without neglecting fundamental duties concerning the improvement of health and quality of life. Individual responsibility was explicitly dismissed as a legitimate principle for priority-setting at the macro level.² Nevertheless, there are still a wide range of health-care services in Sweden that are subjected to rationing which cannot fully be explained by any of the existing guiding principles. For example, dental care for adults, eye-glasses, some vaccinations, over-the-counter drugs and a fair share of assistive devices are all subject to extensive individual responsibility, in the sense that they are available if you want them and have the ability to pay for them. This seems to indicate that, although formally dismissed, the individual capacity of responsibility constitutes an important aspect, in addition to established principles, in decisions on rationing in the Swedish health-care setting. It also illustrates that it is seen as acceptable to allocate health care through market mechanisms under certain circumstances.³ But because the rationing of health-care services most often has taken the form of incremental initiatives rather than direct cutbacks, these circumstances have remained unclear.

At the meso policy-level, there have been no attempts, which we know of, to explicitly take individual responsibility into account. There has, however, been some attempt to explicitly make 'rationing-lists' that are less generic compared to the macro-level policies described above. The most prominent example was undertaken by the state of Oregon; in the early 1990s, Oregon used an explicit process for rationing based on the ranking condition-treatment pairs according to their cost-effectiveness ratio. The initial rankings resulting from this process seemed counterintuitive in some cases and invoked strong negative public reactions. Public consultations were used to adjust the initial ranking. For example, some cancer treatments and appendectomy were initially ranked lower than capping teeth (Hadorn, 1991). In Sweden, explicit rationing at the meso level was undertaken by the County Council of Östergötland, in 2003, which

2 The commission assigned to prepare the law briefly motivated their dismissal by stating that the relationship between behavior and disease is not generally understood, as genetic factors have importance for the onset of certain diseases and harmful behavior to health is usually acquired early in life (Swedish Parliamentary Priorities Commission, 1995). Thus this objection is not relevant to the issue concerning financial responsibility for health care which this paper aims to explore.

3 In addition, the attention of health-care policies is increasingly directed towards the interplay between private and public responsibility. See, for example, the controversy surrounding cost-sharing by patients in the UK National Health Service (National Health Service, 2009) or in Germany, where recently, policies have been implemented within the national insurance system not to fund health care related to certain 'lifestyle choices' (Schmidt, 2007).

developed a set of explicitly defined limitations on health-care provision. Examples of services that got excluded from public funding were: treatments for simple childhood conditions (e.g. head lice, obesity) would no longer be treated in pediatric clinics, and patients had to pay for a second hearing aid privately.

The aim of this paper is to present a tentative conceptual framework, to be used within a system of principally publicly funded health care, for exploring when health-care services exhibit characteristics that facilitate individual responsibility for private financing. Our emphasis here is not to dismantle public funding for health care but rather to move society in a welfare-promoting direction by exploring combinations of public and private financing that may improve overall welfare. The article is divided into five sections: The first defines the concept of individual responsibility that we will focus on. The second section outlines the rationale for exploring health care as an economic commodity, which in some cases may be suitable for private financing. The third section reviews the characteristics that distinguish health care in general from other economic commodities, which potentially could make allocation through market mechanisms less suitable. In the fourth section, we present a tentative conceptual framework, based on the preceding arguments, exploring when the characteristics of health-care services might facilitate individual responsibility for private financing. The last section discusses the potential use and limitation of the conceptual framework.

Defining individual responsibility in health vs health care

Although individual responsibility in general is a prominent topic in the public debate, a common understanding of how this concept may be applied to health and health care is lacking. This is understandable because individual responsibility is a complex concept that may be interpreted in various ways, depending on context.⁴

But in order to have a meaningful discussion concerning individual responsibility as a rationing criterion, we need to clearly distinguish the different conceptions of responsibility and isolate those that are relevant from a health-care policy perspective. To do so, we first need to address the question: responsibility in what domain? Fleurbaey (1995) proposes that individual responsibility in terms of health-care funding can be assigned to two separate domains – factors or outcomes. We will refer to these domains as responsibility for health (factors) and responsibility for health care (outcomes).

The first domain, ‘responsibility for health’, is assigned when individuals are held responsible for the evolution of factors that eventually may affect health outcomes. If applied as a rationing principle, individual responsibility for health would imply that society compensate health disadvantages beyond individual control while implicitly penalizing behavior that is deliberately unhealthy. In

4 See, for example, Hart (1968) for illustrative examples.

most general theories of justice, this control-based notion of responsibility is given a central position. Rawls (1971), for example, points out that any autonomous moral agent should take responsibility for their personal conception of a good life, implying that how people use their equally distributed rights and resources should be a matter of individuals' responsibility. The luck egalitarian approach towards fairness and responsibility centers around the idea that individuals should be held accountable only for the consequences of their own deliberate choices, while the public should compensate individuals suffering as a consequence of brute luck,⁵ that is, factors outside one's own control (Arneson, 1989; Cohen, 1989). This control-based notion of responsibility, represented by luck egalitarianism, has however been criticized as morally impermissible as it could lead to what Anderson (1999: 296) calls, "the abandonment of negligent victims". Furthermore, luck egalitarianism can also be criticized for being too broad, such that it could expand the scope of health policy beyond plausibility if put into practice. For example, an individual who is dissatisfied with the looks of his nose and wishes to have plastic surgery cannot be said to have caused his unattractive nose through deliberate choice. Additionally, many of the choices influencing health might have been influenced by social factors or inadequate information. In a policymaking context, the notion of responsibility for health is therefore arguably of little relevance as it is difficult to settle which health-related choices fall within individual control.

The second domain, 'responsibility for health care', disregards underlying factors and focuses solely on the health outcomes individuals are capable of managing without direct public involvement or subsidies. Within welfare economics, individual responsibility is generally seen as a necessary element for limiting the scope of collective responsibility. The distribution of commodities through the market is the most prominent example of which individuals are assigned responsibility for any combination of goods they end up with. This shows that "responsibility is a necessary consequence of any substantial amount of freedom and is therefore part and parcel of any free society" (Fleurbaey, 2008: 1).

When focusing on responsibility for health care, the delicate but central feature of moral blame that is present when assigning responsibility for health is removed. This is why policies regarding individual responsibility most often are restricted to the health care domain, because this is also politically less sensitive (Le Grand, 1987). Consequently, the most policy-relevant concept of responsibility concerns health care, instead of health, and focuses on whether a person has the capability to take on responsibility for certain health care.

Given these issues, we will not, in this paper, explore the notion of responsibility for health as a potential rationing principle. Instead, we focus on the

5 Dworkin (1981) initially made the now famous distinction between brute luck and option luck, and argued that individuals should be responsible for the latter but not the former. Option luck is defined as those cases where individuals voluntarily and deliberately have submitted to a particular risk.

possibility and suitability of assigning individual responsibility for the financing of health care, that is, to leave it to the individual to privately finance certain kinds of health care. The practical application of individual responsibility for health-care commodities, which we explore in this paper, depends on whether or not individuals have the cognitive and physical capability of exercising their responsibility and whether or not the surrounding settings facilitate the taking of responsibility by the individual. Therefore, the characteristics of specific health-care services and the applicability or shortcomings of market forces in health care are two critical issues to be analyzed when assessing the capability of individuals to take responsibility for the financing for specific health-care services.

Health care as an economic commodity

Before exploring the distinguishing characteristics of health care as an economic commodity that might influence the capability of individuals to take responsibility, we first need to answer the question of whether or not health care really is a commodity. And moreover, what are the advantages of analyzing it as such?

The common definition of an economic commodity is that it is something that is produced and may be sold, exchanged or provided free. The commodity 'health care' can consequently be defined generically as all kinds of services and goods whose primary purpose is to improve or prevent deterioration of health and that may be sold, exchanged or provided free. The commodification of health care is sometimes criticized by individuals who argue that health is a basic right that a society has an obligation to protect from commercialism and immoral market forces, as 'health has no price'. The argument that health care constitutes a basic right and therefore should be distributed in accordance with need, and not in accordance with the ability to pay, is the most common objection to treating and analyzing health care as a commodity. It is easy to agree that need should be of great importance when distributing health care. However, this does not mean that health care is unsuitable for purchase and sale and cannot be analyzed as a commodity. The mere fact that something is considered to be a basic right does not imply that it cannot also be a commodity. For example, while food and shelter are often considered to be basic rights, they are also commonly treated as transferable commodities. More fundamentally, the commodification of health care does not preclude moral considerations. Rather, it highlights the effects of potential transactions so that policymakers can make informed decisions about the allocation of scarce resources.

The second most commonly proposed objection to treating and analyzing health care as an economic commodity is:

The central feature of health care is the personal relationship between a health professional and a person seeking help ... commodities may be used in the process of providing care, but the totality of health care itself is not a commodity.

(Pellegrino, 1999: 247)

Our interpretation of this objection is that it relates to the notion that ‘health’ should not be seen as a commodity to be bought and sold on a market. However, health is not equivalent to health care, which is simply an input to the production of health (Grossman, 1972, 2000). This is an important distinction as there are many factors that influence health, and health care is just one of these. But health care is an input factor that unlike many other factors, such as genetics, may be, and often is, distributed on a free market (which does not necessarily mean that it always should be). Consequently, despite the criticism, there can be no doubt that ‘health care’ is produced, distributed and its demand is influenced by its price, so evidently it can be considered as a commodity.

However, treating or analyzing health care as a commodity does not imply that it constitutes a homogeneous group of goods and services. On the contrary, there is a great deal of heterogeneity among health-care services. This heterogeneity constitutes a major reason why a more structured framework for exploring which services to include in or exclude from the basic health-care package is needed, instead of applying an arbitrary general attitude towards public or private funding. Moreover, treating health care as a commodity does not necessarily imply that private financing is required.

The main advantage of approaching health care as a commodity is that this allows us to consider the substitutability and heterogeneity of health care in a more structural and transparent way. Further, it allows us to study decisions and their underlying incentives as they relate to the production, distribution and consumption of health care when resources are scarce and can be put to alternative uses. Thus, to disregard the applicability and shortcomings of market forces when deciding how to finance and allocate scarce health-care resources would be both unhelpful and morally irresponsible.

Health care as an economic commodity – in what ways is it different?

Despite the fact that health care may be considered a commodity, it differs in important ways from other commodities, which makes it unsuited for distribution through the workings of price mechanisms on a free market. This idea of market failure in the market for health care has been a repeated topic since economists first started to focus on the health-care sector during the 1950s (Mushkin, 1958; Arrow, 1963; Culyer, 1971; McGuire *et al.*, 1988; Pauly, 1988a, 1988b; Hurley, 2000; Folland *et al.*, 2001; Donaldson and Gerard, 2005). In the following section, we will briefly refer to some of the familiar arguments on what distinguishes health care from other commodities, for the purpose of examining how such differences might influence the possibility of individual responsibility for health care.

The distinctiveness of health care compared to other standard commodities may be derived from four characteristics of health care. First, the demand for health care is a derived demand for health. Health care is simply an input to the

production of the desired good, which is health (Grossman, 1972, 2000). Most of the other commodities are usually consumed because they generate instant utility, whereas health care most often generates disutility before better health is produced.

Second, health care involves a great deal of informational asymmetry between patients and doctors. The behavior that a patient expects from doctors clearly differs from what a consumer expects from most other sellers/producers. This is because the health-care commodity does not exist independently of its production, which hinders the potential consumer from testing health care before consuming it. There is an obvious element of trust in the relationship, which implies that the doctor needs to be concerned about the well-being of the patient (or at least appear to be). Such a trust bond between provider and purchaser cannot be expected in other economic areas, at least not to the same extent. Further, as medical knowledge is often very complicated, it creates an asymmetric relationship between the consumer and the provider regarding the need for, and possible consequences of, receiving health care. The patient often does not have (or at least he often believes he does not have) as much knowledge as the doctor and can not thoroughly examine or question the health care prescribed. There is less possibility of accurately judging the need and quality of health care from past experiences because the occurrence of a disease is often as unpredictable as the recovery from it (Arrow, 1963; Weisbrod, 1978; McGuire *et al.*, 1988). To be well-informed, health-care consumers should be able to distinguish differences in health state with and without consumption of the commodity⁶ (Weisbrod, 1978).

The third way health care differs from other commodities is that both the possible future need for and outcome of treatments are often objects of great uncertainty. Widespread uncertainty and the risk of catastrophic expenditures are generally dealt with by insurances in most markets. However, in terms of the health-care market, there are many groups of people with considerable certainty about their future demand for health care, for example, people suffering from chronic diseases or permanent disablement. These groups are uninsurable in a private insurance market with actuarially fair premiums. Public intervention is required in order to secure financing. On the other hand, compulsory insurance against some medical expenditure may also be non-optimal due to the existence of moral hazard, for example, compulsory insurance might create excessive disincentives for individuals to take preventive measures for their health (Pauly, 1968).

Fourth, health care is frequently associated with the existence of positive externalities. Individuals, although behaving perfectly rationally, may not consume

6 In order to limit the extent to which health-care providers deviate from their agency responsibilities, society often uses constraints such as non-profit organizational form, licensure and accreditation, or imposes codes of ethics to discourage exploitation of ill-informed patients.

sufficient health care in the opinion of other individuals in society, as most of the positive effects fall on others. The origin and magnitude of these externalities may however differ depending on circumstances: (i) People may obtain satisfaction from knowing that sick and poor people consume health care (caring externalities). (ii) People may not understand their own good or lack the foresight to purchase health care if it is not publicly provided⁷ (paternalistic externalities). This type of commodities which society desires to provide in quantities greater than what consumers wish to purchase at market price are commonly referred to as merit goods (Musgrave, 1959). (iii) People may obtain increased welfare as a 'technical' positive effect from other people's consumption of health care (selfish externalities). For example, it may be assumed that healthy individuals contribute more to society in terms of production and economic growth than sick people. (iv) Finally, people may place a value on having health care available for potential use, which they may require at some future time (option-value externalities). The irregularity and unpredictability in demand for certain health-care services is a potential problem in ensuring sufficient capacity for those who do not currently require health care, but might require it at some later date.

It is important to note that none of the four characteristics outlined above are entirely unique to health care. Quite the opposite, it is rather easy to recognize other types of commodities that are similar in terms of the asymmetrical distribution of information, uncertainties regarding stochastic events and externalities. Still, the presence of all four of these characteristics at the same time is relatively unique for health care.

In addition to concerns about potential market failures associated with the commodity of health care, the financing of health care also raises serious distributive and social issues. Within welfare economics, concerns about efficiency are traditionally separated from concerns about distribution of wealth and health by maintaining that allocations generated through functioning markets are Pareto optimal, meaning technically and allocatively efficient (Reinhardt, 1998). Hence, economic evaluations most often focus merely on efficiency, leaving distributional concerns to the political process.

Among theories of distributive justice, two common ideas are that: resources should be distributed equally and that the distribution should depend on choices for which individuals can be held responsible. However, these two core ideas – equality in resources and responsibility for one's consumption choices – can conflict somewhat. To a certain extent, full equality⁸ denies individuals the possibility of acting with full agency. This can be viewed as incompatible with

7 By 'publicly provided' we mean health-care that is financed publicly, NOT that it is publicly produced.

8 Full equality is here defined according to resource egalitarianism, in which everyone gets the same share of the goods to be distributed, regardless of an individual's capacity to benefit. See, for example, Arnesson (1989) and Cohen (1989).

the concept of responsibility (Lake, 2001). An extensive amount of literature has covered different aspects of allocation concerning health and health care (Whitehead, 1992; Culyer and Wagstaff, 1993; Barer *et al.*, 1998; Sen, 2002). We will therefore not go into a lengthy discussion regarding equity in health, but be content with pointing out that as long as resources are scarce, health-care inequalities will always exist. It will however be taken for granted that equality of excess is an important policy goal.

Health care is a commodity which in many cases is of special moral importance because it protects the range of capabilities or opportunities open to people and enables them to live up to their own conception of a good life (Sen, 1992; Daniels, 2008). Hence, society has an obligation connected to its distribution that does not apply to most other commodities. This special moral importance does not, however, help us to decide which inequalities in health care are unfair. Consequently, the appropriate question to pose when rationing health care and assigning individual responsibility through private financing should be: which inequalities in health care are acceptable or justifiable from a societal point of view? This question will be explored further in the next section of the paper.

Attributes for individual responsibility – a conceptual framework for exploring the suitability of private financing in a publicly funded health-care system

When is it possible from a market perspective, and reasonable from an equity perspective, to introduce individual responsibility for health care? In the following section, we combine our defined notion of individual responsibility for health care with the distinctive features of health-care commodities in order to present a tentative conceptual framework for determining when health-care services possess such characteristics that they warrant individual responsibility through private financing.⁹ This framework consists of six attributes that should be present in order to make private financing morally justifiable.¹⁰

The concept of sufficient knowledge

The first attribute in the framework centers around the limitations that the existence of informational asymmetry between patients and doctors puts on the feasibility for individual responsibility. To allow private financing for a

⁹ The suitability of specific private financing mechanisms such as co-payments or co-insurances is, however, beyond the scope of this paper.

¹⁰ Norman Daniels and James Sabins' (2008) set of procedural principles for a making health care priority-setting decisions fairly, entitled 'Accountability for Reasonableness', has become the dominant paradigm for resolving disputes about conflicting attributes or values in health policy. The framework presented in this paper should be seen as a set of substantive principles which are needed to complement such procedures. We would like to emphasize the complementarities between procedural and substantive principles, although in this framework, we focus only on the latter.

health-care service, consumers' ability to value benefits both before and after consumption is essential. When individuals consume most other commodities, they are able to test the product and return it if it does not fulfill its purpose or patronize a different producer for future usage, for example, switch to a different restaurant or a plumber. In contrast, for many health-care services, patients are not able to test and obtain this sort of information about quality and possible utility of a certain treatment. For example, given her lack of medical expertise, a patient undergoing cytostatic treatment for cancer or in need of acute lifesaving treatment has little ability to test and obtain information about the treatment's quality and effects. Consequently, if a doctor tells a patient that an expensive treatment is necessary, most people would not be in a position to question this. In addition, the need for many medical services does not recur (at least not frequently), which prevents consumers from building knowledge from past experiences.

At the same time, there are also several health-care services where patients can be considered fairly well informed about quality and possible benefits. Pauly (1978) discusses how widespread information asymmetries in the health sector are, and suggests that at least one-fourth of total private health-care expenditures is associated with situations where patients can be regarded as 'reasonably informed'. If we include expenditures for nursing home services and chronic conditions, this share increases to approximately one-third (Pauly, 1988a, 1988b). Glasses, over-the-counter drugs and most assistive devices are other examples of services for which consumers and providers have the same informational basis as they do for many non-health-care commodities. Given such examples of reasonably informed consumers, it is clear that there exist plenty of health-care services where individuals have sufficient knowledge to facilitate taking responsibility for making informed decisions about what services to consume.

Attribute 1. The considered health-care service should enable individuals to value the need and quality both before and after utilization (consumption).

The concept of individual autonomy

A crucial assumption in welfare economics is that in order for markets to function efficiently, individuals must be capable of knowing and expressing what is best for themselves and make rational choices which maximize their personal well-being. When this assumption is violated, there will typically be inefficient markets. Within behavioral economics, the concept of bounded rationality revises this assumption to account for the fact that perfectly rational decisions are often not feasible in practice, due to the finite computational resources available for making decisions. Although individuals may have sufficient knowledge to assess need and quality, they still may lack insight about how their preferences correlate with their more fundamental needs.

Such bounded rationality extends to health care. When consuming health care, individuals are frequently unable to make rational choices whether to

consume a suggested treatment. Moreover, certain patient groups persistently make inferior decisions in terms of their own welfare – decisions that they would revise if they were equipped with full cognitive ability or full self-control. Consequently, determining the level of individual autonomy of the targeted group is essential when deciding whether it is appropriate to require private financing. Levinsson (2006) sets up four possible conditions that can be used to determine the level of autonomy of health-care consumers: the ability to be self-reflective and well reasoned; the ability to act in coherence with their intentions; the ability to understand relevant information; and the ability to foresee the consequences of their behavior.

Individuals that lack these cognitive capabilities are vulnerable to exploitation, to suffering harms and inability to give adequate consent. We can identify at least four patient groups that illustrate the concept of limited autonomy in connection to health care. The first group consists of individuals with limited cognitive abilities that make it difficult for them to foresee the consequences of their actions. These individuals may, therefore, fail to demand treatment when it is supplied on a free market. The second group consists of patients requiring emergency treatment. These individuals are frequently not in a position to express well-reasoned preferences or self-reflective enough to assess relevant information. The third group comprises individuals with addictive behavior. These individuals often have trouble acting in coherence with their intentions because of their addiction, which arguably might disqualify them as rational consumers.¹¹ The final group consists of children; however, in most cases they have guardians who are able to make rational choices for them (Culyer, 1971).

It is important to note that even if individuals have a high level of autonomy, this does not necessarily imply that they are not influenced by their surroundings but rather that they have the capability to make decisions that can be considered well-reasoned (they might still not always appear well-reasoned in the opinion of others). In addition, it is also important to note that health-care services aimed at a group with low levels of individual autonomy should not automatically be offered free of charge. However, the provision of these health-care services need to be publicly regulated to avoid exploitation of groups that are not capable of acting as rational consumers.

Attribute 2. The considered health-care service should be directed towards individuals with a reasonable level of individual autonomy.

The concept of positive externalities

The third attribute of the framework focuses on the existence of external benefits associated with the consumption of health-care services, that is, benefits

¹¹ Thus, we believe that the concept of rational addiction is problematic when it comes to actual decisions on health-care consumption (Becker and Murphy, 1988).

accruing to persons other than the patient in focus of an intervention or to the public at large. Due to public interest, society may wish to subsidize services with so-called positive externalities, such as vaccinations against epidemic diseases. There are at least three separate types of externalities that need to be considered.

The first type of externalities are what Culyer calls “caring externalities”, which arise because “individuals are affected by others’ health status for the simple reason that most of them care” (Culyer, 1976: 89). In other words, the sick person’s pain and lack of treatment cause disutility for other people in society. It has been argued that the existence of caring externalities is the main reason for the existence of the publicly funded welfare state (Culyer, 1980; Evans, 1984).

The second type of externalities are paternalistic externalities related to the existence of merit goods. These arise when society in general feels that individuals in particular circumstances ought to consume health-care commodities based on norms rather than leaving consumption to consumer rationality. For example, merit goods such as many keep-fit activities are subsidized, while goods viewed as demerit goods, such as tobacco and alcohol are taxed in order to reduce demand (Evans, 1984). These paternalistic attitudes arise because caring for the sick is often seen as a virtue in communities. The individual does not necessarily care about the improved health or utility of others, instead, the positive effect is limited to the instrumental act of affecting the consumption patterns in various ways. This distinguishes paternalistic externalities from caring externalities in view of the fact that the latter are concerned with the actual health status of the persons concerned (also called specific altruism) (Jacobsson, 2005).

The third type of externalities are what we here refer to as ‘selfish externalities’. Some health-care services might, for example, determine whether or not a person is able to proceed with a certain job and thus generate positive externalities to society as a whole. Many assistive devices may increase the potential for individuals to manage without personal assistance from relatives or others, and this benefit could be considered as a selfish externality.

Attribute 3. The considered health-care service should be associated with low levels of positive externalities.

The concept of sufficient demand

The fourth attribute that needs to be taken into consideration when exploring the feasibility of individual responsibility through out-of-pocket payments is whether demand is regular and of a reasonable magnitude. Relying on private financing to support certain health-care services for which demand may be irregular and unpredictable, would lead to insufficient capacity to provide those services in the future (option-value externality). For example, in some sparsely populated areas, the demand for health-care treatment is insufficient to support

a system based exclusively on private financing. Consequently, health care may need to be financed collectively if society wants to ensure health-care capacity outside heavily populated areas, or if it wishes to ensure capacity to treat extremely rare and complicated illnesses.¹²

Attribute 4. The considered health-care service should be associated with a demand of sufficient magnitude to generate a private market.

The concept of affordability

The fifth attribute is the affordability of specific services, that is, payments of a reasonable magnitude for most individuals. Because wealth and income are distributed unequally between individuals it is hard to set a definite boundary for what constitutes an affordable price. However, if society wishes to limit the influence of economic inequality on the consumption of health care, then affordability needs to be taken into consideration. Therefore, a financing system needs to include arrangements for exemptions and/or high-cost protection in order to protect individuals from unaffordable and catastrophic health-care payments. The absence of such arrangements would result in scenarios likely to violate the general public's notions of 'equity' and 'fairness' (Hauck *et al.*, 2004). Consequently, there is a need to make exemptions or price reductions in any generalized system for the use of private financing, in order to gain public acceptability.

The concept of affordability is most commonly used in policy discussions in the context of out-of-pocket payments, but unfortunately the meaning of affordability most often remains obscure, both in general discussions and in more rigorous economic analyses. Bundorf and Pauly (2006) give a normative definition of affordability that can serve as a frame for further specification: "The special good is affordable to an individual if her income after the purchase of the socially acceptable minimum quantity of the good is greater than or equal to the socially defined minimum spending on other goods" (2006: 653). Hence, an affordability threshold needs to be defined in order to further make the concept of affordability operational and this can be done in various ways. Ultimately, defining affordability is as much a political and normative task as it is an economic one.¹³

The stochastic nature of medical expenditures makes it very difficult for individuals to incorporate these expenditures when planning household budget. Consequently, the concept of affordability implies that the services which are appropriate for private financing, are generally those that will not inflict

12 Of course, some health-care products can be kept in storage and therefore do not require high demand. License requirements for treatment and prescriptions may limit the incentive for a local profit-maximizing entrepreneur to enter a low demand market without public funding (Donaldson and Gerard, 2005).

13 In this way, affordability is not very different from applying a societal willingness to pay for a particular service or a quality-adjusted life year (QALY).

widespread catastrophic payments out-of-pocket. Moreover, because private insurance will not be available for everyone in an insurance market with actuarially fair premiums, low-cost services are generally more suitable for private financing.

Attribute 5. The considered health-care service should be associated with payments affordable for most individuals.

The concept of lifestyle enhancement

The final concept, lifestyle enhancement, is to some degree separate from the preceding concepts in the sense that it deals with how to define the core purpose of a health-care system. Rather than eliciting the feasibility of individual responsibility, the concept of lifestyle enhancement aims to avoid over-consumption or moral hazard that could threaten the financial sustainability of any public health-care system.

In order to achieve financial balance, a publicly funded system needs to match revenues against costs for actuarially predictable events. However, moral hazard – defined as over-consumption caused by facing less than full cost of service – threatens to undermine the balancing of revenues and costs. Moral hazard connected to health care is, however, a special case of moral hazard since health care simply is an input to the production of the desired good – health. Compared to other commodities, health care in itself is undesirable as it often generates disutility before better health is produced. Consequently, because moral hazard for medical treatments that deals with pain or symptom avoidance is limited, there is some degree of self-regulation in the health-care market.

There also exist a wide range of health-care services that do not involve any concept of illness or disease. These so-called ‘lifestyle enhancements’ are health-care services that aim to fulfill wishes outside the medical sphere. The risk for moral hazard in lifestyle enhancements is much greater than in other health-care areas, as the lack of disutility associated with such health condition or generated by the health-care service do restrain demand. Therefore, to maintain financial sustainability in public health-care systems, services associated with lifestyle enhancements need to be funded privately to some extent.¹⁴

In many countries, despite having been shown as highly cost-effective, health care for reasons other than disease or injury has been excluded from public funding. The most prominent rationing debate has concerned so-called ‘lifestyle drugs’. The exclusion of such health-care services from public funding has most often been justified by rationales implying that such health-care services are lifestyle enhancements or luxuries rather than necessities. Despite how commonly such terms are used, it is difficult to define what is meant by lifestyle enhancements or

¹⁴ See Pauly (1968) and Zweifel and Manning (2000) for empirical examples on moral hazard related to health-care consumption.

luxury health-care, as these concepts are influenced by social and cultural norms (Reissman, 1999). The theory developed by the economist Tibor Scitovsky (1977) could be helpful in refining these concepts further. He distinguishes two sorts of utility gain, utility arising from ‘pleasure seeking’ (lifestyle enhancement) and utility arising from ‘pain avoiding’ (necessity consumption). ‘Pleasure seeking’ connected to health can be considered as those services that people desire in order to exceed a societal reference point for health (Stolk *et al.*, 2002). Such services could, for instance, be cosmetic treatments, doping or medication for erectile dysfunction.¹⁵ Similarly, pain avoidance connected to health care can be considered as those services desired because patients are currently below a societal reference point for health, for example, cancer treatments. The reference point for health might, however, differ between societies and patient groups. For example, older people probably have a lower reference point than younger people, concerning many health states (Williams, 1997). A decline in mobility that is considered unacceptable for a 20-year-old is in many cases acceptable for an 80-year-old. Obviously, grey areas will arise in which it is unclear whether a service constitutes a lifestyle enhancement or a medical necessity. For example, it can be debated whether treatments that increase height, muscle capacity or intelligence, constitute enhancements rather than treatments for medical necessities. Distinguishing between what should be considered reconstructive surgery, and what cosmetic, is similarly difficult. However, certain ‘non-medical’ cases that seem to come close to treating an illness, for example, in cases when the desire to look differently causes significant psychological distress that it leads to seeking surgery. How then, should the line between treating an illness and treating a non-medical wish to be drawn? How do we determine which unwanted conditions are worthy of treatment? This, of course, is a very delicate issue, but it is the type of normative judgement that any public health-care system has to confront. As Sandel (2004) argues:

In order to grapple with the ethics of enhancement, we need to confront questions largely lost from view – questions about the moral status of nature, and about the proper stance of human beings toward the given world.

(Sandel, 2004: 51)

When these kinds of normative judgments have been made for a specific health-care service, it is important that the rationale for this decision is clearly stated and defended publicly so that this can serve as a ‘case law’ for similar policy decisions in the future. This will demonstrate coherence and consistency, which will strengthen the legitimacy of public decision makers in the long run.¹⁶

¹⁵ How to distinguish medical usage of medication for erectile dysfunction, from lifestyle purposes is however debated vigorously; see for example, Gilbert *et al.* (2000).

¹⁶ One approach for making such rationales transparent is offered by Norman Daniels and Jim Sabin in their framework Accountability for Reasonableness (A4R). For a more detailed discussion concerning A4R, see chapter four in Daniels (2008).

Attribute 6. Health-care services associated with lifestyle enhancements rather than medical necessities are eligible for private financing.

Discussion

Decisions concerning the rationing of health-care services are made in all publicly funded health-care systems, and may significantly affect both individuals' health and broader economic situation. Therefore, it is of great importance to explicitly discuss the underlying logic for such decisions to avoid arbitrary and unfair allocation of health-care resources. Although individual responsibility connected to health care evidently is an influential factor for decisions on rationing, individual responsibility is seldom explicitly tackled in the literature concerning economics and public policy. This article is an attempt to bring some attention to the role of individual responsibility in health-care rationing – a matter we firmly believe benefits from multidisciplinary approaches to fill the gap between existing guiding ethical principles and economic approaches. The main objective of this article is to explore how the characteristics of health care, as an economic commodity, can be combined with distributive considerations into a conceptual framework for determining which types of health-care service can be left to individual responsibility through private financing.

As emphasized in this paper, health care does not consist of a homogeneous group of commodities, and consequently, decisions on rationing cannot be computed through a simple formula. The framework developed in the paper should be used to explore health care as a heterogeneous group of commodities, rather than applying an all-or-nothing approach to inclusion or exclusion in the publicly funded health-care package. We argue that adopting this approach to health-care rationing has many advantages, especially in that it forces decision makers to explicitly define the objectives of the rationing or priority-setting process, even if these objectives cannot always be easily observed or measured. By supporting decision makers with substantive concepts and attributes focusing on health care as a commodity, we hope that our tentative framework will constitute a good point of departure for exploring the feasibility of individual responsibility when it comes to funding health care. However, it is important to note that this does not imply that patients should be prevented from being treated in accordance with their specific prerequisites. We merely focus here on the need of a more systematic approach to assess the appropriateness of private financing for certain services, as there are health-care services where public funding is more appropriate than private funding, just as there are many health-care services where private funding is more appropriate than public funding. For this reason the societal objective should be to seek the optimal combination of public and private funding within the public health-care system.

The tentative conceptual framework will hopefully support policymakers in accomplishing two key objectives; (i) to identify the feasibility of individual

responsibility through the characteristics of health care as commodities; and (ii) to facilitate a transparent and systematic discussion on health-care rationing through a fair deliberative procedure. Attributes 1 to 3 in our framework (sufficient knowledge, individual autonomy and positive externalities) are all connected to the specific characteristics of health care that violate crucial assumptions for a well-functioning market. These attributes allow us to identify those specific services that can be allocated through the working of market mechanisms without resulting in sub optimal outcomes due to market failures. Attribute 4, sufficient demand, is further connected to specific characteristics concerning the provision of health care, which may call for public funding in order to ensure that individuals are able to take individual responsibility. Health-care services that (for reasons other than market failures) are not available through private alternatives are unsuited for individual responsibility. The lack of these services would consequently make them high-priority candidates for inclusion in the publicly funded health-care system. In addition, policymakers also have an obligation to protect people from unbearable financial costs associated with poor health. Rationales for collective responsibility begin with some concept of fairness, usually involving the belief that no one should be denied health care due to inability to pay, or the idea that public funds should not be spent on 'pleasure seeking' rather than pain or symptom avoidance. Such beliefs about equality and fairness constitute an important basis for including the concepts of affordability and 'lifestyle enhancements' into the framework. The attributes related to them are of utmost importance for any approach exploring the feasibility of individual responsibility for financing consumption of health-care commodities.

What is the potential use for this conceptual framework? Past experiences in the Netherlands and Sweden have shown that any rationing principle that is poorly operational is doomed to be of limited use in priority setting. We hope that our conceptual framework will contribute to better operationalization of the notion of individual responsibility so that it can be of more explicit use in public policy. Experience in the Netherlands also shows that criteria intended to give a clear yes or no to funding are judged to be inappropriate. Hence, it is not our intention that the attributes, or the framework as a whole, should provide clear-cut answers. Rather, the six attributes need to be weighed against each other when exploring the feasibility for individual responsibility for health care. Considering health care as a set of economic commodities is just one element of the broader priority-setting process. Rationing will always involve conflicts between different values. But instead of obscuring such conflicts, we hope that our framework will provide a transparent and structural way to approach the inescapable issues of individual responsibility.

When formulating rationing principles at a policy level, there are two key risks: either the formulation is so tightly framed that it overrides clinical freedom and excludes health care for deserving cases, or alternatively it is too

loosely framed, which forces clinicians to ration health care on a case-to-case basis and renders the formulation useless as a support to clinical practice (Newdick, 2005). The ethical core values, such as human dignity and need, that are supposed to guide local authorities in setting health-care priorities in Sweden, seem to suffer from the latter. And although cost-effectiveness models have become consistently more intuitive and accessible over the past decade, there is still ground to cover before they will be widely accepted as a rationing principle. Since “if explicit decisions on priority setting, albeit based on strong economic evidence, are felt to be politically unacceptable they are unlikely to be implemented” (Robinson, 1999: 23). Ultimately, the political process in democratic societies will determine which services to exclude from the publicly funded health-care system. Consequently, we are faced with two alternatives: we can either ignore the fact that we can not afford everything and just let rationing take its course by default, or we can make decisions on rationing explicitly by design. This paper has provided a tentative framework to aid policymakers in conducting rationing by design, filling the gap between existing ethical principles and cost-effectiveness, by explicitly addressing the question: individual responsibility for what?

Acknowledgements

We are grateful to Adriane Gelpi, Alena Buyx, Fredric Jacobsson, Lars Sandman, Norman Daniels, Elizabeth Fenton, Kristin Voigt, Dan Wikler, Keren Ladin, Dan Brock, Petter Tinghög and two anonymous reviewers for helpful comments and suggestions on earlier drafts of this paper.

References

- Anderson, S. E. (1999), ‘What is the point of equality?’, *Ethics*, **109**: 287–337.
- Arneson, R. J. (1989), ‘Equality and equal opportunity for welfare’, *Philosophical Studies*, **56**: 77–93.
- Arrow, K. J. (1963), ‘Uncertainty and the welfare economics of medical care’, *American Economic Review*, **53**: 941–973.
- Barer, M. L., T. E. Getzen and G. L. Stoddart (1998), *Health, Health Care and Health Economics: Perspectives on Distribution*, Chichester: John Wiley.
- Barry, B. (2005), *Why Social Justice Matters*, Oxford: Polity.
- Becker, G. S. and K. M. Murphy (1988), ‘A theory of rational addiction’, *Journal of Political Economy*, **96**(4): 675–700.
- Bundorf, M. K. and M. V. Pauly (2006), ‘Is health insurance affordable for the uninsured?’, *Journal of Health Economics*, **25**(4): 650–673.
- Buyx, A. M. (2008), ‘Personal responsibility for health as a rationing criterion: why we don’t like it and why maybe we should’, *Journal of Medical Ethics*, **34**(12): 871–874.
- Cappelen, A. W. and O. F. Norheim (2006), ‘Responsibility, fairness and rationing in health care’, *Health Policy*, **76**(3): 312–319.
- Cohen, G. A. (1989), ‘On the currency of egalitarian justice’, *Ethics*, **99**: 906–944.

- Culyer, A. (1971), 'The nature of the commodity health care and its efficient allocation', *Oxford Economic Papers*, 23: 189–211.
- Culyer, A. J. (1976), *Need and the National Health Service: Economics and Social Choice*, London: M. Robertson.
- Culyer, A. (1980), *The Political Economy of Social Policy*, Oxford: Martin Robertson.
- Culyer, A. J. and A. Wagstaff (1993), 'Equity and equality in health and health care', *Journal of Health Economics*, 12(4): 431–457.
- Daniels, N. (2008), *Just Health: Meeting Health Needs Fairly*, Cambridge; New York: Cambridge University Press.
- Daniels, N. and J. E. Sabin (2008), *Setting Limits Fairly: Learning to Share Resources for Health*, New York; Oxford: Oxford University Press.
- Donaldson, C. and K. Gerard (2005), *Economics of Health Care Financing: The Visible Hand*, Basingstoke: Palgrave Macmillan.
- Drummond, M., G. Torrance and J. Mason (1993), 'Cost-effectiveness league tables: more harm than good?', *Social Science and Medicine*, 37(1): 33–40.
- Dworkin, R. (1981), 'What is equality? Part II: Equality of resources', *Philosophy and Public Affairs*, 10(4): 283–345.
- Evans, R. G. (1984), *Strained Mercy: The Economics of Canadian Health Care*, Toronto: Butterworths.
- Fleurbaey, M. (1995), 'Equality and responsibility', *European Economic Review*, 39: 683–689.
- Fleurbaey, M. (2008), *Fairness, Responsibility, and Welfare*, Oxford: Oxford University Press.
- Folland, S., A. C. Goodman and M. Stano (2001), *The Economics of Health and Health Care*, Upper Saddle River, NJ: Prentice Hall.
- Garber, A. M. and C. E. Phelps (1997), 'Economic foundations of cost-effectiveness analysis', *Journal of Health Economics*, 16(1): 1–31.
- Garber, A. M. and C. E. Phelps (2008), 'Future costs and the future of cost-effectiveness analysis', *Journal of Health Economics*, 27(4): 819–821.
- Gilbert, D., T. Walley and B. New (2000), 'Lifestyle medicines', *British Medical Journal*, 321(7272): 1341–1344.
- Government Committee on Choices in Health Care (1992), *Choices in Health Care*, Rijswijk: Ministry of Welfare, Health and Cultural Affairs.
- Grossman, M. (1972), 'On the concept of health capital and the demand for health', *Journal of Political Economy*, 80(2): 223–255.
- Grossman, M. (2000), 'The Human Capital Model', in A. J. Culyer and J. P. Newhouse (eds), *Handbook of Health Economics*, volume 1, Amsterdam: Elsevier: 347–408.
- Hadorn, D. C. (1991), 'Setting health care priorities in Oregon. Cost-effectiveness meets the rule of rescue', *Journal of the American Medical Association*, 265(17): 2218–2225.
- Hart, H. L. A. (1968), *Punishment and Responsibility: Essays in the Philosophy of Law*, Oxford: Oxford University Press.
- Hauck, K., P. C. Smith and M. Goddard (2004), *The Economics of Priority Setting for Health Care: A Literature Review*, Washington: The International Bank for Reconstruction and Development.
- Hoel, M. (2007), 'What should (public) health insurance cover?', *Journal of Health Economics*, 26: 251–262.
- Hurley, J. (2000), 'An Overview of the Normative Economics of the Health Sector', in A. J. Culyer and J. P. Newhouse (eds), *Handbook of Health Economics*, volume 1, Amsterdam: Elsevier: 56–118.

- Jacobsson, F. (2005), 'The good economaritan – efficiency and equity considerations concerning altruism and health' (in Swedish), Dissertation, Linköping University, Linköping.
- Lake, C. (2001), *Equality and Responsibility*, Oxford: Oxford University Press.
- Le Grand, J. (1987), 'Equity, health and health care', *Social Justice Research*, 1: 257–274.
- Le Grand, J. (2003), *Motivation, Agency, and Public Policy: Of Knights and Knaves, Pawns and Queens*, Oxford: Oxford University Press.
- Levinsson, H. (2006), 'Den självbestämmande individen', in K. Hansson (ed.), *Etiska utmaningar: i hälso- och sjukvården*, Lund: Studentlitteratur: 191.
- McGuire, A., J. Henderson and G. Mooney (1988), *The Economics of Health Care: An Introductory Text*, London: Routledge & Kegan Paul.
- Meltzer, D. (1997), 'Accounting for future costs in medical cost-effectiveness analysis', *Journal of Health Economics*, 16(1): 33–64.
- Minkler, M. (1999), 'Personal responsibility for health? A review of the arguments and the evidence at century's end', *Health Education and Behaviour*, 26(1): 121–140.
- Musgrave, R. A. (1959), *The Theory of Public Finance: A Study in Public Economy*, New York; London: McGraw-Hill.
- Mushkin, S. J. (1958), 'Toward a definition of health economics', *Public Health Reports*, 73(9): 785–793.
- National Health Service (2009), 'Guidance on NHS patients who wish to pay for additional private care – a consultation', http://www.dh.gov.uk/en/Consultations/Liveconsultations/DH_089926 [3 March 2009].
- Newdick, C. (2005), *Who Should We Treat? Rights, Rationing, and the Resources in the NHS*, New York: Oxford University Press.
- Olsen, J. A., J. Richardson, P. Dolan and P. Menzel (2003), 'The moral relevance of personal characteristics in setting health care priorities', *Social Science and Medicine*, 57(7): 1163–1172.
- Pauly, M. V. (1968), 'The economics of moral hazard', *The American Economic Review*, 58(3): 531–537.
- Pauly, M. V. (1978), 'Is Medical Care Really Different?', in W. Greenberg (ed.), *Competition in the Health Care Sector*, Germantown: Aspen Systems.
- Pauly, M. V. (1988a), 'A Primer on Competition in Medical Markets', in H. E. Frech (ed.), *Health Care in America: The Political Economy of Hospitals and Health Insurance*, San Francisco, CA: Pacific Research Institute for Public Policy: 401.
- Pauly, M. V. (1988b), 'Is medical care different? Old question, new answers', *Journal of Health Politics Policy and Law*, 13(2): 227–237.
- Pellegrino, E. D. (1999), 'The commodification of medical and health care: the moral consequences of a paradigm shift from a professional to a market ethic', *Journal of Medical Philosophy*, 24(3): 243–266.
- Rawls, J. (1971), *A Theory of Justice*, Cambridge, MA: Belknap Press of Harvard University Press.
- Reinhardt, U. (1998), 'Abstracting from Distributional Effects, this Policy is Efficient', in M. L. Barer, T. E. Getzen and G. L. Stoddart (eds), *Health, Health Care and Health Economics: Perspectives on Distribution*, Chichester: John Wiley.
- Reissman, D. (1999), 'The pros and cons of covering lifestyle drugs', *Disease Management and Health Outcomes*, 6: 249–251.
- Robinson, R. (1999), 'Limits to rationality: economics, economists and priority setting', *Health Policy*, 49(1–2): 13–26.

- Sandel, M. J. (2004), 'The case against perfection: what's wrong with designer children, bionic athletes, and genetic engineering', *Atlantic Monthly*, 292(3): 50–54, 56–60, 62.
- Schmidt, H. (2007), 'Personal responsibility for health – developments under the German Healthcare Reform 2007', *European Journal of Health Law*, 14(3): 241–250.
- Scitovsky, T. (1977), *The Joyless Economy: An Inquiry into Human Satisfaction and Consumer Dissatisfaction*, Oxford: Oxford University Press.
- Sen, A. (1992), *Inequality Reexamined*, Oxford: Oxford University Press.
- Sen, A. (2002), 'Why health equity?', *Health Economics*, 11(8): 659–666.
- Stolk, E. A., W. B. Brouwer and J. J. Busschbach (2002), 'Rationalising rationing: economic and other considerations in the debate about funding of Viagra', *Health Policy*, 59(1): 53–63.
- Swedish Parliamentary Priorities Commission (1995), 'Priorities in health care: ethics, economy, implementation', *Statens offentliga utredningar*, Stockholm: Fritze: 5E, 180.
- Weisbrod, B. A. (1978), 'Comments on M. V. Pauly', in W. Greenberg (ed.), *Competition in the Health Care Sector: Past, Present, and Future*, Washington, DC: Government Printing Office.
- Whitehead, M. (1992), 'The concepts and principles of equity and health', *International Journal of Health Services*, 22(3): 429–445.
- Wikler, D. (2002), 'Personal and social responsibility for health', *Ethics and International Affairs*, 16(2): 47–55.
- Williams, A. (1997), 'Intergenerational equity: an exploration of the "fair innings" argument', *Health Economics*, 6(2): 117–132.
- Zweifel, P. and G. W. Manning (2000), 'Moral Hazard and Consumer Incentives in Health Care', in A. J. Culyer and J. P. Newhouse (eds), *Handbook of Health Economics*, volume 1A, Amsterdam: Elsevier: 1–890.

PAPER

II

Discounting, Preferences, and Paternalism in Cost-Effectiveness Analysis

Author: Gustav Tinghög, M.S.^{1,2}

1. Center for Medical Technology Assessment, Department of Medical and Health Sciences,
Linköping University, Linköping, Sweden
2. The National Center for Priority Setting in Health Care, Linköping, Sweden

Corresponding Author:

Gustav Tinghög

Department of Medical and Health Sciences

Linköping University

SE-581 83 Linköping

Sweden

Email: gustav.tinghog@liu.se

Phone: +46 13 221751

Fax: +46 13 224995

Abstract

When assessing the cost effectiveness of health care programmes, health economists typically presume that distant events should be given less weight than present events. This article examines the moral reasonableness of arguments advanced for positive discounting in cost-effectiveness analysis both from an *intergenerational* and an *intrapersonal* perspective and assesses if arguments are equally applicable to health and monetary outcomes. The article concludes that behavioral effects related to time preferences give little or no reason for why society at large should favour the present over the future when making intergenerational choices regarding health. The strongest argument for discounting stems from the combined argument of diminishing marginal utility in the presence of growth. However, this hinges on the assumption of actual growth in the relevant good. Moreover, current modern democracy may be insufficiently sensitive to the concerns of future generations. The second part of the article categorises preference failures (which justify paternalistic responses) into two distinct groups, *myopic* and *acritic*. The existence of these types of preference failures makes elicited time preferences of little normative relevance when making decisions regarding the social discount rate, even in an intrapersonal context. As with intergenerational discounting, the combined arguments of growth and diminishing marginal utility offer the strongest arguments for discounting in the intrapersonal context. However, there is no *prima facie* reason to assume that this argument should apply equally to health and monetary values. To be sure, selecting an approach towards discounting health is a complex matter. However, the life-or-death implications of any approach require that the discussion not be downplayed to merely a technical matter for economists to settle.

1. Introduction

When conducting cost-effectiveness analysis (CEA), discounting enables policymakers to compare competing health programmes that entail intertemporal trade-offs between present and future values, typically by assigning less value to future events in comparison to events closer to the present. Although the discount rate is a crucial parameter in CEA, it is often employed arbitrarily and with little or no explicit justification. As Krahn and Gafni (1993) remark: "Most analyses, including those who take great care to measure costs and consequences, pull their discount rates either out of the air or off the shelf, and the lucky number is most often 5%." (p. 415). The widespread lack of explicit justification reflects persistent disagreement about the legitimate basis for discounting future events, a debate that has lasted for decades (e.g. Pigou 1932; Baumol 1968; Olsen 1993). The discipline of discounting has thus become a battleground where heavyweight intellectuals from different academic fields wield powerful and impressively diverse arguments. It has also become a discipline for great soliloquists – skilled at presenting their own arguments, but less prone to consider others.

Against the background of widespread disagreement, this article begins by categorising the rationales advanced for discounting health into two distinct theoretical groups based on Ramsey's (1928) formalisation of the social discount rate. Thereby, I intend to address three recurrent questions with particular relevance in the context of applied cost-effectiveness analysis: i) Which of the motives outlined in the literature constitute morally legitimate reasons for discounting future values in CEA?; ii) Do these motives apply differently when making intertemporal valuations in comparing health to wealth?; and iii) To what extent should the social discount rate used in CEA incorporate considerations for individuals' time preferences?

The main conclusion of this article is that the behavioral effects related to time preferences advanced in the literature provide little or no reason why society at large should favour the present over the future when making *intergenerational* choices regarding health. From a fairness perspective it seems impermissible that society should favour current over future generations purely on the basis of position in time. Moreover, contemporary democracy may be insufficiently sensitive to the concerns of future generations, rendering a paternalistic

approach towards the social discount rate preferable. However, the fact that behavioural effects give little support to discounting future health does not mean that discounting in CEA should be abandoned altogether. Factors such as growth and diminishing marginal utility, which are likely to change as functions of temporal distance, may still justify assigning less value to future outcomes. This does not mean that temporal distance in itself offers any legitimate reason for discounting, and there is no *prima facie* reason to assume that these factors should apply equally to health and money. For *intrapersonal*, intertemporal choices regarding health, the issues of rationality and justifiable forms of paternalism become more apparent aspects when discussing the normative relevance of individuals' time preferences. This article categorises preference failures, which justify paternalistic intervention, into two distinct groups, *myopic* and *acratia*. The existence of these types of preference failures renders elicited time preferences of little normative relevance when making decisions regarding the social discount rate. Some factors of individuals' time preferences, such as dread and anticipation, could however potentially constitute a legitimate basis for discounting in the intrapersonal context once these have been separated from myopic and acratia elements.

Another important argument concerning the relevant basis for social discounting will also be presented, specifically: Although uncertainty is a valid reason for assigning less value to future outcomes, it should not be dealt with through the discount factor since this mis-states our moral view – future events are not less important, but less likely. From a methodological perspective, incorporating uncertainty as a basis for social discounting would result in double counting since life tables and uncertainty parameters included in CEA are supposed to take the likelihood of future events into account.

The article is divided into five sections: Section one introduces the discounted utility (DU) model to illustrate how discounting works in an applied context. Section two follows the structure of Ramsey's formula for the social discount rate to outline the rationales for discounting future health, based on arguments related to time preferences. Section three continues by outlining arguments advanced on the basis of growth and diminishing marginal utility. Section four builds on the preceding sections as it transcends into analysing and justifying rationales advanced for discounting future health as opposed to monetary outcomes in CEA. Section five discusses the extent to which the social discount rate should incorporate considerations for individuals' time preferences for health, and presents cases where

paternalism could be justified. The article concludes with some non-utilitarian remarks concerning the DU model and CEA.

2. Formal models of discounting

Frank Ramsey (1928) presented one of the earliest formalizations of the wide range of reasons expressed for why resources available in the future should be valued less than the same amount available today. His early formula of the social discount rate offers a good starting point for exploring the normative basis for discounting future health:

Social discount rate: $D = \delta + \epsilon_x g_x$

This formula helps distinguish three main components of the rationale for discounting. First, individual time preference (δ) is motivated by the assumption that individuals are impatient and generally prefer a given amount of resources sooner rather than later. Second, the growth rate of good x (g) motivates discounting because if society can invest to transform resources available today into greater amounts in the future, there are opportunity costs that a social time preference should take into account. The third part of the formula, the elasticity of the marginal utility of good x (ϵ), serves as a link between the other two main components. It implies that if there will be more of a good (x) in the future, the utility that an individual derive from a given unit of that good in the future would be less than it is today.

Before exploring the legitimacy these categories of reasons might give to discounting future health, it is important to have an understanding of how intertemporal choices are analyzed in practice. The discounted utility (DU) model is the primary framework through which intertemporal decisions are analysed in economic evaluations today. The model was initially developed by Paul Samuelson (1937), who condensed the aforementioned categories of reasons for giving less weight to future resources or events into a single parameter – the discount factor. The DU model makes it possible to calculate overall intertemporal utility simply by multiplying each outcome by a discount factor. Due to the appealing simplicity of the DU model, and its resemblance to the interest rate formula, it quickly became (and remains) the standard model for intertemporal public decisions. Formally, the DU model can be expressed as follows:

$$U(x) = \sum_{t=0}^{\infty} \delta^t U(x_t)$$

$$\text{where } \delta^t = \frac{1}{1 + p}$$

Where, $U(x)$ represents the utility of outcome x , $U(x_t)$ represents individuals' instantaneous utility, and δ^t represents the discount factor, which can vary according to the relative weight placed on utility in period t compared to future periods. Box 1 further illustrates how the DU model operates in the case of CEA. The general conclusion derived from the literature is that the DU model has little descriptive validity when it comes to describing actual intertemporal behavior among individuals. This has been confirmed by a wide range of experimental studies on time preferences that have displayed numerous anomalies. For example, it has been observed that people do not appear to discount future outcomes at a rate that is constant over time (often referred to as hyperbolic discounting), and that gains tend to be discounted at a higher rate than losses.¹ Although the DU model has little descriptive validity, this does not necessarily imply that it has no normative validity.

Insert box 1

Although discounting might appear to be a merely technical procedure to non-economists, it is important to acknowledge that discounting implies making decisions about delicate moral issues that have crucial practical implications. For instance, if we apply the DU model and assume that the benefit is lives saved, and we assign a discount factor of 3%, this would mean that we state that 1 life saved today is equivalent to saving 19.2 lives in 100 years. Thus, with a 3% discount factor it is preferable to save 1 life today rather than 18 lives in 100 years according to the DU model. As Table 1 demonstrates, a larger discount factor is associated with giving more weight to the present. Hence, different decisions on discount rates will lead to different recommendations on how to allocate health care resources – recommendations that carry life or death implications. It is important not to downplay the normative discussion

¹ For a systematic review of DU anomalies see Fredrick et al. (2003)

on discounting as merely a technical matter to be settled by health economists ‘who know the price of everything but the value of nothing’.

Table 1 Number of future lives equal to one present life at different discount factors

Years in the future	3%	5%	10%
10	1.3	1.6	2.6
50	4.4	11.5	117.4
100	19.2	131.5	13780.6

3. Time preferences as a basis for discounting health (δ)

Two principal behavioural effects have been advanced to explain why individuals’ time preferences should ascribe lower value to future benefits: i) the pure time preference effect and ii) the uncertainty effect.

3.1 Pure time preference

The first argument, the *pure time preference effect*, refers solely to the effect of remoteness in time, i.e. individuals generally prefer benefits sooner rather than later, or prefer immediate utility to delayed utility. Already in the economic literature of the 19th century, Eugen von Böhm-Bawerk acknowledged individuals’ systematic impatience:

We must now consider a /.../ phenomenon of human experience – one that is heavily fraught with consequence. That is the fact that we feel less concerned about future sensations of joy and sorrow simply because they do lie in the future, and the lessening of our concern is in proportion to the remoteness of that future (Cited in Van Liedekerke 2004, p. 74).

When considering the concept of pure time preference, it is important to distinguish it from preferences for time discounting, which more broadly include any reason for giving less weight to future events. Pure time preferences refer solely to individuals’ degree of impatience and do not take into account any uncertainties about the future or any misguided preferences that individuals might have. People may act impatiently, unaware of the future consequences. This, however, should not be regarded as a case of positive pure time preferences, but of information failure. The concept of pure time preferences simply states

that I prefer to have X today rather than having X at some point in the future, for no other reason than the fact that I dislike waiting.

The arguments of intergenerational fairness and personal identity

The pure time preference effect as a basis for social discounting of future health has been subjected to criticism, mainly because it disregards the preferences of future generations. From a fairness perspective it could be argued that a social discount rate should protect future generations from the selfish preferences of current generations – that it is unfair to count a quality-adjusted life-year (QALY) gained by a person in 40 years as only 1/7th of a QALY today (using a 5% discount rate). Hence, the fact that current generations favour themselves over their descendants should not be considered an ethically valid basis for discounting. In a theory of justice, Rawls (1973) argues that decision makers should not give any weight to mere position in time since no one knows when they will be alive behind the veil of ignorance. Following Rawls' argument, individuals would give equal weight to all periods in time if they were behind the veil of ignorance, not knowing in which period of time they would find themselves.

The matter of personal identity, as proposed most notably by Parfit (1984), extends the fairness criticism beyond the case of intergenerational discounting. In short, Parfit uses the concept of personal identity to link a person in one time period with the same person in a later period. Parfit uses this account of personal identity to contend that there is no long-term self or "I" to which future benefits can be ascribed. Instead, the decrease in psychological connections gives the future self a status of other people. For example, a 21-year-old person will – through psychological features such as memory and intent – be closely connected to his actions as a 20-year-old. However, the psychological link between the 20-year-old and his future self as a 65-year-old will be much weaker since his personal identity has changed. He treats his future selves as if they were different individuals to whom he is more or less attached. Consequently, the 20-year-old can impose *costs* on his future self. But the 65-year-old future self has had little influence on the actions his younger self imposed on him, although he must bear much of the cost. The implication from Parfit's argument on discounting health would be that a certain degree of short-sighted, pure time preference is rational because individuals become progressively less psychologically attached to their future selves. Therefore, it would not be a sign of irrationality for individuals to give future

health less weight than their present health. In this respect, the utility experienced by our future selves is not fully one's "own". Parfit thereby extends the fairness argument to include future selves as future generations. Although Parfit's claim about personal identity has received considerable attention among philosophers, economists have largely ignored the topic when discussing intertemporal choices.²

3.2 Uncertainty

The *uncertainty effect* is the second principal argument advanced to explain why individuals' time preferences should ascribe lower weight to future benefits than to present benefits. The common way to express this effect in relation to the positive discount rate is that individuals tend to choose some good that is closer in the future simply because they are uncertain whether some good that lies in the more distant future will be worth as much to them (in relation to their future preferences), or indeed if they will even be around to reap the benefits of that good. Hence, for a welfare-maximising individual to account for some uncertainty it would be rational to discount future benefits at least by the probability of being alive in later periods. Uncertainty is not necessarily related only to mortality risk; it can also be linked to potential changes in taste. Some events might be more or less likely to be enjoyable in the future. Going to the circus is probably more enjoyable as a child than as an adult, eating is probably more enjoyable when you are hungry, and so on. However, many of these changes in taste are difficult to anticipate.

The probability argument

The main criticism against uncertainty as a legitimate basis for discounting is that uncertainty does not necessarily follow remoteness in time. Derek Parfit (1984) proposes what he calls *probabilistic discounting*, arguing that we should not discount over time but over probabilities. If we discount over time we misrepresent one of the core reasons for discounting, suggesting that we attach lesser importance to what happens in the future while in fact we attach lesser importance to what is uncertain.

² The work by Shane Fredrick (2003) being the exception.

³ See Gravelle & Smith (2001) for more detailed discussion.

We ought to discount those predictions that are more likely to be false. Call this a *probabilistic discount rate*. Predictions about the further future are more likely to be false. So the two kinds of discount rate, *Temporal* and *Probabilistic*, roughly correlate. But they are quite different. It is therefore a mistake to discount for time rather than probability. One objection is that this mis-states our moral view. It makes us claim, not that more remote bad consequences are less likely, but that they are less important. (Parfit 1984, p. 482)

What Parfit suggests is that if uncertainty is a main reason for discounting future values, then the discount rate should be differentiated depending on the risk involved in the project – not on when we can reap the benefits of some (present) cost. For health this would mean the risk of not living in the next period, which has several immediate implications. Since the risk of not living in the next period steadily increases as a person ages, this would imply a significant higher rate when discounting health benefits for older persons compared to younger persons. Similarly, disadvantaged groups with a lower probability of surviving should discount at a higher rate with probabilistic discounting.

4. Diminishing marginal utility and growth as a basis for discounting health ($\epsilon_x g_x$)

The other part of Ramsey's formula for the social discount rate stems from the combination of growth and diminishing marginal utility. In a growing economy, people will generally be better off in the future since they will have more of some relevant good. Diminishing marginal utility simply states that utility derived from a fixed unit of good will decrease as the amount of this specific good increases. Consequently, in the presence of growth, society should place higher value on present consumption, and this should be done through positive discounting. What should be noted is that marginal utility does not really say anything about the value of some good *per se*, but rather about its relative value. The first Euro is not of any less value than the 1 millionth Euro, other than in relation to some person (who already has 1 million dollars). The diminishing marginal utility element of the argument can be viewed as a link between time preference and growth, since it requires both growth in the relevant good

and altered individual preferences towards a fixed unit of that good resulting from this growth.

The basic rationale for why growth constitutes a legitimate basis for discounting future values is a variant of the classical economic notion of opportunity costs. If it is possible for society to transform resources available today into greater amounts in the future through savings or investments, this constitutes a social opportunity cost. Such a social opportunity cost related to early consumption should be accounted for through discounting when conducting CEA. From an individual's perspective, the opportunity cost associated with consuming something now rather than in the future would be the intertemporal arbitrage that he or she would make through investing a given unit of resource as wisely as possible. From a societal perspective, however, the opportunity cost regarding intertemporal choices becomes equivalent to the growth rate in the economy, since this is the rate at which society would have been able to transform resources had they not been consumed during an earlier period.

If the argument of growth and diminishing marginal utility is applied to health, and if a future generation is healthier than the current one, this should be accounted for simply because the future generation will have more of this good (health) and will value increased health less. But it is also possible, as Gravelle and Smith (2001) argue, that the relative value of health compared to other goods might increase over time as society becomes wealthier. This should, in such a case, be accounted for by applying a lower discount rate to health effects compared to monetary effects when discounting future values – which leads us to the unsettled issue of whether or not health and cost should be discounted at differentiated or uniform rates. Since the growth argument states that in the presence of economic growth a dollar tomorrow is worth less than a dollar today, many economists have, by extension, argued that anything convertible into money (and back again) should be discounted at an equal rate. Those who favour uniform discounting often refer to the quotation by Fuchs and Zeckenhauser (1987) that “self-respecting economics should not/.../ use different rates because it is health that is being valued”. The logic underlying such a statement is that the health effects in CEA must always be valued, implicitly or explicitly, in monetary terms. Hence, it would be inconsistent and misleading not to discount at a rate that is equal for health and monetary outcomes. The reason for this can be ascribed to two closely related arguments for uniform discounting: *the consistency argument* and the *postponing paradox*.

The consistency argument

The *consistency argument* was first presented in an article by Weinstein and Stason (1977) and is still used as one of the major arguments for advocating a uniform discount rate for costs and effects. Weinstein and Stason argue that inconsistencies arise when using different rates to discount costs and effects. They state that:

The reason for discounting future life years is precisely that they are being valued relative to dollars and, since a dollar in the future is discounted relative to a present dollar, so must a life year in the future be discounted relative to a present dollar /.../ It is the discounting of dollar costs, and the assumed steady-state between dollars and health benefits, that mandates the discounting of health benefits as well as dollars. (Weinstein & Stason 1977, p. 720)

Insert Box 2

In the argument forwarded by Weinstein and Stason, there is an exact equivalence regarding the CE ratio between the two programmes presented in Box 2. This is underlined by the assumption of a steady state relation between cost and health benefits, which implies that purchasing health benefits for dollars does not change over time. Consequently, if decision makers want to give equal priority to programmes that cost the same and yield the same benefit in a particular year (regardless of year), this can only be achieved if costs and effects are discounted at the same rate. For the purpose of remaining consistent, and to avoid confusion, health and costs should therefore be discounted at a uniform rate according to Weinstein and Stason. Alan Williams elucidates this reasoning further:

Because it is possible, at the margin, to transform health into wealth, and vice versa, at any point in time, and since “wealth” is (ideally) allocated through time with reference to the rate of social time preference, then it would be inconsistent to apply a different rate of discount to “health” from that being applied to “wealth” (Cited in Gold 1996, p. 277)

Consequently, the consistency argument hinges on the crucial assumption that our relative valuation of health and money remain stable over time. van Hout (1998) has argued that there is no obvious reason for this assumption to be true. Instead, he has suggested that discount rates for costs and effects should be based on different growth rates and the marginal utilities related to this growth. In other words, if the value of health effects grows over time at a different rate from our valuation of wealth, this should be incorporated in the evaluation method by using differentiated discount rates. If we abandon the constant relationship between monetary valuation and health states, the argument of consistency disappears.³

The postponing paradox

Keeler and Cretin (1983) first presented the *postponing paradox*, which is based on the fact that if one applies lower discount rates for health outcomes than for costs, postponing any given programme will improve its CE ratio. Hence, for any attractive programme there is always a superior delayed programme that should be funded first. Consequently, no programmes would ever be implemented because even the initially worst postponed programme will appear better than the best immediate programme.

Insert Box 3

Gravelle and Smith (2001) have questioned the postponing paradox as a valid argument for not using differentiated discount rates since it does not recognise “... that the CEA decision rule is inherently incomplete and cannot cope with issues of timing decisions”. After discussion, van Hout (1998) acknowledges this line of argument as correct in principle. He agrees that calculations based on differentiated rates (like that in Box 3) might lead to counterintuitive results. However, he argues that these theoretical conclusions have no real bearing in practice since the type of decision problem that policymakers face is not whether programme A should be implemented now or next year, but rather whether to implement programme A instead of an already existing programme B. Hence, if decision makers have no scope to delay current budgets to future periods, selecting programmes from a ranked list cannot be blocked.

³ See Gravelle & Smith (2001) for more detailed discussion.

5. What is the legitimate basis for discounting future values in CEA, and how might it influence valuation of health and wealth differently?

It is evident from the empirical literature that most people care less about their future health than they do about their current health.⁴ From a welfare-maximising perspective it is also obvious that individuals have rational reasons for doing so, especially given the uncertain nature associated with future events. Still, this does not offer sufficient grounds to proclaim that society as a whole should adopt a short-sighted attitude towards the future as regards health outcomes. In fact, if we assume that it is rational from the perspective of the individual, it could still be irrational from the perspective of society. Moreover, it might also be morally impermissible. In this section, I will return to explore the first two questions raised in the introductory section: i) Which of the motives presented in the preceding sections constitute morally legitimate reasons for discounting future values in CEA?; and ii) Do these motives apply differently when making intertemporal valuations in comparing health to wealth?

Before approaching these questions it is crucial to recognise some of the conceptual ambiguity that could potentially obstruct understanding core issues related to assessing the legitimate basis for discounting health. As indicated above, it is especially important to note that arguments related to discounting do not necessarily apply equally to intrapersonal discounting and intergenerational discounting. Issues related to the legitimacy of discounting in the *intrapersonal* setting stem principally from questions relating to rationality and justifiable cases of paternalism. Issues related to the legitimacy of discounting in the *intergenerational* setting stem principally from questions relating to fairness and limits of democracy. Recognising this conceptual ambiguity, and raising awareness about it, should help improve the quality of the discourse on discounting. Much of the literature on discounting fails to recognise the distinction between intrapersonal and intergenerational discounting, and authors frequently use the same concepts in referring to different underlying phenomena. In the economic literature, discounting is most often discussed in an intrapersonal context and justified by appealing to concepts of growth and opportunity costs. In the philosophic literature, however, discounting tends to be discussed mainly in an

⁴ Even though the empirical literature on time preferences and health show a perplexing range for annual discount rates, ranging from negative values to infinity. For a systematic review of the empirical literature on time preferences for health, see Cairns (2006).

intergenerational context. Much of the disagreement concerning discounting, especially in the discourse between economists and philosophers (where typically philosophers find it impermissible while economists find it inevitable), seems to stem from differences in the discounting perspectives they apply. Misunderstanding is the inevitable result.

The pure preferences effect is the first argument advanced to explain why individuals' time preferences should ascribe lower value to the future. It could be argued that a societal discount rate in a modern democratic society should reflect the preferences of its citizens, and if they for whatever reason feel that the present is more important than the future, this is the attitude that society should adopt (e.g. Marglin 1963). This type of appeal to democracy, however, lends little support to the claim that pure time preference is a morally permissible basis for intergenerational discounting, since the future generations affected by present choices will have no vote. When discussing time preferences we are generally referring to individuals' personal preferences within their own lifespan. When incorporating the pure time preference effect as a legitimate basis for social discounting, however, we also assert that the welfare of future generations is less important than that of the present generation. Intuitively this seems impermissible from a fairness perspective. Few people would, for example, find it morally permissible to apply a spatial discount rate that assigns a lower value to benefits that accrue on individuals at a further distance, based on remoteness in miles. The fact that current generations favour themselves over future generations, and are therefore willing to impose greater burdens on future generations to avoid lesser burdens on their own, cannot be considered an ethically valid basis for discounting – particularly as regards policymakers who are supposed to adopt an impartial stance and protect the interests of all individuals in society, current and future generations alike. Consequently, there is a strong argument against populist forms of democracy as processes for deciding social discount rates. Democracy is, in essence, a system for respecting citizens' political preferences. Contemporary democracy is insufficiently sensitive to the concerns of future generations if individuals in general tend to favour themselves over future generations, which seems highly plausible. Considering that the life span of governments is usually much shorter than the life span of individuals, this would entail directing attention to decisions that have immediate and visible effects as opposed to more long-term effects, implying a high discount rate. A potential way for policymakers to resolve the conflict between democracy and the concern for future generations is to increase the technical complexity of the issue, preventing people from understanding the serious practical implications that the social discount rate can have. I would argue that this is what has

happened in the case of discounting. Discounting is often portrayed (falsely I would say) as a highly technical matter, and this has had obvious consequences for transparency and the value judgements it involves.

Still, some could object that many policy decisions affect only the current generation, and in such scenarios it cannot be considered morally impermissible to favour the present over the future. This objection has some bearing and moves us towards the issue of irrationality and paternalism in discounting – an issue that I will return to in the next section. For the moment, I will assume that the behavioural effects that together form individual time preferences as a basis for discounting future health constitute rational reasons for discounting future health from an intrapersonal perspective. However, when comparing the pure time preference effect between health and wealth in the intrapersonal context, it should be noted that although most individuals reveal positive time preferences in empirical studies, this is not necessarily true for all individuals or all choices (e.g. Van der Pol and Cairns 2000). I would argue that pure time preferences for health are more likely to be negative compared to monetary values, since individuals might prefer ill-health sooner rather than later in order to eliminate dread – just as some individuals prefer to postpone future monetary benefits in order to derive pleasure from anticipation. Consequently, there is no *prima facie* reason for assuming that the pure time preference effect should apply equally to health and wealth in the intrapersonal context.

The uncertainty effect is a second argument advanced to explain why individuals' time preferences should ascribe lower value to future outcomes. From an intrapersonal perspective it could be argued that given the brevity and uncertainty of human life, it is rational to attribute less value to future events than to immediate ones if they have a different probability of occurring. The more distant an event, the less certain we are likely to be about its value to us. This is clearly a relevant factor to consider when making intertemporal decisions. Analogously, when making intergenerational policy recommendations, society ought to take into account that predictions about the future are less likely to be true. However, I would argue that the proper way of dealing with uncertainty over time in CEA is not through the social discount rate, regardless of perspective. When conducting CEA, we typically assume that measures of uncertainty are incorporated through life tables and other parameters of uncertainty regarding the future. For that reason, taking uncertainty into account when assigning a societal discount rate would present a case of double counting. Another strong reason for not dealing with uncertainty through the social discount rate is that uncertainty

often correlates poorly with time and is highly unstable across different groups of individuals. For example, the mortality risks associated with the old or disabled are much higher than for the young and healthy. Although we may want to consider this when making intertemporal policy decisions, including uncertainty as a basis for discounting would mis-state our moral view – it makes the statement that remote events and older people are less important, but what we want to state is that those events are less likely.

As with pure time preferences, there seems to be no *prima facie* reason for assuming that the uncertainty effect should apply equally to health and wealth in the intrapersonal context. Uncertainty associated with future monetary outcomes involves more than the risk of early death. There is also the risk (or opportunity) that individuals might live an unexpectedly long life, in which case it would be rational to save benefits needed for future consumption. Consequently, the uncertainty effect, as regards future wealth, might influence the discount rate not only in the direction of putting less value on future goods, but also by putting higher relative value on future goods. In addition, at least two substantial differences regarding uncertainty are found between health and wealth, which could cause them to differ further. First, although individuals cannot insure themselves against their own death, they can insure themselves against future loss of wealth. This makes the uncertainty effect potentially higher for health compared to wealth. Second, regarding wealth, some utility arises from bequests when a person suffers an early death. This bequest effect is not applicable to health, and could potentially increase the uncertainty effect in health compared to wealth.

When discussing the role of uncertainty in an intergenerational context, we should note that uncertainty is likely to have much less impact at the societal level compared to the individual level. Societies are not mortal in the same sense that individuals are. Although, potential natural catastrophes and nuclear wars may invoke some uncertainty regarding survival of society at large, this is very different from the individual perspective where death is an inevitable fact, and the risk of being absent in the next time period steadily increases over time. Consequently, the argument for discounting arising from uncertainty is not as valid when applied to social decision making in contrast to individual decision making.

The combined argument from *growth* and *diminishing marginal utility* possibly presents the strongest reason why society should apply a positive discount rate in an intergenerational context. However, this hinges on the assumption of continuous *growth* in the relevant good, implying that we derive more satisfaction from any given unit of a good now rather than later, *if* we have more of it in the future. To avoid redistribution from a present that is relatively poor, to a future that is relatively rich, one can make a strong fairness argument for encouraging intergenerational discounting in the presence of growth. However, in the absence of growth this fairness argument disappears. In the case of recession, the logic of growth and diminishing marginal utility would, on the contrary, imply that we should redistribute from the relatively rich present to the relatively poor future. Hence, if we expect individuals to be worse off in terms of the relevant good in the future (*ceteris paribus*), society should assign a negative discount rate if the rate is to be based solely on the marginal utility effect.

As with pure time preferences and uncertainty, there appears to be no *prima facie* reason to assume that the combined argument from growth and diminishing marginal utility should apply equally to health and wealth in an intrapersonal context. Regarding health, most people expect to be less healthy as they grow older due to natural ageing. In such a scenario the effect of diminishing marginal utility would imply that future health benefits will be of higher relative value to individuals since we can assume that people generally become less healthy the older they get. Regarding wealth, in contrast, most individuals do not anticipate becoming less wealthy the older they get. Consequently, in the intrapersonal context the combined argument from diminishing utility and growth moves in opposite directions for health compared to wealth. For intergenerational discounting, this difference between health and wealth seems less obvious since we are dealing with the future life of others and not our own stock of health or wealth.

The time-relative relationship between health and wealth

Justifications for applying a uniform discount rate for health outcomes and costs in CEA are usually underpinned by a consequentialistic rationale, i.e. that the consequences of not discounting or discounting at a differential rate would generate unacceptable outcomes. However, it is important to notice that neither the time paradox nor the consistency argument present arguments for discounting future health *per se*, but arguments why health should be discounted at the same rate as cost. Both the consistency argument and the postponing

paradox present arguments implying that the use of differential discounting would lead to policy decisions considered to be undesirable and counterintuitive. The reason is that outcomes from CEA would be such that beneficiaries, who are identical in all relevant aspects except for their position in time, would be assigned unequal CE ratios when present values are calculated.

Both the postponing paradox and the consistency argument rely on the crucial assumption of a steady state relationship between money and health – i.e. the constant value of a unit of health (i.e. QALY or DALY). This assumption appears to hinge on convenience rather than reflection. Whether or not the growth rate for health and wealth is the same is an empirical question for which we are unable to give any clear-cut answer. But given the special nature of health compared to other *goods*, it does not seem unreasonable to believe that the growth rate for health could differ substantially from the growth rate elsewhere in the economy. And if the relationship between health and wealth is indeed unstable over time, and we fail to take this into account, Pareto-dominated health programmes might be recommended for implementation on the basis of CEA. Hence, if the relative valuation of health increases over time, society should correct for this through differential discounting to reach proper conclusions regarding the relative welfare improvements of different health care interventions. By doing so we may find that we have to re-evaluate the normative consequential framework underpinning CEA.

6. Normative relevance of individuals' time preferences

The normative relevance of empirical insights regarding elicited time preferences crucially depends on whether or not they address the question they are intended to inform. Harsanyi (1955) suggested that individuals have two sets of preferences – one set that reflects individuals' preferences for their own wellbeing and a second set where individuals (in the role of citizens) take social considerations into account. These two sets of preferences may come into conflict, which complicates the concept of welfare maximisation. More specifically, it is highly likely that individuals' time preferences for their own health differ from the preferences they (in the role of citizens) have for social intertemporal choices of health. Hence, just as it was important to recognise the conceptual distinction between intrapersonal and intergenerational discounting when assessing the legitimate basis for

discounting, it is necessary to recognise the distinction between *personal* and *societal time preferences* when discussing the normative relevance of individuals' time preferences in the context of the social discount rate.

The basic difference is that personal time preferences refer to those preferences elicited by asking individuals to make intertemporal choices regarding themselves, while social time preferences involve making intertemporal choices from the perspective of society at large.⁶ Personal time preferences are well in accordance with the individualistic foundations of welfare economics. The traditional wisdom amongst welfare economists has been that individuals have clear, well-defined preferences, and societal value is maximised when aggregating these personal preferences (utilities) within a community. In recent years, however, interest has increased amongst economists for eliciting individuals' societal preferences (Nord et al., 1999). The rationale for this is that personal preferences do not incorporate the distributional concern that individuals might have for others, and when allocating resources within a policy context such distributional concerns have substantial relevance. Given that CEA primarily aims at informing decision makers when dealing with allocation at a population-level, it seems illogical that personal time preferences regarding their own health should be used as input. Instead, it seems reasonable to argue that individuals' *societal* preferences are the ones most relevant when discussing the basis for *social* discount, since choices between competing health care programmes are *social* choices, not private ones.

Preference failures and paternalism

So far I have dismissed time preference as a morally permissible basis for intergenerational discounting since it would lead to unfair allocations between generations. But let's return to the objection that many policy decisions are intragenerational, i.e. they affect only the current generation. In intragenerational discounting, it is difficult to see why favouring the present over the future should be considered morally impermissible. Libertarians have long argued that although individuals' choices will not always maximise their own wellbeing, they remain the best judges of how to satisfy their own preferences. From a general perspective it is hard to disagree with this claim – why should the state or someone else be a better judge of my

⁶ Note that it is possible to combine these two perspectives when eliciting preferences. See, e.g. Dolan et al. (2003)

welfare? Indeed, one of the underlying value judgements for achieving Pareto optimality is that individuals are the best judges of their own welfare. However, from a less general perspective it is obvious that this claim is not true in all cases. Paternalism might be justified in some non-standard cases.

Paternalism is generally viewed as state action directed towards the individual against their will, motivated by a claim that the affected person will be better off or protected from harm (Dworkin, 2010). Consequently, the issue of paternalism requires careful balancing of which value (i.e. respecting subjects' autonomy, or increasing subjects' well-being) is more important in the case at hand. A necessary (although not necessarily sufficient) prerequisite for justifying paternalism is that the state in fact has a better conception of what constitutes welfare-promoting actions than what stems from individuals' basic aims and values. In economics, preferences based on insufficient information or inadequate facts are generally viewed as market failures where society should intervene. For instance, if a harmful drug is put on the market, the state should take action to prevent individuals from unknowingly taking it. The inefficient outcome of taking harmful drugs would otherwise be a result of individuals' inability to assess how to promote their own welfare. This type of paternalism, where the individual is not regarded as a fully capable decision maker, is generally considered justifiable even among libertarians. In the case of positive pure time preferences, however, we are dealing with individuals who act in an imprudent or shortsighted manner while fully understanding the future consequences of their behavior. In these cases, state intervention is harder to justify than in the case of market failure. Consequently, we are confronted with the question when, if ever, is it justifiable for policymakers to override individuals' fully informed preferences?

From an economist's perspective, the obvious answer would be that it is justifiable to override an individual's fully informed preferences when it increases the individual's welfare. Hence, we need to identify those preference failures (as opposed to market failures), where individuals sufficiently fail to act in their own best interest. For positive pure time preferences we find two typical categories of preference failures that might help justify paternalistic state interventions to correct for the individual's fully informed preferences. I will label these two categories *myopic* and *acritic* preference failures.

The first category, *myopic* preference failure, relates to a cognitive inability to compute information adequately to make choices that maximise the welfare of the individual. This typically refers to cases where the logical or causal connections between choice and outcome are difficult to make (Goodin 1993), or alternatively to cases where the quantity of information is too great for the individual to assess (New, 1999). The information itself might, however, be perfectly adequate. Judgments about probabilities and likely future outcomes of actions are particularly common in this category of preference failure. The work of Tversky and Kahneman (1982) has shown empirically how people with perfect information commonly display inadequate reasoning power to interpret and calculate the implications of certain statistical or probabilistic situations. This further strengthens the argument that individuals' assessment of uncertainties associated with their own future should not be a basis for the social discount rate.

The second category, *acratia* preference failure, relates to reduced willpower. This refers to instances where individuals act against their own better judgment and fail to do what is in their long-term best interest. It could, for example, include cases where the individual often lacks the ability to emotionally assess the force of information – knowing the facts, but not feeling the force of an outcome psychologically. Hence, even though these individuals might make good judgments as to what is necessary to optimise welfare, they still frequently choose a different course of action. For example, people who maintain an unhealthy lifestyle often assert that they would like not to engage in unhealthy activity, but find it impossible to stop doing so due to an insufficient capacity to self-regulate. Another example of acratia preference failure would be individuals that spend all their money in the pub rather than saving for future necessities, despite being fully aware of the negative long-term effects this will have on their welfare. An analogous way of explaining acratia preferences would be to say that individuals have 'preferences for preferences' and acrasia occurs when their second-order evaluative desires differ from their first-order motivational desires (Frankfurt, 1971). Thus, individuals wish that their preferences would be different from what they are, and in such cases paternalism could be justified on the grounds that the state is merely respecting a person's own higher preferences.

In many countries, this line of argument is commonly used to motivate mandatory public programmes to ensure individual saving for retirement and avoid procrastination, or to deny an addicted person's preferences once he or she has checked into a rehab center.

The main difference between myopic and an acratia preference failure is that myopia involves a systematic tendency to overestimate the value of earlier benefits in comparison to later benefits. A myopic individual sincerely believes that he or she is making a correct choice when choosing a smaller early benefit instead of a larger later benefit. But the individual's tendency of imprudence leads to choices that give inadequate weight to future events. An acratia individual is well aware that he or she will prefer a later benefit, but still choose the early benefit rather than the later one. From an economist's perspective, both myopic and acratia preferences are by definition irrational since they are associated with regret in the sense that individuals wish they had chosen differently retrospectively.

Acrasia can be further linked to Parfit's reductionist account on personal identity, which provides a plausible explanation for why this seemingly irrational behaviour is, in fact, rational. Parfit's claim is that although being fully informed, individuals might be truly *unable* (psychologically) to resist the temptations of choosing immediate benefits rather than greater benefits later even though their judgment says that this is not the best option. According to Parfit, a person may be truly unable to act in coherence with his or her fully informed judgment since the psychological link with the future self is too weak, making such preferences fully rational. However, this has no immediate implications for view presented here of when paternalism is justifiable since the existence of acratia preferences would be categorised as market failures (negative externalities) due to the inability of some persons affected by the decision (future selves) to participate in the decision-making process. Further, with Parfit's account of personal identity, acrasia transcends from being an issue of irrationality to being an issue about fairness. Arguably, fairness would be a stronger justification for overriding informed preferences than irrationality would.

In cases where individuals' preferences suffer from myopia or acrasia, society occasionally would have an advantage over the individual in appreciating what is in that individual's best interests in the long term. Decisions that involve weighing future benefits against present ones offer both a significant temptation to satisfy immediate wants above those with long-term benefits (acrasia) and unusual technical difficulties involving probabilities that make it hard for individuals to compute the actual trade-off (myopia). By extension, if policymakers have the knowledge that most individuals would like to change their choices *ex post* it would not be a case of overriding individuals' informed preferences. Instead, it would merely be a case of

giving some of the person's own preferences less consideration than others in attempting to prevent individuals from being caught short by their own inconsistent patterns of pure time discounting. Hence, this softer kind of paternalism aims primarily at helping people make choices that are congruent with their long-term interests.

To conclude, a person's immediate preference satisfaction does not necessarily coincide with a person's good when it comes to intertemporal choices. Hence, some laundering of preferences characterised by myopia and acrasia intuitively seems both necessary and legitimate when discounting future values (for both health and monetary outcomes) in a welfare-promoting society. We must acknowledge, however, that the idea of positive pure time preferences is not self-evidently connected to market or preference failures. Just as people might prefer red wine to white wine, or a Big Mac to a Whopper, they could have a similar preference for the present compared to the future. Two aspects related to today's preferences about future events are particularly relevant when assessing the value of the future state of affairs. The values placed on anticipation and dread are experienced in the present, and I would argue that these are values that do not stem from myopia or acrasia. We often postpone things of a good nature to experience good value from anticipation. Likewise, at times we choose to bring forward unpleasant events, not because we expect them to be less unpleasant, but because we want to avoid dread. For intrapersonal discounting, individuals' time preferences could potentially constitute a legitimate basis for discounting future events once such preferences have been laundered from myopic and acratice elements.

Final non-utilitarian remarks about cost-effectiveness analysis and discounting

This article has been written mainly from a traditional utilitarian perspective. As an economist, I feel most comfortable within this type of framework since maximising welfare offers a straight-forward solution regarding which policies to recommend. However, I would argue that it is worth stepping outside of the traditional framework of utilitarianism when considering the issue of discounting future health. If the societal objective in making intertemporal choices is solely to maximise net welfare over time, this could easily lead to a very unequal distribution across generations, forcing one generation to make excessive sacrifices for the sake of future generations. For example, it would be unjust to require the current generation to commit itself to large-scale medical experiments with severe immediate consequences for the purpose of benefiting future generations – even though this would

increase the net sum of welfare over time. No generation can be morally obligated to exceed a certain level of sacrifice for the sake of future generations. This is an issue that the utilitarian framework for discounting within CEA does not consider – but then, CEA in general incorporates little or no consideration for how to distribute burdens or benefits. Cost-effectiveness analysis does not reflect every aspect of importance in the allocation of scarce health care resources, which could be considered a weakness if one chooses to view CEA as a decision rule rather than decision input. In my opinion, the future credibility of CEA as valid input to intertemporal public policy hinges on its ability to be transparent in the value judgments it incorporates. Decisions regarding allocation of health care resources (temporal or not) involve balancing of conflicting moral values on which reasonable people will disagree. Therefore, rather than trying to incorporate distributional and other non-utilitarian concerns into the CEA, we should keep these concerns separate while concurrently informing decision makers in a manner that makes inevitable moral judgments as transparent as possible.

Acknowledgements

I am grateful to Greg Bogner, Dan Brock, Alex Broadbent, Dan Wikler, Lennart Nordenfelt, Matthew Frank, Adriane Gelpi, Elizabeth Fenton, Carl Hampus Lyttkens, and Lisa Furberg for helpful comments and suggestions on earlier drafts of this article.

References

- Baumol, W. J. 1968. On the social rate of discount. *American Economic Review* 81: 334-340.
- Brouwer, W. B. Niessen, L. W. et al. 2005. Need for differential discounting of costs and health effects in cost effectiveness analyses. *BMJ* 331: 446-448.
- Cairns, J. 2006. Developments in discounting: With special reference to future health events. *Resources and Energy Economics* 28: 282-297.

- Dolan, P., Olsen, J. A., Menzel, P. & Richardson, J. (2003) An Inquiry into the different perspectives that can be used when eliciting preferences in health. *Health Economics*, 12: 545-51.
- Dworkin, G. (2010) Paternalism. In Zalta, E. N. (Ed.) *The Stanford Encyclopedia of Philosophy (Summer 2010 Edition)*.
- Frankfurt, H. G. (1971) Freedom of the will and the concept of a person. *Journal of Philosophy*, 68: 5-20.
- Fredrick, S. (2003) Time Preference and Personal Identity. In Loewenstein, G., Read, D. & Baumeister, R. F. (Eds.) *Time and decisions*. New York, Russel Sage Foundation.
- Fuchs, V.R. and Zeckenhauser, R. 1987. Valuing health- a “priceless” commodity. *American Economic Review* 77: 263-68
- Gold, M. R. 1996. *Cost-effectiveness in Health and Medicine*. New York: Oxford University Press.
- Goodin, R. 1993. Democracy, preferences and paternalism. *Policy Sciences* 26: 229-47.
- Gravelle, H. and Smith, D. 2001. Discounting for health effects in cost-benefit and cost-effectiveness analysis. *Health Economics* 10: 587-99.
- Harsanyi, J. 1955. Cardinal Welfare, individualistic ethics and interpersonal comparison of utility. *Journal of Political Economy* 63: 309-321.
- Kahneman, D. and Tversky, A. 1979. Prospect theory: An analysis of decisions under risk. *Econometrica* 47: 313-327.
- Keeler, E. B. and Cretin, S. 1983 Discounting of life-saving and other nonmonetary effects. *Management Science* 29:300-306.
- Krahn, M. and Gafni, A. 1993. Discounting in the economic evaluation of health care interventions. *Medical Care* 31: 403-18.

- Marglin, S. A. (1963) The social rate of discount and the optimal rate of investment *Quarterly Journal of Economics* 77: 95-111.
- New, B. 1999. Paternalism and public policy. *Economics and Philosophy* 15: 63-83.
- Nord, E., Pinto, J., Richardson, J., Menzel, P. & Ubel, P. (1999) Incorporating societal concerns for fairness in numerical valuations of health programmes. *Health Economics*, 8: 25-39.
- Olsen, J. A. 1993. On what basis should health be discounted? *Journal of Health Economics* 12: 39-53.
- Parfit, D. 1984. *Reasons and Persons*. Oxford: Oxford University Press.
- Pigou, A. C. 1932. *The Economics of Welfare*. London: Macmillan and Co.
- Ramsey, F. P. 1928. A mathematical theory of saving. *The Economic Journal* 38: 543-549.
- Rawls, J. 1973. *A Theory of Justice*. Cambridge, Massachusetts; Belknap Press.
- Samuelson, P. (1937) A Note on Measurement of Utility. *Review of Economic Studies*. 155-161.
- van Der Pol, M and Cairns, J. 2000. Negative and zero time preference for health. *Health Economics* 9: 171-175.
- van Hout, B. A. 1998. Discounting costs and effects: a reconsideration. *Health Economics* 7: 581-94.
- van Liedekerke, L. 2004. John Rawls and Derek Parfit's critique of the discount rate. *Ethical Perspectives* 11: 72-83.
- Weinstein, M.C and Stason, W.B. 1977. Foundations of cost-effectiveness analysis for health and medical practices. *NEJM* 296: 716-721.

Box 1 Discounting in cost-effectiveness analysis

The main purpose of cost-effectiveness analysis (CEA) is to provide information about costs and effects summarised in a so-called cost-effectiveness ratio (CE ratio), e.g. cost per additional quality-adjusted life-year (QALY) or cost per additional life-year gained. This CE ratio can be seen as a price estimate for achieving a certain outcome that together with other considerations may help guide policymakers in making decisions on whether or not to implement a new medical treatment (or exclude an old treatment). To do this a present value of future costs must first be calculated using a discount rate. In algebra this can be defined as:

$$PV_C = \sum_{t=0}^{t=\infty} \frac{C_t}{(1+i)^t}$$

Where C_t measures the cost in year t after the start of the programme and i is the discount rate. Example: A cost of \$100 000 is supposed to occur 5 years ahead. If the discount rate for cost (i) is set at 3%, the present value of this sum will be $\$100\,000/1.03^5 = \$86\,261$. An almost identical formula can be used in calculating the present value for a stream of health effects

$$PV_H = \sum_{t=0}^{t=\infty} \frac{H_t}{(1+r)^t}$$

Here H_t measures the health effect in year t after the start of the programme and r is the discount rate. Example: A preventive health care programme will yield a health benefit of 10 QALYs 3 years ahead. If the discount rate for health effects (r) is set at 3% the present value of this health benefit will be $10\text{ QALY}/1.03^3 = 9.15\text{ QALYs}$. The CE ratio for a programme is then calculated by dividing PV_C (\$86 261) with PV_H (9.15 QALYs). Hence the CE ratio can be thought of as the average cost per unit of effectiveness (\$9427 per QALY in the example). The programme that is the most cost effective is the one that has the lowest average cost per unit of effectiveness. The higher the discount rate applied, the lower the priority given to outcomes distant in time.

Box 2 Numerical example: Consistency argument

Consider the following health programmes:

Programme A costs \$30 000 this year and has an effect of 1 QALY this year.

Programme B costs \$30 000 in year 40 and has an effect of 1 QALY in year 40.

If we assume that the monetary valuation of health outcomes remains stable over time, as the consistency argument does, it is easy to see that these two programmes should receive equal priority since their CE ratio is identical. Hence, if we want programmes A and B to receive equal priority, the discount rate of costs (r) should equal that of effects (i), since only then does the CE ratio of B $((\$30\,000/(1+r)^{40})/(1/(1+i)^{40})) = \$30\,000/1$ equal that of A $(\$30\,000/1)$.

Box 3 Numerical example: Postponing paradox

Consider the same health programmes as in Box 2:

Programme A from Box 1 has a cost-effectiveness ratio of \$30 000 per QALY. Assuming a 5% discount rate for costs and a zero discount rate for effects, postponing the programme by 1 year would improve its cost-effectiveness ratio to $\$30\,000/(1.05)^1 = \$28\,571$ per QALY. Further postponement would again improve the ratio, since the cost component declines more rapidly than the effect. The paradox is therefore that a lower discount rate for effects makes infinite postponement theoretically optimal, but it is practically undesirable.

PAPER

III



ORIGINAL ARTICLE

Incentivizing deceased organ donation: A Swedish priority-setting perspective

FAISAL OMAR*, GUSTAV TINGHÖG* & STELLAN WELIN

Department of Medical and Health Sciences, Linköping University, Linköping, Sweden

Abstract

Aims: The established deceased organ donation models in many countries, relying chiefly on altruism, have failed to motivate a sufficient number of donors. As a consequence organs that could save lives are routinely missed leading to a growing gap between demand and supply. The aim of this paper is twofold; firstly to develop a proposal for compensated deceased organ donation that could potentially address the organ shortage; secondly to examine the compatibility of the proposal with the ethical values of the Swedish healthcare system. **Methods:** The proposal for compensating deceased donation is grounded in behavioural agency theory and combines extrinsic, intrinsic and signalling incentives in order to increase prosocial behaviour. Furthermore the compatibility of our proposal with the values of the Swedish healthcare system is evaluated in reference to the principles of human dignity, needs and solidarity, and cost effectiveness. **Results:** Extrinsic incentives in the form of a €5,000 compensation towards funeral expenses paid to the estate of the deceased or family is proposed. Intrinsic and signalling incentives are incorporated by allowing all or part of the compensation to be diverted as a donation to a reputable charity. The decision for organ donation must not be against the explicit will of the donor. **Conclusions:** We find that our proposal for compensated deceased donation is compatible with the values of the Swedish healthcare system, and therefore merits serious consideration. It is however important to acknowledge issues relating to coercion, commodification and loss of public trust and the ethical challenges that they might pose.

Key Words: *Cost effectiveness, deceased organ donation, ethics, incentives, priority setting*

The impetus for this paper partly arose from an interview with a Swedish organ procurement coordinator during a project exploring priority setting in kidney transplantation (11 May 2009). During the dialogue the coordinator recalled an incident that occurred in her procurement catchment some years back. A sudden spike in the number of calls to her office from individuals volunteering to donate their bodies as cadavers for scientific purposes roused her suspicions and prompted her to investigate their motivation. It became clear rather quickly that there had indeed been motives other than altruism. Callers had been partly motivated by an advert that appeared in a local seniors' newsletter pledging an urn to store the

ashes of all those who donate their bodies for scientific purposes. We began to reflect on the kind of compensation that would be both appropriate and effective, and what challenges this presents to the values of Swedish health care.

Introduction

Discourse on healthcare priority setting is rooted in the recognition that limits must be set in the face of resource scarcities [1,2]. While financial constraints are the most prominent reasons for priority setting, non-financial factors also lead to limitations in healthcare provisioning as clearly demonstrated by organ scarcities in the field of transplantation.

*Faisal Omar and Gustav Tinghög share first authorship on this paper.

Correspondence: Faisal Omar, Division of Health and Society, Department of Medical and Health Sciences, Linköping University, SE-581 83 Linköping, Sweden.

E-mail: faisal.omar@liu.se

(Accepted 26 October 2010)

© 2010 the Nordic Societies of Public Health
DOI: 10.1177/1403494810391522

Organ transplantation has over the past few decades developed into the treatment of choice for end stage organ diseases; unfortunately demand for new organs outstrips supply in nearly all parts of the world [3,4]. Health care must prioritize among potential recipients as a means of rationing the limited supply of organs. The result can be devastating for those who do not receive a transplant as they face the prospect of death. A very unfortunate consequence of the unmet demand has been the proliferation of organ black markets, which callously jeopardize the health of poor organ vendors in the developing world [5,6]. The 2008 Declaration of Istanbul on Organ Trafficking and Transplant Tourism recognized the responsibility of individual countries to reach organ self-sufficiency to address the needs of its citizens, which is also hoped to reduce the demand driving the unregulated commerce in organs [7]. As a consequence of transplantation's efficacy and aging populations there is an emerging trend of accepting older patients as transplant candidates meaning the number of patients waiting for new organs is likely to increase further in the coming years. Thus, finding new ways of addressing the shortage in organs is an important issue for global public health. This paper explores the potential role of a diverse set of incentives to increase the rate of deceased organ donation within the context of Swedish health care. Such a diverse set of incentives would try to capture both self-promoting motives as well as more altruistically based motives in order to increase the supply of organs.

Deceased donation is the cornerstone of organ transplantation, because donors simultaneously provide multiple organ meeting needs within diverse disease categories [3,8]. For many organs, living donation is either not possible (e.g. heart) or decidedly risky for the donor, making deceased donation preferable for bridging the gap in needs. The typical deceased donor is someone who is "brain dead" – breathing on a mechanical ventilator but with irreversible end of all brain activity, after massive trauma to the brain or stroke. Consequently, the situation is decidedly different between deceased donation and living donation as there is no risk to the deceased donor as a result of the donation because the organs are no longer of benefit to him or her. Given these advantages, increasing the availability of deceased donor organs is a goal for public health that merits serious consideration.

Sheehey et al. (2008) indicate that "the greatest opportunity for increasing the rate of donation from the pool of brain-dead potential donors lies in increasing the consent rate" [8]. While seeking ways to increase the rate of deceased organ donation

it is important that such initiatives acknowledge local particularities and ethical considerations. This was underlined in a resolution passed at the Munich Congress on the Ethics of Organ Transplantation in 2002: "Individual countries will need to study alternative, locally relevant models, considered ethical in their societies, which would increase the number of transplants, protect and respect the donor" [9,10]. This is because naturally, "Each country's health care system reflects its own history, culture, political system, and society" [11,12] making it necessary to take a contextual approach to organ donation policies by tailoring them to particular healthcare systems.

In the Swedish context any policy level proposal to increase deceased donation must be in accordance with what is known as the Swedish ethical platform for healthcare priority setting. This platform is comprised of the principles of human dignity, needs and solidarity, and cost-effectiveness. They are legislated as the guiding principles for priority-setting decisions and also represent the core values of Swedish health care [13]. The principles are lexically ordered. The human dignity principle, the highest ranked, is concerned with the equal distribution of value and rights while needs and solidarity stipulates that healthcare resources should be distributed to the areas representing greatest need. Finally, cost-effectiveness calls for a reasonable relationship between cost and effect measured as improvement in health and quality of life.

The aim of this paper is twofold; firstly to develop a proposal for compensated deceased organ donation based on a nuanced incentive scheme that could potentially bridge the gap in transplantation needs. Secondly to examine the compatibility of this proposal with the Swedish priority-setting ethical platform which represents the core values of the Swedish healthcare system and is intended to guide decision making at various levels of the healthcare system. Structurally the paper is divided into four sections; in the first part above we have introduced our contextual approach with the Swedish priority-setting ethical platform. The second section presents a compensation based organ donation model for increasing the number of deceased donor organs, grounded in incentive theory and empirical evidence regarding prosocial behaviour. In the third section we draw together the preceding sections to explore the compatibility of our proposed model with the principles of the Swedish ethical platform. Finally we devote some space to discussing the potential ethical challenges the proposal might present.

Incentivizing deceased organ donation

Using incentives to increase the supply of organs is often associated with a mental imagery of the rich buying organs from the desperately poor. In reality, however, incentives are any factor (financial or non-financial) that motivates individuals to take a particular course of action. In other words, incentives are the expectations that encourage people to behave in a certain way and could hence be tailored in order to achieve organizational success. Financial incentives such as salaries and bonuses are the most obvious types of incentives, but there are also non-financial incentives which in certain settings may be far more powerful tools when trying to influence individual behaviour. When people act against economic self-interest in order to promote wider societal interest, various types of non-financial incentives can usually explain why people choose to behave prosocially.

Incentives and prosocial behaviour

When trying to explain why individuals engage in activity that cannot be explained by economic self-interest alone (for example, doing volunteer work or becoming an organ donor) two major approaches from behavioural agency theory can be identified. The social preference approach, which focuses on how motives such as altruism, competitiveness, reciprocity and fairness influence prosocial behaviour; the other major approach is the social esteem approach with a focus on self-regarding motives such as pride and shame and their influence on prosocial behaviour. While these two approaches are overlapping to some extent the key distinction which is critical to our discussion is that the social preference approach focuses on the individual's concern for others, while the social esteem approach extends this approach to also include concerns for what others think, as important factors [14].

The incentives for individuals to behave prosocially can, by combining these two major approaches be divided into three sub-categories: extrinsic, intrinsic and signalling motivation [15]. Extrinsic motivation refers to motivation from factors external to the individual. For example, individuals may be motivated by material rewards such as money. These rewards provide satisfaction and pleasure that the behaviour or task itself may not provide. Intrinsic motivation refers to motivation from factors internal to an individual such as the pleasure one derives from the task itself or from the sense of satisfaction in completing or even working on a task. Pure-altruism, meaning a selfless concern for the welfare of others

without regard for reward or the benefits of recognition, could be one such intrinsic motivational factor for behaving prosocially. Finally, signalling motivation refers to motivation from an individual's concern with how he or she is perceived by others i.e. we are more inclined to feel good about ourselves when others hold us in high esteem. These categories of motivations are mutually interdependent, and may interact and possibly crowd-out one another, resulting in less prosocial activity in some instances [16]. Consequently, it is important for policy makers to have a sufficient understanding of how different types of incentives interact in order to increase an individual's willingness to engage in prosocial activities.

Within the economic literature on prosocial behaviour it has been widely established that social esteem plays a key role in explaining this type of behaviour. For example, Glazer and Konrad [17] illustrate that although anonymous donations are widely considered to be most admirable they typically only constitute 1% of the total number of donations given to charity. Another very telling experiment by Lacetera and Macis (2008) showed that people were more prone to donate blood when awarded bronze, silver and gold "medals" according to the frequency of their donation [18]. However, this positive effect remained only when the rankings were published in newspapers. Mellström and Johannesson (2008) tested if a small monetary compensation (50 SEK – €5) for donating blood crowds-out the supply of blood donors [19]. They found a significant decrease in blood donors when people are offered a small monetary compensation. However they also found that there was no crowding-out effect in cases where people were allowed to donate the payment to charity. This violation of basic price theory results from what economists call, a signal-extraction problem. When there is no financial reward, a prosocial act is often interpreted as genuinely altruistic behaviour. As financial rewards become more substantial, the "meaning" of the act changes: it becomes more difficult to know to what extent it is motivated by altruism or by financial motivation (since individuals differ along both dimensions). The signalling value of prosocial behavior thus weakens, offsetting or even reversing the direct effect of higher financial reward. However, when payments are allowed to be explicitly donated to charity this signalling value is restored.

Establishing appropriate incentives: our proposal for compensated deceased organ donation

The current system for deceased organ donation in Sweden relies heavily on intrinsic motivation in the

form of altruism as a selfless concern for the welfare of others without regard for financial reward, to generate the required donation rates. Given the increasing gap between need and supply it is apparent that intrinsic motivation alone is insufficient as a tool for bridging the gap. It is our belief that extrinsic and signalling incentives could help increase donation rates. We will in this part introduce our proposal for incentive-based deceased organ donation.

A possible option for compensating deceased organ donation is as a onetime payment to all potential donors at the time of registration for organ donation. However such a model is complicated because payments would be made to all registrants without guarantee of the future usability of organs, or assurance that donors will not renege or families of the deceased will not veto the donation. A model that is likely to be more effective is an on-the-spot contribution offered to the estate/family of the deceased. Such a financial contribution could take different forms. One option is to make the usage of such a financial contribution earmarked for a specific purpose; another is to allow the deceased (as detailed in their will) or family to use the money in any manner they wish. However, in order to limit the influence of coercion on the family of the deceased we believe an earmarked contribution is the preferable option. We propose that this earmarked contribution be approximately €5,000 paid by the state, towards funeral expenses.

While financial compensation in such a scenario would be offered in all instances where the donation process is initiated, we believe alternatives should also be presented so as not to alienate groups of donors who due to the decreased signalling value feel less inclined to become donors. In order to offset this potential crowding-out effect, the family should be able to forego the payment if they feel more comfortable doing so. A further option is that the entire or part of the monetary compensation can be directed as a donation to a reputable charitable organization chosen by the family, or the deceased if their will is known in honour of the deceased and the family. This can, for example, be a contribution to a cancer foundation, or an organization that tackles poverty in the developing world. For practical reasons a list of handpicked charities should be provided from which the donor or the family can choose. This will lighten the administrative load. This potential list of charities would have to be vetted by the authorities but should include a variety of charities dealing with popular charitable causes. These charities would have to be evaluated on a number of characteristics such as efficacy, history of charitable work, connections to banned groups (which can be a concern in

conflict areas). Charities which have a broad appeal, inclusive in their mandate, and work across religious, cultural and political lines should be chosen. Certainly a more sophisticated methodology would be developed when implementing such a proposal to ascertain the type of charities which will qualify. We believe this opportunity in which the act of donation takes on a dual charitable role, saving lives of patients, and contributing to a cause dear to the donor or the family will be an attractive proposal, with outstanding social utility. Providing these alternatives will insure that the greatest numbers of preferences are accommodated since as the theoretical approaches suggest, motivation to partake in prosocial behaviours can be driven by a combination of factors. Several field studies have demonstrated that a donation to charity increases if it is associated with a high degree of publicity [20,21]. According to these findings, a publicized donation, acknowledged through a plaque or certificate, to a reputable charity in honour of the deceased and the family could serve to increase the rates of consent to organ donation.

The components and details of the compensation proposal can be summarized as follows:

- €5,000 contribution to the estate or family of the deceased towards funeral expenses.
- Deceased (advanced directive) or family could decline the contribution in full.
- All or part of the €5,000 contribution could be directed as a publicized donation to a reputable charity of the deceased (if the will is known) or family's choice.
- In all circumstances the right to self determination of the deceased must be respected; and the decision of the family cannot be in contradiction to the will of the deceased if it is explicitly stated.

Compensation and the Swedish priority-setting ethical platform

Now that we have presented our proposal for compensation, we return to the aforementioned priority-setting ethical platform to explore the compatibility between our proposal and the values of the Swedish healthcare system.

Human dignity

The human dignity principle holds the foremost position, outweighing other principles, in the priority-setting ethical platform; it denotes that in health care all individuals should have equal value and equal rights regardless of personal characteristics or role in society [13].

Human dignity is a broad concept, lending itself to diverse applications both in health care and beyond. Nordenfelt (2007) distinguishes between four varieties of human dignity. These are the dignities of merit, moral stature, identity, and finally a universal type of human dignity (*Menschenwürde*) [22]. The dignity of merit is attached to an individual's social rank or position, for example that which is derived from one's occupation. The dignity of moral stature is tied to an individual's moral actions or thoughts. The dignity of identity is tied to our image of ourselves as autonomous persons with a history, future, and relationships with other persons [22]. These first three varieties exist in different degrees and are not evenly distributed among individuals in society, and they are subject to change over time depending on shifts in one's social position or actions; consequently they do not fully reflect the intention of the Swedish ethical platform. It is the final variety of human dignity as presented by Nordenfelt, *Menschenwürde* referring to a common and equal dignity based on human value and rights, which is most closely aligned with the intention of the Swedish priority-setting platform. This is a basic dignity independent of personal characteristics and is the basis for equitable health care. Under this conception of dignity all individuals should have equal opportunity in getting their health needs met, that is, patients should not be discriminated based on non-medical factors such as income, gender, race and so on.

Recipient and donor perspectives

Universal health care in welfare states such as Sweden aims to meet population health care needs irrespective of a patient's non-medically relevant personal characteristics. For instance a patient in need of bypass surgery should be prioritized independent of occupational status, religious observance, or social and biological relationships, attributes which are tied to the dignities of merit, moral stature, and identity. This does not mean that all patients in the same medical situation will receive identical treatment. Living donor transplantation presents a challenge to the aim of the human dignity principle since health needs can be met insofar as the patient is able to bring forth a donor to overcome the systemic shortage in organs. Many people might intuitively find it permissible that the healthcare system encourages patients to rely on their social capital to locate willing donors. However, this can unfortunately create great strain on the patient who must rely on others or face grave health consequences. Understandably, factors determining access to

living donor transplantation fall outside of matters that health can influence. For example, social capital in the form of family and external social relationships enhance the likelihood of accessing living donors [23]. On a financial level those with superior financial capital may circumvent the entire system by travelling to the developing world for transplantation [6]. The current transplantation system can only strictly meet the human dignity principle in the distribution of cadaveric organs, which it can exercise discretion over, through equal access and prioritization on waiting lists irrespective of one's non-medically relevant characteristics. A successful incentive-based organ donation model which results in an increase of deceased donor organs is more able to effectively meet the goal of the human dignity principle, since it will level out differences in access to organs between patients with low social capital and patients with high social capital.

A concern which surfaces in compensated donation is that donors may disproportionately come from lower socioeconomic strata, coerced by the financial incentives, which would mean that personal characteristics – in this case material wealth or the lack thereof – plays a role in the likelihood that one becomes a donor, which would be contrary to the human dignity principle. With regard to our incentive-based proposal it needs to be acknowledged that the concern about coercion does not hold the same implications for deceased donation as it does for living donation owing to the presence of risk associated with the medical procedure for living donors [24]. Furthermore, there are no obvious reasons to believe that the influence of coercion should be any weaker in systems relying on altruism to motivate deceased donation. In such a system social recognition and social esteem (as opposed to money) is more likely to be the potential coercive factors. The amount we propose is not large enough in our opinion that it risks becoming a strongly coercive factor to donors or families to donate against their best judgment.

Some might also raise the objection that the commodification of the human body is incompatible with the idea of human dignity. This would have been a fair objection if we were proposing a free market for organs where individuals are allowed to offer lucrative bids for organs. However we propose that compensation should be equal for all donors regardless of characteristics such as age, race, sex and so on. There will be no bidding for organs, which would clearly violate the human dignity principle and would increase the potential for coercion and commodification. In this proposal all individuals are given equal value, the amount is paid by the state, and all

organs are allocated by the healthcare system according to relevant medically established criteria.

Needs and solidarity

Needs. The needs component of this principle signifies that healthcare resources should be committed to the people or activities representing the greatest need [13].

End stage organ diseases without treatment invariably lead to death, making the need and urgency for care arising from these conditions universally accepted. It is for this reason that end stage organ diseases are found in the highest priority grouping amongst various conditions in Sweden's official priority ranking. It is also well established that transplantation more effectively meets patient needs by providing both better long term survival, and quality of life [25,26]. The degree to which this need can be addressed is, however, constrained by the availability of organs required for transplantation. From a needs perspective, it is self evident that a model which contributes to alleviating the current shortage in organs for transplantation would better meet patient needs.

Although our proposal incorporates the possibility of financial compensation, this does not mean that organs will either be obtained or allocated based on the ability to pay. On the contrary organs will be provided on the basis of relevant medical criteria ensuring adherence to the needs principle in the provision of care.

Solidarity. The solidarity component of the need and solidarity principle prescribes equal opportunities for health care, and argues that inequalities in the condition of living a good life should be equalized to the greatest extent possible; it also suggests that those who are more fortunate should demonstrate solidarity by allowing the prioritization of the needs of the less fortunate in the context of health [13].

Healthcare funds, generated through taxation, in welfare states are viewed as shared societal resources. These healthcare funds are utilized to finance other resources such as equipment, infrastructure, and medical personnel which are then also regarded as shared societal resources. According to the solidarity principle these shared resources would be directed at first to the neediest.

To apply this principle to organ donation, deceased donor organs, once available, may also be considered a shared societal resource and distributed to those most urgently in need. Offering payment or incentives for such posthumously donated organs does not defy the principle of solidarity for the following reasons. Firstly, incentives can increase the amount

of shared resources (a necessary element for practising solidarity). Secondly, it increases our ability to distribute these resources according to needs, thus meeting the principle of solidarity. And finally, it provides distinct advantages to relying on living donors. This is because living donors almost always donate an organ to a particular individual (e.g. to whom they have an emotional attachment) rather than donating to a pool of shared resources without a specific recipient in mind, and thus (almost without exception) exercise non-solidarity based preferences in the allocation of their organs.

Cost effectiveness

Cost effectiveness is the final principle in the priority-setting platform, and denotes that when choosing between different medical interventions there should be a reasonable relationship between costs and effects measured as improvements in health and quality of life. The principle is applied when comparing treatments within the same disease category; for example the comparative cost effectiveness of dialysis vs. transplantation in the treatment of end stage renal disease [13].

Economic analysis plays an increasingly important role in contemporary healthcare decision making [27]. With the growing range of new health technologies, decision makers use this tool to calculate the return on investment for various procedures as measured both by direct financial costs as well as gains in survival, and quality of life for patients. We will rely here on figures presented in the international literature to illustrate the reasonable relationship between costs and effects of our proposed incentives model.

Mendeloff et al. (2004) have specifically investigated the amount healthcare systems can cost effectively invest for the addition of each donor to the donor pool [28]. After a review of the costs and effects associated with various forms of organ transplantation the authors conclude that an investment of roughly \$1.3 million for each additional deceased donor would be considered cost effective. The exact value can change depending on the assigned value per quality-adjusted life year (QALY); the authors assign a value of \$100,000 per QALY, which is a widely accepted and used value per QALY. Whiting et al., (2004) have investigated the cost effectiveness of an initiative called Donor Action [29]. The authors conclude that in the Canadian context an investment of up to \$1 million can be considered cost effective so long as it generates as little as three additional donors. Matas and Schnitzler (2004) calculate that each living kidney donor saves the US medical system

approximately \$100,000 [30]. However the cost saving arising from finding new deceased donors is likely to be even bigger. If one takes into account the multiple organs which are retrieved from a single deceased donor the potential savings multiply across a number of recipients. In addition, the costs of donation are also likely to be much lower for diseased organ donations because there will be less cost associated with medical treatment of the donor. Consequently, from a pure cost perspective compensation will prove cost-saving. Furthermore recipients of the transplants will contribute to the cost effectiveness through gains in survival, and gains in the quality of life during those years of additional survival.

The literature uniformly indicates that efforts to increase organ donation are highly desirable even at a high cost per additional donor. The figures above are much higher than the €5,000 we have suggested as a contribution to the funeral expenses or as a donation on behalf of the deceased person and his or her family. The introduction of small incentives to increase deceased organ donation with its large potential for cost effectiveness makes it attractive according to the cost effectiveness principle. Contrary to most health interventions on a purely financial basis introducing compensation for deceased organ donation would be both cost saving and life saving.

Final discussion and conclusion

Although it is greatly admirable to donate organs based on a pure will to help others in need, it cannot be denied that the current system, which relies solely on altruism to motivate donation has failed to bridge the gap between demand and supply for organs. As a consequence organs that could save lives are routinely missed. And despite various efforts to increase rates of organ donation the gap persists; it is obvious that a new approach is required. This paper has discussed compensated deceased organ donation as one such approach. Based on this discussion we conclude that our proposal to introduce compensated deceased organ donation is compatible with the values of the Swedish healthcare system, and that it therefore merits serious consideration. While coming to this conclusion, it is however important to acknowledge the moral challenges that introducing compensation for deceased organ donation might bring. Three of these challenges that we wish to re-emphasize are the risk that families will be coerced by the compensation, concerns related to

commodification of the body, and the potential loss of trust in the healthcare system.

Introducing compensated deceased organ donation may be deemed impermissible if the incentives which it creates are so forceful that it makes individuals or their families donate against their true wishes. It cannot be ruled out that some people may feel coerced by the compensation. However we believe this effect associated with our proposal will be limited. We see no obvious reason why our proposal should be more coercive than the current system, where individuals might feel coerced through factors related to social recognition. Furthermore the proposal is varied to attract individuals motivated by various incentives. The concern about coercion is, however, an issue which requires careful consideration and monitoring.

Secondly, concerns regarding commodification also deserve some attention with regard to our proposal. It is important to stress that our proposal does not support a free market for organs. We believe that it should strictly remain a buyer's market in the sense that the buyer should be the state with a fixed programme for how donors should be compensated. This will avoid price mechanisms making organs from younger individuals more lucrative to donate, for example, which would indeed lead to undesirable commodification of the body.

Thirdly, potential loss of trust in the healthcare system merits recognition. Without the support and maintained trust of the public in the healthcare system such a proposal cannot be successful. As a result it is of utmost importance that any initiative to introduce compensation is accompanied by an open and transparent public discussion. It must be made clear that this proposal is part of the universal healthcare system for all citizens and will not benefit only a subset of society.

The issue of compensated donation is without any doubt a difficult one. However the issue is unlikely to go away given the growing need for organs. As illustrated, there are features related to compensation and deceased organ donation that can be viewed as ethically objectionable, where the risk of coercion is probably the most significant in our view. The fundamental question that we need to ask ourselves is: are these objections so strong that it should hinder society from introducing compensation in order to increase the pool of organs (something which undeniably will save lives)? In the case of our incentive-based proposal we do not view these objections as strong enough for allowing people to die. Consequently we suggest that the potential benefits of introducing a regulated system of compensated deceased organ donation outweigh

potential disadvantages in the context of Swedish health care.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Acknowledgements

We are grateful to Zeynep Gürtin-Broadbent, Alena Buyx and three anonymous reviewers for helpful comments and suggestions on earlier drafts of this paper. GT would also like to send a special thank you to the Brocher Foundation which supplied him with excellent research facilities and a stimulating environment while working on this paper.

References

- [1] Daniels N, Sabin JE. Setting limits fairly: can we learn to share medical resources? Oxford, New York: Oxford University Press; 2002.
- [2] Sabik LM, Lie RK. Priority setting in health care: lessons from the experiences of eight countries. *Int J Equity Health* 2008;7:4.
- [3] Ojo AO, Heinrichs D, Emond JC, McGowan JJ, Guidinger MK, Delmonico FL, et al. Organ donation and utilization in the USA. *Am J Transplant* 2004;4:27–37.
- [4] Randhawa G. Procuring organs for transplantation – a European perspective. *Eur J Public Health* 1998;8:299–304.
- [5] Budiani-Saber DA, Delmonico FL. Organ trafficking and transplant tourism: a commentary on the global realities. *Am J Transplant* 2008;8:925–9.
- [6] Bramstedt KA, Xu J. Checklist: passport, plane ticket, organ transplant. *Am J Transplant* 2007;7(7):1698–1701.
- [7] The Declaration of Istanbul on Organ Trafficking and Transplant Tourism. *Clin J Am Soc Nephrol* 2008;3:1227–31.
- [8] Sheehy E, Conrad SL, Brigham LE, Luskin R, Weber P, Eakin M, et al. Estimating the number of potential organ donors in the United States. *N Engl J Med* 2003;349(7):667–74.
- [9] Daar AS. Paid organ procurement: pragmatic and ethical viewpoints. *Transplant Proc* 2004;36:1876–7.
- [10] Warren J. Commerce in organs acceptable in some cultures, guidelines needed, ethics congress recommends. *Transplant News* 2003;13:1–3.
- [11] Hasman A, Holm S. Accountability for reasonableness: opening the black box of process. *Health Care Anal* 2005;13:261–73.
- [12] Enthoven AC. What can Europeans learn from Americans? *Health Care Financ Rev* 1989;Spec No:49–63; discussion –77.
- [13] The Swedish National Center for Priority Setting in Health Care. Health care's all too difficult choices? Survey of priority setting in Sweden and an analysis of the Riksdag's principles and guidelines on priorities in health care. SNCPSHC; March 2007.
- [14] Ellingsen T, Johannesson M. Paying respect. *J Econ Perspect* 2007;21:135–49.
- [15] Bénabou R, Tirole J. Incentives and prosocial behavior. *Am Econ Rev* 2006;96:1652–78.
- [16] Deci EL, Ryan RM. Intrinsic motivation and self-determination in human behavior. New York: Plenum; 1985.
- [17] Glazer A, Konrad K. A signaling explanation for charity. *Am Econ Rev* 1996;86:1019–28.
- [18] Lacetera N, Macis M. Motivating altruism: a field study. SSRN 2008, Discussion Paper No 3770. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1290039# (accessed 2 September 2009).
- [19] Mellström C, Johannesson M. Crowding out in blood donation: was Titmuss right? *J Eur Econ Ass* 2008;6(4):845–63.
- [20] Harbaugh TW. What do donations buy? A model of philanthropy based on prestige and warm glow. *J Public Economics* 1998;67:269–84.
- [21] Soetevent AR. Anonymity in giving in a natural context – a field experiment in 30 churches. *J Public Economics* 2005;11–12:2301–23.
- [22] Nordenfelt L. The varieties of dignity. *Health Care Analysis* 2004;12:69–81.
- [23] Giles S. An antidote to the emerging two tier organ donation policy in Canada: the Public Cadaveric Organ Donation Program. *J Med Ethics* 2005;31:188–91.
- [24] Omar F, Tufveson G, Welin S. Compensated living kidney donation: a plea for pragmatism. *Health Care Analysis online*; Jan 2009.
- [25] Wolfe RA, Ashby VB, Milford EL, Ojo AO, Ettenger RE, Agodoa LY, et al. Comparison of mortality in all patients on dialysis, patients on dialysis awaiting transplantation, and recipients of a first cadaveric transplant. *N Engl J Med* 1999;341:1725–30.
- [26] Laupacis A, Keown P, Pus N, Krueger H, Ferguson B, Wong C, Muirhead N. A study of the quality of life and cost-utility of renal transplantation. *Kidney Int* 1996;50:235–42.
- [27] Sculpher M, Claxton K. Establishing the cost-effectiveness of new pharmaceuticals under conditions of uncertainty – when is there sufficient evidence? *Value Health* 2005;8:433–46.
- [28] Mendeloff J, Ko K, Roberts MS, Byrne M, Dew MA. Procuring organ donors as a health investment: how much should we be willing to spend? *Transplantation* 2004;78:1704–10.
- [29] Whiting JF, Kiberd B, Kalo Z, Keown P, Roels L, Kjerulf M. Cost-effectiveness of organ donation: evaluating investment into donor action and other donor initiatives. *Am J Transpl* 2004;4:569–73.
- [30] Matas AJ, Schnitzler M. Payment for living donor (vendor) kidneys: a cost-effectiveness analysis. *Am J Transpl* 2004 Feb;4(2):216–21.

PAPER

IV

Horizontal Inequality in Rationing by Waiting Lists

Gustav Tinghög^{1,2}, David Andersson¹, Petter Tinghög^{3,4}, Carl Hampus Lyttkens⁵

¹Center for Medical Technology Assessment, Department of Medical and Health Sciences, Linköping University, Linköping, Sweden

²The National Center for Priority Setting in Health Care, Linköping University, Linköping, Sweden

³Department of Bioscience and Nutrition, Karolinska Institute, Stockholm, Sweden

⁴Department of Neuroradiology, Karolinska University Hospital, Stockholm, Sweden

⁵Department of Economics, School of Economics and Management, Lund University, Lund, Sweden

Corresponding Author:

Gustav Tinghög

Department of Medical and Health Sciences

Linköping University

SE-581 83 Linköping

Sweden

Email: gustav.tinghog@liu.se

Phone: +46 13 221751

Fax: +46 13 224995

ABSTRACT

Objective: To investigate the existence of horizontal inequality in access to care for six categories of elective surgery in a publicly funded system, when care is rationed through waiting lists.

Data Source: Administrative waiting-time data on all elective surgeries (N=4634) performed in Östergötland, Sweden during 2007 were linked to national registers containing variables on socioeconomic indicators.

Study design: By using multiple regression, we tested five hypotheses reflecting that more resourceful groups receive priority when rationing by waiting lists.

Principal findings: Low disposable household income predicted longer waiting times for orthopedic surgery (27%, $p<0.01$) and general surgery (34%, $p<0.05$). However, no significant differences on the basis of ethnicity and gender were detected. A particularly noteworthy finding was that disposable household income appeared to be an increasingly influential factor when the waiting times were longer.

Conclusions: Our findings reveal horizontal inequalities in access to elective surgeries, but only to a limited extent. Whether this is good or bad depends on one's moral inclination. From a policymaker's perspective it is nevertheless important to recognize that horizontal inequalities arise even though care is not rationed through ability to pay.

Keywords: rationing, waiting list, horizontal equity, elective surgery, Sweden

1. INTRODUCTION

All healthcare systems ration care in one way or another. Rationing aims at bridging the gap between demand and available resources by limiting the possibilities to fully meet the need for health care. The method for conducting rationing can take different forms. When care is rationed by price, demand is restrained by the monetary price, which the patient faces at the point of demand. According to standard economic models price is increased until demand and supply converge at a market-clearing price. In public systems where care often is free or priced well below market-clearing level, demand and supply must be reconciled through other rationing devices.

Rationing by price raise delicate normative issues related to equality in general and different socioeconomic groups' equal access to health services in particular. Within publicly funded health-care systems distributional issues related to some notion of equality are often core concerns¹ (van Doorslaer et al., 1993, Hurst, 2002). The notion of horizontal equity is typically interpreted to mean that: people in equal need are to be treated equally, irrespective of personal characteristics such as age, gender, income, and ethnicity (Wagstaff and van Doorslaer, 2000). Adherence to the principle of horizontal equity is difficult to maintain when care is rationed by price. Therefore, public funded systems commonly rely on waiting lists as a mechanism to limit availability, while still trying to maintain that persons in equal need are treated equally. However, since rationing by waiting lists is most often done implicitly, there are reasons to suspect the presence of systematic socioeconomic differences in waiting times for patients with equal need, i.e. similar medical condition. This study investigates the

¹ We will in this paper refer to equality as a descriptive notion and equity as a normative notion. Much of the existing literature has failed to recognize this important difference. For a more thorough discussion with regards to the relationship between equality and equity see Tinghög (2011)

existence of horizontal inequality in waiting times for six types of elective surgery in the Swedish healthcare system.

Waiting lists are conceptually related to waiting lines². When care is rationed through waiting lines, individuals are required to physically remain in a queue until being admitted to care. The time during which individuals are restrained from performing other activities constitutes an opportunity cost, making individuals less prone to seek care. Waiting lists, on the other hand, do not require individuals to physically remain in a queue while waiting. Instead waiting lists operate as a rationing mechanism mainly because the value of future treatments declines the longer the individual has to wait; this produces convergence between supply and demand (Lindsay and Feigenbaum, 1984, Martin and Smith, 1999). To illustrate the different rationing mechanisms at play, one can think of the time spent in line at a fast food restaurant compared to the extra waiting time associated with ordering take-out. While standing in line at the fast food restaurant prohibits persons from doing most other activities, ordering take-out allows one to participate in other activities of their choice.

It is important to acknowledge that the declining value of future treatment is not merely an economic discounting device. Being on a waiting list also involves increased burdens as a result of the disutility associated with remaining on the waiting list (Propper, 1995). Returning to the example of the fast food restaurant; while the person who orders take-out is able to participate in other activities, they are also likely to experience increased hunger as a result of the extra waiting time. This study focuses on inequalities in access related to rationing by waiting lists.

² Both are dynamic rationing tools in the sense that the number of patients waiting at a given moment in time will be dependent on the expected waiting time.

1.1 Previous research on waiting times in public healthcare systems

When investigating inequalities in access to health services it is also important to recognize that waiting times constitute only a small, albeit important, part of the wider concept of *access* to health services. Much of the empirical literature on horizontal inequality and access to health care has focused on other components of access, such as the utilization, provision, and delivery of healthcare services (van Doorslaer et al., 2000, d'Uva et al., 2009, Wagstaff and Vandoorslaer, 1992). Horizontal inequality as regards to waiting times in publicly funded systems is, by comparison, a less systematically explored area. A vast, but disparate, body of literature exists concerning the association between various categorizations of socioeconomic status and waiting time for specific patient groups or treatments (Pell et al., 2000, Yong et al., 2004, Fitzpatrick et al., 2004, Milner et al., 2004, Hacker and Stanistreet, 2004, Pettinger, 1999, Shortt and Shaw, 2003). Only a few have investigated such associations within the context of the Nordic healthcare systems (Hojmark Sorensen et al., 2009, Carlsen and Kaarboe, 2010, Lofvendahl et al., 2005, Arnesen et al., 2002). Of these, some have confirmed the existence of a link between lower socioeconomic status and increased waiting time within specific treatment domains (Hojmark Sorensen et al., 2009, Carlsen and Kaarboe, 2010), while others found no such link (Arnesen et al., 2002, Lofvendahl et al., 2005). Moreover, all previous studies have varied extensively in terms of setting, methodology, and objectives, making it difficult to draw general conclusions from previous studies about the association between waiting time and socioeconomic status.

In terms of setting and objective, the study by Siciliani and Verzulli (2009) is the one most similar to this study. Hence, their results are also the most relevant for comparison purposes. Their study uses data from the Survey of Health, Ageing, and Retirement in Europe (SHARE)

when investigating how socioeconomic status among elderly Europeans influences waiting times (at a given level of need) for specialist consultation and non-emergent surgery. According to their findings, patients with higher levels of education experienced shorter waiting times for non-emergent surgery in Sweden. Their results also showed that an increase in income by 10 000 Euros (EUR) decreased waiting time by 3% (but this association failed to reach the significance level). Seemingly contradictory, they also found that an increase in income of EUR 10 000 increased waits for specialist consultation by 11%. Siciliani and Verzullis' findings were based on a self-reported data from a limited sample population (n=247) with a high non-response rate. Hence, it will be interesting to see if their findings can be replicated within a larger and more reliable sample.

1.2 Why do we expect horizontal inequalities in waiting times?

At least three types of reason can be given for suspecting systematic socioeconomic differences in waiting times for patients with equal need. From the perspective of economic efficiency it would seem reasonable prioritize individuals that are more productive. Hence, one could argue that it is more efficient to prioritize individuals who currently hold a job. Further, if assuming that income is related to productivity, it would be more efficient to prioritize high-income individuals. One could also argue that prioritizing younger individuals is more efficient since this would most likely maximize the number of healthy life-years gained. Although rationales related to productivity and 'ageism' is explicitly discouraged at the macro level within the Swedish healthcare system, it does not seem far-fetched that it might play a role in practice when rationing care at the patient level.

The second type of reason is that less resourceful groups are less able to successfully communicate and express their needs, which is an empirically well-established fact (e.g. Brewer, 2002). The occurrence of this phenomenon may, however, be understood as a result of several non-mutually exclusive factors. One possible explanatory factor is that the working of the lists has been shaped largely by the most advantaged segments of the community, making the lists (either consciously or unconsciously) better suited to accommodate the needs of their peers. This type of explanation may also be advocated in relation to the hypotheses that women and patients of foreign origin experience longer waiting times. A second explanation, to the above describe phenomenon, may be that resourceful groups are more knowledgeable about how the system works, or can at least more effectively acquire such knowledge. This, in theory, would make them better equipped to ‘manipulate’ the system, e.g. by arguing in ways that strike a chord with personnel that can influence one’s time spent on the waiting list.

Finally, the psychological literature establishes that individuals generally tend to have an unconscious (implicit) in-group bias, which makes them more aversive toward groups dissimilar to one’s own group (Olsson et al., 2005). As an example, psychological testing of physicians has indicated that unconscious biases may contribute to racial/ethnic disparities in setting healthcare priorities (Green et al., 2007). Although this type of reason is not always easily distinguished in practice from the former reason, it puts a more direct focus on the behaviors of the actors involved in decision making at the patient level.

Given the background described above, this study aims to investigate whether horizontal inequalities exist in waiting times for six types of elective surgeries in the Swedish healthcare

system. To achieve this aim we set out to test five hypotheses that more resourceful groups receive priority when rationing is done through waiting lists:

- Male patients receive priority over women
- Younger patients receive priority over older patients
- Patients with higher incomes receive priority over patients with lower incomes
- Patients of Swedish origin receive priority over patients of foreign origin
- Patients currently active in the workforce receive priority over patients currently inactive.

2. METHODS & MATERIALS

2.1 Study setting

Patients in Sweden are assigned to waiting lists as follows: Patients with non-acute conditions first consult their general practitioner, who will refer them to a specialist if needed. In some cases the patient can visit the specialist directly for medical assessment. If patients are considered to require surgery, they will be placed on a waiting list. When investigating inequalities regarding waiting lists there are two main stages where discrimination can occur. First, general practitioner or/and referral specialists might have different standards for different socioeconomic groups which might lead to unequal opportunities to get on the waiting list. Secondly, patients might be treated unequally once they have been placed on the waiting list. This study focuses on this second stage.

In large part the Swedish system is publicly managed, with 21 county councils owning and operating most healthcare facilities. Each county council has a defined geographical

catchment area. Patients are restrained from seeking non-acute care outside their own county councils. Although county councils are formally self-governed, Sweden has a long tradition of promoting egalitarian social justice and reducing differences across county councils (Gerdtham and Sundberg, 1998, Anell, 2005). This is highlighted further in the Swedish Health and Medical Services Act (1982), which declares that the objective of the Swedish healthcare system is to promote “good health on equal terms”.

In 1992 a maximum waiting time guarantee for 12 elective procedures³ was introduced to reduce geographic variations in Sweden. According to this guarantee, patients waiting for surgical procedures should wait no longer than 90 days from referral to surgery, unless an agreement is reached between the patient and the doctor. If the county council fails to meet this requirement the patient has the right to seek surgery in another county council, or at a private clinic, at the expense of the home county council (Hanning, 1996).

Östergötland County Council, from which the administrative waiting time data used in this study was retrieved, is located in the southeastern part of Sweden and has 420 000 inhabitants. Three hospitals are located in the catchment area. Most publicly funded elective procedures are performed at these public hospitals, although one private clinic also performs some elective procedures that are publicly funded. Since we were unable to retrieve reliable waiting time data from this private clinic we have excluded elective procedures where the private clinic performed any share of the services.

The catchment area of Östergötland County Council has an age and sex distribution similar to the national distribution (Statistics Sweden, 2010a), and 19.2% of the inhabitants have higher

³ Additional elective procedures have been added since 1992.

education (the national average is 18.0%) (Statistics Sweden, 2010b). Comparing nationwide waiting times for the operation codes included in this study, Östergötland's pattern is similar to the nationwide averages. In Östergötland, 70% of elective procedures reported in 2007 were performed within the limit of the waiting time guarantee, compared to 73% for Sweden in total⁴.

2.2 Study design

This was a retrospective study in the sense that the lengths of the waiting times were calculated for all patients who received an elective surgery in 2007. This implies that patients who remained on the list or exited the list for reasons other than operation is not included in the data set. Data on waiting time for elective procedures were obtained from the administrative healthcare register in the county of Östergötland, Sweden. The elective procedures were based on the ICD-10 codes used to classify non-acute surgical procedures. In total, 24 non-acute surgical procedures were included. We aggregated these 24 elective procedures into six surgical categories; vascular, orthopedic, gynecology, urology, ophthalmology, and general surgery. In this paper, the 24 non-acute surgical procedures will be referred to as *elective procedures* while the six categories will be referred to as *specialties*. The full data set included 4634 patients living within boundaries of the county and treated between January 1, 2007 and December 31, 2007. Patients below 18 years of age and those waiting for a second and presumably related elective surgery were excluded from the data set.

⁴ Data comparing waiting times between different county councils can be retrieved from the publicly available database at <http://www.vantetider.se>. This database is updated and maintained by the Swedish Association of Local Authorities and Regions (a governmental agency) to monitor and minimize differences in waiting times between county councils.

Our dependent variable, *waiting time*, was defined as the number of days between referral and surgery. The independent variables age, gender, disposable income, ethnicity, and workforce activity were retrieved from Statistics Sweden (SCB) and merged with individual data on waiting time. The income concept used in this study was *disposable household income* per equivalent adult during 2007. Disposable household income equals the sum of net income in a household divided by the number of people in the household, which is adjusted according to an equivalent scale⁵. Income data are highly reliable since they have been linked to national income tax statistics. Subjects were categorized into three groups using income quartile limits to identify the lowest and highest income groups. In categorizing the level of disposable income for patients above 65 years of age and retired, we used separate quartile limits than for patients of working age. This was done because retired individuals' current income could be misleading as an indicator of socioeconomic status when compared to that of working individuals. The *foreign origin* variable was defined as patients whose parents were both born outside Sweden. *Workforce activity* was used as a dichotomized variable, where individuals in retirement, early retirement, unemployment, and education were defined as being non-active in the workforce.

2.3 Data analysis

We used multiple linear regression models to investigate the associations between waiting times and our predictors. Although our dependent variable, waiting time (days), was found normally distributed at $\alpha < 0.05$ using a Shapiro-Wilk test, we chose to transform the variable to the logarithmic scale. This was done to enable presentation of relative differences between

⁵ The equivalent scale used is: one adult = 1; two adults = 1.51; additional adult = 0.60; first child aged 0-19 years = 0.52; and second and following children = 0.42.

the groups with greater precision and thwart potential outliers from disturbing the analysis conducted for smaller groups of patients (Kirkwood and Sterne, 2003). A multiple regression model was run separately for the six specialties under investigation. To adjust for differences in urgency or severity between elective procedures we entered elective procedures (ICD-10 codes) as covariates into the regression models. Finally, to validate the results from the regression analyses, all two-way interactions were tested. All non-significant interactions at ($\alpha < 0.05$) were removed and the analyses were re-run. Interactions that remained statistically significant were then carefully examined and interpreted.

3. RESULTS

The study included 4636 completed waiting-time episodes, and 59% of the patients were women (Table I). Ophthalmology and orthopedics constituted the two largest surgical specialties and contributed with 39% and 21% of the stock of patients respectively. The average proportion of patients currently not active in the workforce was 71%. Ophthalmology was the surgical specialty with the highest mean age (74.7 years). Consequently, ophthalmology also had the highest proportion of patients categorized as having no workforce activity (88%). Disposable household income was highest among patients in urology and lowest among patients in gynecology.

Insert table I

Table II shows mean waiting times for the six surgical specialties by demographics. On average, 100 days were spent on the waiting lists. The means ranged from 150 days in orthopedics to 37 days in gynecology.

Insert table II

Table III presents the results of the multiple regression models. Concerning orthopedic surgery, it was found that men waited 18% ($p < 0.01$) longer than women. This pattern was, however, not found in other surgical specialties.

Table III further shows that *age* predicted waiting time in orthopedics ($p < 0.01$), ophthalmology ($p < 0.05$), and general surgery ($p < 0.05$). However, the direction of the associations differed to some extent among the surgical specialties. In ophthalmology, we found a clear trend that older patients waited longer. In orthopedics and general surgery we found a bell-shaped tendency, i.e. patients between the ages of 40 and 69 years stayed on the waiting list longer in comparison to younger and older patient groups.

Lower *disposable household income* was found to predict longer waiting times in orthopedics ($p < 0.05$) and general surgery ($p < 0.05$), but not in the surgical specialties vascular, gynecology, urology and ophthalmology. In orthopedics, the lowest income group waited on average 28% ($p < 0.01$) longer than the highest income group, while low-income patients in general surgery waited on average 34% ($p < 0.05$), longer compared to individuals in the highest income group.

Ethnicity was associated with waiting time in the case of gynecology, where waiting time for patients of foreign origin was on average 40% ($p < 0.01$) shorter than for patients of Swedish origin. The interaction analysis revealed that this difference could be mostly ascribed to differences between patients who are active in the workforce. Hence, this association is only true for patients of foreign origin who are currently active in the workforce. In all other fields

of surgery we found a contrasting but non-significant trend that patients of foreign origin waited longer than patients of Swedish origin.

When examining *workforce activity* as a potential predictor, we found that patients not active in the workforce experienced 29% ($p<0.01$) longer waiting times in ophthalmology. Otherwise no significant associations between workforce activity and waiting time were detected.

Insert table III

4. DISCUSSION

In many OECD countries, societal preferences for *solidarity* have made rationing by ability to pay an unviable policy option. Limiting access to care through waiting lists has therefore become a widely used rationing tool. Even though equity of access is a central theme in the health services literature, few studies have systematically investigated how rationing through waiting lists violates the principle of horizontal equity. This study provides a systematic analysis of socioeconomic-related inequalities when care is rationed through waiting lists in the Swedish healthcare system.

4.1 Main findings

We found some support for the assertion that horizontal inequality occurs when rationing elective surgeries through waiting lists. This picture, however, is not clear-cut, as inequality is not present in all surgical specialties. In vascular and urologic surgery we found no evidence

of horizontal inequalities. Concerning gynecology, the results contradicted the stipulated hypothesis, i.e. that more resourceful groups receive priority when rationing by waiting lists. In the specialties of orthopedics, ophthalmology, and general surgery the data supported the presence of horizontal inequalities. Particularly noteworthy in this respect is that lower household income predicted longer waits in both orthopedics and general surgery. These findings are in line with the results presented by Siciliani and Verzulli (2009).

In connection with the finding that low income predicted longer waiting times in orthopedics and general surgery, it is worth noting that these were also the surgical specialties where patients in general experienced the longest waiting times. This may suggest that income becomes an increasingly influential factor as waiting time becomes longer. Alternatively, it could suggest the existence of a potential threshold effect, where disutility associated with waiting is negligible as long as the waiting time does not exceed a certain length. This may be explained by the existence of an initial ‘adjustment period’ allowing individuals to mentally prepare and make necessary practical arrangements. But in cases where time on the waiting list exceeds this threshold, the disutility associated with continued discomfort and anxiety take over, making individuals increasingly active in trying to gain priority, and thereby favoring more resourceful groups.

Following this line of reasoning, the existence of an ‘adjustment period’ could explain why we mainly find horizontal inequalities in accordance with our hypotheses in the specialties with relatively long waiting times. Further studies need to investigate the existence of such a threshold effect and whether the disutility associated with being on a waiting list increases at a hyperbolic or constant rate over time. However, the credibility of the argumentation above could possibly be attributed to differences in levels of urgency and/or severity between the

underlying medical conditions related to the investigated surgical specialties. Still, all procedures included in the study were part of a national initiative to harmonize waiting times, indicating that levels of urgency and severity of the related medical conditions are fairly equal.

This study did not confirm our hypotheses concerning ethnicity and gender. Hence, men and patients of Swedish origin did not appear to receive priority. Our finding of no gender inequality contradicts recent findings by Carlsen and Kaarboe (2010), who found that male patients are overrepresented among non-acute Norwegian patients categorized as high-priority cases. This discrepancy may suggest a difference between the Norwegian and the Swedish healthcare systems concerning gender inequality. Another, and probably more likely, explanation for the contradictory results is that Carlsen and Kaarboe (2010) focus on patients who are given the right to treatment versus patients who are declined the right to treatment, as determined by the referring physician. Our study, on the other hand, focuses solely on differences among patients who have already been referred to a specific elective procedure. Hence, female patients may have been discriminated before even getting on the waiting list, but not once they have been placed on the waiting list. This may also explain why the stated hypothesis regarding ethnicity could not be confirmed.

4.2 Strengths and limitations

This study has several advantages compared to earlier studies focusing on equity and waiting times. First, Sweden has a long tradition of collecting high-quality data on diseases and healthcare utilization (Rosen, 2002). By using the civic person identification number, we could link waiting-time data from an administrative database to socioeconomic registers. In

contrast to earlier studies, this study does not rely on self-reported data or broad, unspecific categories of socioeconomic status. Second, the data in this study allows for comparisons between surgical specialties. Since the data included surgical indications, adjustments concerning elective procedures within surgical specialties could be made. These advantages, we believe, make this dataset unique allowing us to investigate waiting lists as a rationing tool in a more detailed way than previously done.

Regarding limitations, first it should be noted that disposable household income is not a flawless indicator of socioeconomic status. Elderly patients might not currently have very high disposable incomes, but could potentially have accumulated wealth during earlier stages in life. This makes current disposable household income potentially misleading as a socioeconomic indicator. Unfortunately, neither accumulated wealth nor educational levels are available as socioeconomic indicators for elderly individuals in Sweden. Nevertheless, there is good reason to assume that the disposable income of retired persons correlates with their disposable income during working age. Since we used separate quartile limits for disposable income of retired patients above 65 years of age we believe that we have captured socioeconomic differences among elderly individuals fairly accurately.

Second, studies of waiting lists as a rationing tool should, in principle, also examine all potential entrants to the waiting list and their responses to expected waiting times. Such data were however not available. This limitation has most likely decreased the socioeconomic differences detected in this study.

Third, the study is based on a retrospective approach in that it only includes individuals who have received an operation. This approach fails to take account of the fact that some

individuals might have exited the waiting lists prior to the execution of the awaited procedure. Reasons for exiting could be voluntary or involuntary. A voluntary reason could be that patients adapt to their medical condition and choose to give up their place on the waiting list, or the patient might choose to pay out-of-pocket for privately funded care to gain quicker access. Here, we need to point out that patients who pay out-of-pocket for a procedure covered by public health insurance are extremely rare in Sweden. Death before being admitted to care is the predominant involuntary reason for exiting a waiting list. As there are no obvious reasons to suspect an association between exiting a waiting list and the patients' socioeconomic characteristics, our results would probably not be substantially affected.

Finally, this study is based on waiting-time data from Östergötland County Council, which has a defined geographic catchment area covering less than 5% of the Swedish population. Despite several national initiatives to harmonize clinical practice and waiting times across different county councils, some regional differences within the Swedish context cannot be ruled out.

4.3 Final remarks

It is important to acknowledge that the mere association between socioeconomic factors and horizontal inequalities in waiting times cannot alone support judgments about injustice. Economists and philosophers have offered a wide array of justifications for inequality in access on grounds of efficiency and social justice. However, before approaching the delicate normative question concerning which inequalities should be deemed unfair, it is necessary to address the descriptive question concerning which horizontal inequalities actually exist.

This study documents that access is unequal when care is rationed by waiting lists, but only to a limited extent. Whether this is good or bad news depends on your worldview, i.e. your expectations on the functioning of public health care and your normative view of how it should function. From a policymaker's perspective it is important to recognize that horizontal inequalities arise even though care is not rationed through ability to pay. Consequently, horizontal equality in waiting times cannot be assumed in publicly funded healthcare systems unless they are closely monitored and openly discussed. Although our findings are context-specific to some extent, they are relevant for decision makers in other countries that, like Sweden, rely heavily on waiting lists as a mechanism to ration care.

REFERENCES

- Anell, A. (2005) Swedish healthcare under pressure. *Health Economics*, **14**, S237-S254.
- Arnesen, K. E., Erikssen, J. & Stavem, K. (2002) Gender and socioeconomic status as determinants of waiting time for inpatient surgery in a system with implicit queue management. *Health Policy*, **62**, 329-41.
- Brewer, M. B. (2002) The Psychology of Prejudice: Ingroup Love and Outgroup Hate? *Journal of Social Issues* **55**, 429-444.
- Carlsen, F. & Kaarboe, O. M. (2010) Norwegian priority guidelines: estimating the distributional implications across age, gender and SES. *Health Policy*, **95**, 264-70.
- D'uva, T. B., Jones, A. M. & Van Doorslaer, E. (2009) Measurement of horizontal inequity in health care utilization using European panel data. *Journal of Health Economics*, **28**, 280-289.
- Fitzpatrick, R., Norquist, J. M., Reeves, B. C., Morris, R. W., Murray, D. W. & Gregg, P. J. (2004) Equity and need when waiting for total hip replacement surgery. *Journal of Evaluation in Clinical Practice*, **10**, 3-9.

- Swedish Health and Medical Services Act. (1982) in Svensk Författningssamling, SFS Stockholm.
- Gerdtham, U. G. & Sundberg, G. (1998) Equity in the delivery of health care in Sweden. *Scandinavian Journal of Social Medicine*, **26**, 259-264.
- Green, A. R., Carney, D. R., Pallin, D. J., Ngo, L. H., Raymond, K. L., Iezzoni, L. I. & Banaji, M. R. (2007) Implicit Bias among Physicians and its Prediction of Thrombolysis Decisions for Black and White Patients *Journal of General Internal Medicine* **22**, 1231-1238.
- Hacker, J. & Stanistreet, D. (2004) Equity in waiting times for two surgical specialties: a case study at a hospital in the North West of England. *Journal of Public Health*, **26**, 56-60.
- Hanning, M. (1996) Maximum waiting-time guarantee--an attempt to reduce waiting lists in Sweden. *Health Policy*, **36**, 17-35.
- Hojmark Sorensen, H. T., Olsen, K. & Vedsted, P. (2009) Association Between General Practice Referral Rates and Patients' Socioeconomic Status and Access to Specialised Health Care: A Population-based Nationwide Study. *Health Policy*, **92**, 180-186.
- Hurst, J. W. (2002) Performance measurement and improvement in OECD health systems: overview of issues and challenges. IN OECD (Ed.) *Measuring up: improving health systems performance in OECD countries*. Paris, OECD.
- Kirkwood, B., Sterne, J. (2003) *Essential Medical Statistics* (Second ed). Blackwell Oxford.
- Lindsay, C. M. & Feigenbaum, B. (1984) Rationing by waiting lists. *American Economic Review*, **74**, 404-17.
- Lofvendahl, S., Eckerlund, I., Hansagi, H., Malmqvist, B., Resch, S. & Hanning, M. (2005) Waiting for orthopaedic surgery: factors associated with waiting times and patients' opinion. *International Journal for Quality in Health Care*, **17**, 133-40.

- Martin, S. & Smith, P. (1999) Rationing by waiting lists: an empirical investigation. *Journal of Public Economics*, **71**, 141-164.
- Milner, P. C., Payne, J. N., Stanfield, R. C., Lewis, P. A., Jennison, C. & Saul, C. (2004) Inequalities in accessing hip joint replacement for people in need. *European Journal of Public Health*, **14**, 58-62.
- Olsson, A., Ebert, J. P., Banaji, M. R. & Phelps, E. A. (2005) The role of social groups in the persistence of learned fear. *Science*, 309, 785-7.
- Pell, J. P., Pell, A. C., Norrie, J., Ford, I. & Cobbe, S. M. (2000) Effect of socioeconomic deprivation on waiting time for cardiac surgery: retrospective cohort study. *British Medical Journal*, **320**, 15-8.
- Pettinger, N. (1999) Waiting lists. Richly deserving. *Health Serv J*, **109**, 20-1.
- Propper, C. (1995) The Disutility of Time Spent on the United Kingdom's National Health Service Waiting Lists *The Journal of Human Resources*, **30**, 677-700.
- Rosen, M. (2002) National Health Data Registers: a Nordic heritage to public health. *Scandinavian Journal of Public Health*, **30**, 81-5.
- Shortt, S. E. & Shaw, R. A. (2003) Equity in Canadian health care: does socioeconomic status affect waiting times for elective surgery? *Canadian Medical Association Journal*, **168**, 413-6.
- Siciliani, L. & Verzuilli, R. (2009) Waiting times and socioeconomic status among elderly Europeans: evidence from SHARE. *Health Economics*, **18**, 1295-306.
- Statistics Sweden. (2010a) Folkmängden efter län, civilstånd och kön. År 1968-2009. Available from www.scb.se/Pages/ProductTables____25795.aspx, accessed 12 May 2010.

- Statistics Sweden. (2010b) Befolkning 16-74 år efter Kommun, Utbildningsnivå och Kön. År 1985-2009. Available from www.ssd.scb.se/databaser/makro/Produkt.asp?produktid=UF0506, accessed 14 May 2010.
- Tinghög, G. (2011) The Art of Saying No – The Economics and Ethics of Healthcare Rationing. Linköping, Linköping University Medical Dissertation.
- Wagstaff, A. & Van Doorslaer, E. (2000) Equity in health care finance and delivery. IN Culyer, A. J. & Newhouse, J. P. (Eds.) *Handbook of Health Economics*. Amsterdam, Elsevier.
- Wagstaff, A. & Vandoorslaer, E. (1992) Equity in the Finance of Health-Care - Some International Comparisons. *Journal of Health Economics*, **11**, 361-387.
- Van Doorslaer, E., Wagstaff, A. & Rutten, F. (Eds.) (1993) *Equity in the Finance and Delivery of Health Care: an International Perspective*, Oxford, Oxford University Press.
- Van Doorslaer, E., Wagstaff, A., Van Der Burg, H., Christiansen, T., De Graeve, D., Duchesne, I., Gerdtham, U. G., Gerfin, M., Geurts, J., Gross, L., Hakkinen, U., John, J., Klavus, J., Leu, R. E., Nolan, B., O'donnell, O., Propper, C., Puffer, F., Schellhorn, M., Sundberg, G. & Winkelhake, O. (2000) Equity in the Delivery of Health Care in Europe and the US. *Journal of Health Economics*, **19**, 553-583.
- Yong, P. F., Milner, P. C., Payne, J. N., Lewis, P. A. & Jennison, C. (2004) Inequalities in access to knee joint replacements for people in need. *Annals of the Rheumatic Diseases*, **63**, 1483-9.

Table I Sample characteristics

	n(%)	Mean Age	Mean disposable Income(€*)	Proportion women	Proportion of foreign origin	Proportion out of work force
Vascular ¹	417 (9.0)	66.9	19172	0.30	0.13	0.71
Orthopedic ²	990 (21.4)	60.5	18 762	0.54	0.09	0.68
Gynecology ³	357 (7.7)	58.2	17 171	1.00	0.10	0.56
Urology ⁴	204 (4.4)	69.9	23 226	0.00	0.07	0.77
Ophthalmology ⁵	1792 (38.7)	74.7	17 322	0.62	0.10	0.88
General Surgery ⁶	874 (18.9)	53.8	17 668	0.66	0.16	0.46
TOTAL	4634	65.5	18 112	0.59	0.11	0.71

* 1 Euro was calculated as 9.25 SEK according to the annual average exchange rate for 2007.

1) Cardiac valve replacement surgery (211) Coronary artery bypass grafting (206). 2) Carpal tunnel release surgery (342) Hip replacement (269) Knee replacement (230) Synovitis (177) Slipped disc surgery (57) Needle aponeurotomy for dupuytren's contracture (80) Shoulder instability surgery (27) Anterior cruciate ligament surgery (69) Subacromial (27) Spondylolisthesis (8) Rheumatoid arthritis surgery (21) Thumb arthritis (25). 3) Uterine prolapse (195). Operation for incontinence (87) Hysterectomy on benign indications (75). 4). Benign prostatic hyperplasia (188) Scrotal hernia surgery (16). 5) Cataract surgery (1760) Strabismus surgery (18) Operation of lacrimal canal (14) .6) Anal surgery (182) Cholecystectomy (165) Thyroid surgery (60) Varicose veins (77) Scar hernia (48)

Table II Mean waiting times (in days)[†]

		Mean waiting time (days)						
		n	% (n= 4634)	Vascular	Orthopedic	Gynecology	Urology	Ophthalmology
Gender	Women	2717	58.6	78.7	140.7	37.2		77.8
	Men	1917	41.4	61.8	161.3		103.9	78.4
Age group	18-39	351	7.6	87.2	121.4	42.5	76.3	75.2
	40-69	2108	45.5	64.8	152.1	36.5	91.3	71.4
	70+	2175	46.9	68.4	158.9	38.9	115.2	80.3
Income group	Lowest	1164	25.1	75.7	150.6	33.2	122.5	79.4
	Middle	2315	50.0	63.2	153.9	38.4	102.9	78.5
	Highest	1155	24.9	67.7	143.2	40.5	100.6	75.5
Ethnicity	Foreign	508	11.0	75.1	152.3	25.7	112.5	81.1
	Swedish	4126	89.0	65.7	149.9	38.6	103.2	77.7
Workforce activity	No	3312	71.5	69.3	156.5	36.8	109.9	79.7
	Yes	1322	28.5	60.9	136.9	37.8	83.6	65.4
TOTAL		4634	100.0	66.9	150.1	37.2	103.9	78.0

Table III Multiple regression model with relative values of waiting times[†]

	Vascular	Orthopedics	Gynecology	Urology	Ophthalmology	General Surgery
Gender		**	NA	NA		
Women	1.12 (0.84 - 1.48)	0.83 (0.73 - 0.94)	NA	NA	0.99 (0.92 - 1.05)	1.00 (0.85 - 1.18)
Men	1.00 (REF)	1.00 (REF)	NA	NA	1.00 (REF)	1.00 (REF)
Age group		*			*	*
18-39	0.49 (0.20 - 1.17)	0.95 (0.73 - 1.24)	1.08 (0.68 - 1.72)	0.81 (0.34 - 1.90)	0.75 (0.54 - 1.05)	0.91 (0.69 - 1.21)
40-69	1.09 (0.81 - 1.48)	1.18 (1.02 - 1.36)	0.98 (0.78 - 1.23)	0.83 (0.66 - 1.04)	0.89 (0.82 - 0.97)	1.19 (0.94 - 1.51)
70+	1.00 (REF)	1.00 (REF)	1.00 (REF)	1.00 (REF)	1.00 (REF)	1.00 (REF)
Income group		*				*
Lowest	1.06 (0.73 - 1.54)	1.27 (1.07 - 1.49)	0.86 (0.67 - 1.10)	1.25 (0.89 - 1.77)	1.02 (0.93 - 1.12)	1.34 (1.06 - 1.68)
Middle	1.07 (0.81 - 1.43)	1.19 (1.04 - 1.37)	0.92 (0.74 - 1.16)	0.91 (0.74 - 1.11)	0.98 (0.90 - 1.06)	1.10 (0.90 - 1.33)
Highest	1.00 (REF)	1.00 (REF)	1.00 (REF)	1.00 (REF)	1.00 (REF)	1.00 (REF)
Ethnicity [‡]			**	**		
Foreign	1.24 (0.85 - 1.8)	1.10 (0.90 - 1.34)	0.59 (0.45 - 0.77)	1.18 (0.81 - 1.72)	1.07 (0.96 - 1.18)	1.05 (0.86 - 1.29)
Swedish	1.00 (REF)	1.00 (REF)	1.00 (REF)	1.00 (REF)	1.00 (REF)	1.00 (REF)
Workforce activity					**	
No	1.28 (0.92 - 1.79)	0.96 (0.84 - 1.11)	1.09 (0.90 - 1.33)	1.22 (0.93 - 1.61)	1.24 (1.11 - 1.40)	0.98 (0.83 - 1.15)
Yes	1.00 (REF)	1.00 (REF)	1.00 (REF)	1.00 (REF)	1.00 (REF)	1.00 (REF)

[†]Adjusted for elective procedures

*= sig. level 0.05

**= sig. level 0.01