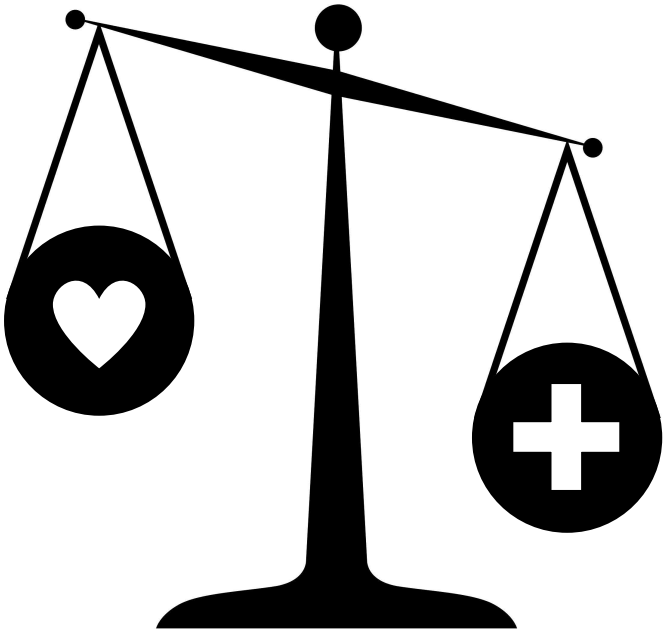
Linköping University Medical Dissertation No. 1795

Opportunity cost in healthcare priority setting

Jonathan Siverskog



LINKÖPING UNIVERSITY

Linköping University Medical Dissertations No. 1795

Opportunity cost in healthcare priority setting

Jonathan Siverskog

Department of Health, Medicine and Caring Sciences
Linköping University, Sweden
Linköping 2022

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.
https://creativecommons.org/licenses/by-nc/4.0/

©Jonathan Siverskog, 2022

Cover design: Martin Pettersson

Published articles have been reprinted with the permission of the copyright holders.

Printed in Sweden by LiU-Tryck, Linköping, Sweden, 2022

ISBN 978-91-7929-135-8
ISBN 978-91-7929-136-5 (PDF)
ISSN 0345-0082

Till Malina, Bob och Harry

"Still, in the end he worked out a method which would at least produce a result. He decided not to mind the fact that with the extraordinary jumble of rules of thumb, wild approximations, and arcane guesswork he was using he would be lucky to hit the right galaxy; he just went ahead and got a result. He would call it the right result. Who would know? As it happened, through the myriad and unfathomable chances of fate, he got it exactly right, though he of course would never know that."

- Douglas Adams

CONTENTS

ABSTRACT vii
SVENSK SAMMANFATTNING ix
LIST OF PAPERS xi
I. INTRODUCTION 1
Thesis aim 3
Outline and papers 3
A small glossary 4
II. OPPORTUNITY COST 5
Best alternative use 5
QALYs and WTP 8
The bookshelf and the meaning of "displacement" 11
Implications of inefficiency 18
Expected alternative use 26
Summary and implications 27
III. EMPIRICAL ESTIMATES 31
Methods for estimating the marginal cost of producing health 31
Evidence from Sweden 35
Credibility and experimental evidence 40
More evidence from Sweden 43
IV. ROLE IN PRIORITY SETTING 51
Equity-efficiency trade-offs 51
A societal perspective and non-health benefits 56
Sensitivity analysis 57
Do high prices imply high priority? 65
V. CONCLUSIONS 67
ACKNOWLEDGEMENTS 69
REFERENCES 71

ABSTRACT

The resources available for the public provision of health care are not unlimited. Cost-effectiveness evidence on new healthcare interventions can help us prioritise in order to use scarce resources wisely, but to interpret cost-effectiveness evidence, it may appear as if we must make trade-offs between life and money. This is not so. If we are able to quantify the health improvements that resources would or could have generated in alternative use, a decision about providing or denying treatment can instead be framed as a trade-off between health gained and health forgone. In this thesis, I seek to provide a more robust basis for this way of reporting and interpreting cost-effectiveness evidence.

In Chapter II, I discuss the definition of opportunity cost in economic evaluation. The opportunity cost of providing an intervention is what we must give up to provide it. More precisely, it is typically defined as the value of the best alternative forgone. In economic evaluation of health care, opportunity cost has been understood in terms of the least cost-effective, currently funded intervention, which should be displaced when funding new interventions subject to a fixed budget. I show that alternative uses forgone may be neither currently funded nor well-defined, which implies that we should not look to cost-effectiveness evidence on specific interventions for information on opportunity cost. Further, identifying a best alternative use assumes that priority setting is based on objectives that can be summarised into a single measure of value. If economic evaluation is used to inform trade-offs between one measure of value (e.g., quality-adjusted life years, QALYs) and other, unquantified objectives, I suggest that it would be more appropriate to define opportunity cost as value in expected alternative use.

To quantify opportunity cost as health forgone, we need evidence on the health that resources would or could have generated in alternative use. In Chapter III, I use panel data on health spending and life expectancy in Swedish regions to estimate the marginal cost of producing a QALY. My findings imply that Swedish health care can produce health at a marginal cost of SEK 180,000 per QALY, which could be used as an expectation on how productive health spending would be in alternative use. I discuss methodological issues with this approach and identify some credibility problems with selection-on-observables strategies plaguing this and similar research to date. I address (some of) these problems by assessing coefficient stability and the causal mechanisms between healthcare resource use and health outcomes, using a second panel on hospital bed capacity and

mortality. This analysis finds that health could be gained at a cost of SEK 420,000 per QALY by providing more hospital beds.

To illustrate the role of this evidence in healthcare priority setting, Chapter IV considers how it could have been used to inform decision making in a case of pharmaceutical reimbursement. I propose that economic evaluation report cost-effectiveness evidence as QALYs forgone per QALY gained. This frames a decision about providing or denying treatment as a judgement on the relative priority of QALYs gained and QALYs forgone, which is more transparent about a trade-off between equity and efficiency than deciding whether the monetary cost per QALY is too high. Framing decisions as health gained versus health forgone could also lead to better decision making by making opportunity costs more salient to decision makers and the reason for sometimes denying costly treatments easier to communicate.

In summary, cost-effectiveness evidence can be used to achieve the theoretical objective of health maximisation, but economic evaluations rarely report opportunity costs explicitly as health forgone. This thesis provides the practical means to be explicit and implications for the definition of opportunity cost and the interpretation of cost-effectiveness evidence when health maximisation is not the sole objective of healthcare priority setting.

SVENSK SAMMANFATTNING

Hälsoekonomiska utvärderingar av nya behandlingar är tänkta att fungera som beslutsstöd och bidra till att hälso- och sjukvårdens begränsade resurser används på ett klokt sätt. En typisk utvärdering visar hur mycket det kostar att vinna ett år i full hälsa (ett kvalitetsjusterat levnadsår, QALY) med en ny behandling. Det är dock tveksamt om den här informationen är särskilt användbar för vårdens beslutsfattare eftersom den tycks kräva att de kan och vill värdera människors hälsa i pengar. Men detta är inte det enda sättet att tolka en hälsoekonomisk utvärdering. Om resurserna som krävs för att tillhandahålla en ny behandling i stället hade kunnat användas till annan vård, kan vi förstå ett beslut om huruvida en ny behandling bör tillhandahållas som en avvägning mellan hälsan vi vinner och hälsan vi går miste om. Syftet med denna avhandling är att bidra till den här sortens tolkning av hälsoekonomiska utvärderingar, såväl teoretiskt som empiriskt.

Skillnaden mellan ett perspektiv som ställer pengar mot hälsa och ett som ställer hälsa mot hälsa speglar olika syn på vad alternativkostnaden är. Alternativkostnaden för att finansiera en ny behandling är, enligt gängse definition, värdet av det bästa alternativet vi går miste om. För att begripliggöra detta begrepp inom hälsoekonomisk utvärdering tänker vi oss ofta att hälso- och sjukvården består av en mängd olika behandlingar som redan finansierats. Om man beslutar om att tillhandahålla en ny behandling utan ett motsvarande tillskott till budget, vore det bästa sättet att finansiera den nya behandlingen att sluta tillhandhålla den behandling som i nuläget gav den minsta hälsovinsten per spenderad krona; alternativkostnaden skulle vara hälsan vi då gick miste om. I Kapitel II diskuterar jag både definitionen av alternativkostnad och denna tankefigur. Eftersom prioriteringar inte enbart bygger på en princip om ”hälsomaximering” föreslår jag att det vore rimligare att definiera alternativkostnaden som värdet av förväntad alternativ användning. Med denna definition är det möjligt att mäta alternativkostnaden som exempelvis ett förväntat antal förlorade QALYs, utan att för den skull anta att hälso- och sjukvårdens enda mål är att maximera QALYs. Vidare konstaterar jag att det är svårt att i verkligheten beskriva hälso- och sjukvården som ett antal av varandra oberoende behandlingar. Även om detta kan vara ett pedagogiskt sätt att illustrera varför nya behandlingar skulle kunna tränga undan annan vård bör vi – åtminstone när vi försöker uppskatta alternativkostnaden – snarare tänka på undanträngning som små nedskärningar över flera delar av hälso- och sjukvården, till exempel i form av personal.

Kapitel III bidrar med evidens som möjliggör kvantifiering av hälsan vi går miste om när nya behandlingar finansieras. Jag använder data på medellivslängd och hälso- och sjukvårdsutgifter i svenska regioner över tid för att skatta sambandet mellan hur mycket en region spenderar och dess hälsoutfall. Enligt dessa resultat kan svensk hälso- och sjukvård vinna ett QALY genom att spendera 180 000 kronor, vilket skulle kunna fungera som en förväntning på hur många QALYs vårdens resurser hade producerat i alternativ användning. Denna analys tar hänsyn till svårigheterna att dra slutsatser om kausalitet med hjälp av icke-experimentella data, men jag konstaterar att det finns vissa trovärdighetsproblem. I en andra analys försöker jag underbygga trovärdigheten i sambandet mellan hälso- och sjukvårdens resursanvändning och befolkningens hälsoutfall med hjälp av liknande data på mortalitet och antalet vårdplatser. Resultaten från denna studie tyder på att hälso- och sjukvården skulle kunna producera ett QALY för 420 000 kronor genom att tillhandahålla fler vårdplatser.

I Kapitel IV illustrerar jag till sist den praktiska tillämpningen av denna typ av evidens. Med ett fall av läkemedelssubvention som exempel visar jag hur den skulle kunna göra hälsoekonomiska utvärderingar till ett bättre stöd för prioriteringar inom hälso- och sjukvården. Mitt förslag är att utvärderingar explicit rapporterar antalet förlorade QALYs per vunnen QALY. Eftersom svensk hälso- och sjukvård ska fördela resurser efter behov ska inte fler förlorade än vunna QALYs automatiskt tolkas som att en behandling inte bör tillhandahållas. En av fördelarna med att presentera ekonomiska utvärderingar på det här sättet är att vi kan förstå beslutet som en bedömning av hur prioriterade patienterna vars hälsa skulle förbättras är i förhållande till de patienter som skulle gå miste om vård. Detta tydliggör avvägningen mellan effektivitet och rättvisa som hälso- och sjukvårdens beslutsfattare står inför jämfört med ett ställningstagande om huruvida den monetära kostnaden per QALY är för hög. Att rapportera kostnader som förlorad hälsa skulle också kunna leda till bättre beslutsfattande genom att framhäva alternativkostnaden, som annars är lätt att glömma eller bortse från, samt göra det lättare att kommunicera varför man ibland väljer att inte tillhandahålla effektiva men dyra behandlingar.

Hälsoekonomiska utvärderingar kan alltså hjälpa hälso- och sjukvårdens beslutsfattare att avgöra om vi vinner mer hälsa än vi går miste om när nya behandlingar finansieras, men utvärderingar rapporterar sällan explicit hur mycket hälsa som går förlorad. Den här avhandlingen bidrar med evidens som gör det möjligt att vara explicit i detta avseende, samt visar på konsekvenser för definitionen av alternativkostnad och tolkningen av hälsoekonomiska utvärderingar när prioriteringar inom hälso- och sjukvården inte handlar om att maximera QALYs.

LIST OF PAPERS

I. Siverskog, J., Henriksson, M. (2019). Estimating the marginal cost of a life year in Sweden's public healthcare sector. *European Journal of Health Economics,* 20(5), 751–62. https://doi.org/10.1007/s10198-019-01039-0

II. Siverskog, J., Henriksson, M. (2020). Mutually Exclusive Interventions in the Cost-Effectiveness Bookshelf. *Medical Decision Making*, 40(3), 399-403. https://doi.org/10.1177/0272989X20912261

III. Siverskog, J., Henriksson, M. (2021). On the role of cost-effectiveness thresholds in healthcare priority setting. *International Journal of Technology Assessment in Health Care*, 37(1), E23. https://doi.org/10.1017/S0266462321000015

IV. Siverskog, J., Henriksson, M. (2021). The health cost of reducing hospital bed capacity. Manuscript.

I. INTRODUCTION

It is a popular fact that you cannot always get what you want (Jagger and Richards, 1969). A contributing factor to this frustration is scarcity of resources. There are a lot of things that we would like to do or have, had we but the time and money. Because we do not, we must prioritise to make sure that our limited means tend to satisfy pressing needs before not so pressing desires. This is true of all human affairs, both private and public. Making wise choices about how to spend our time, money, and efforts can be hard enough as private individuals, but the difficulty pales in comparison to that of making wise choices about public spending. This is because we are fairly good judges of our own wants and needs and can be expected to make choices that align with them when left to our own devices. The difficulty in making wise choices lies, rather, in having wise preferences and making sense of the information available to us. Public sector decision makers make choices on behalf of other people. To them, the priority of different ends is not internal and the people whose interests they are elected or employed to represent may be competing. Further, it is probably fair to say that the potential consequences of their choices are often more complex (e.g., the decision to buy a new car versus the decision to build a new motorway). In this context, it is easy to understand the need to evaluate the costs and benefits of public projects to promote rational, non-wasteful public spending.

In this thesis, my focus is specifically the evaluation of new medical treatments. Health economists engage in evaluation of such interventions to produce c*ost-effectiveness evidence*, which is most often reported as a monetary cost of gaining some measure of health (typically a QALY, quality-adjusted life year). Cost-effectiveness evidence can help healthcare decision makers determine the order of priority in a list of potential investments, but when presented with the decision to provide or deny a single new treatment they are forced to face another popular fact: the value of life is immeasurable. It would seem, that unless it is possible to put a price tag on life, there is no way of telling whether the benefits of a potentially life-saving treatment justify its costs and cost-effectiveness evidence would not be of much use. Many economists would argue that this type of decision forces us to put a price tag on life (at least implicitly), whether we like it or not. That is, if the decision is to deny treatment, then apparently, the lives it could have saved were not valuable enough.

Here, I will consider another way of interpreting cost-effectiveness evidence, which I believe is more helpful to those with the difficult task of

priority setting in health care. To sum this interpretation up, I borrow words (my own translation to English) from the late, great Hans Rosling on his time as a physician in Mozambique during the 1980s, trying to explain to a visiting colleague why he did not provide the best possible treatment to a very sick child

> I explained that more deaths among children probably could be averted if I devoted more time to improving neighbourhood health care services, care centres, and small clinics [...] If we put our efforts and our staff into providing the best care in hospital, fewer children would be vaccinated, the care centres would be understaffed, and in total more children would die. I was responsible for the children I saw die, as well as those I did not see, but who nevertheless died (Rosling and Hägersten, 2017).

A decision to deny a potentially life-saving treatment does not have to be a judgement about the monetary value of life. As Hans Rosling argued, it could reflect that the resources required to provide the treatment had some other use that could save more lives. Although the consequences of resource scarcity are less palpable in contemporary high-income countries than in 1980s Mozambique, it is a difference of degree and not of kind. There are a lot of competing ends in health care, and when we use resources to help one patient, we forgo their use to help another.

The difference between these two ways of interpreting cost-effectiveness evidence reflects a difference in view on what the relevant *opportunity cost* is. Shortly, I will discuss the meaning of this word at some length, but in essence, the opportunity cost of providing a medical treatment is what we must give up to provide it. To make sense of cost-effectiveness evidence, what we gain and what we forgo must be commensurable. If what we gain is appropriately measured in health, we should try to find out how much health we would or could have gained by using resources in another way. In economic evaluation, health gains are typically measured in QALYs, which suggests that costs should be measured as QALYs forgone from this perspective. However, healthcare priority setting is not based solely on concerns about efficiency (i.e., maximising QALYs). Most notably, equity (e.g., allocating resources according to need) is an important objective for decision makers (Guindo et al., 2012) and even an explicit principle for healthcare priority setting in Swedish legislation (SFS 2017:30). Thus, there is much to be said for trying to incorporate equity concerns in economic evaluation (Williams and Cookson, 2006, Cookson et al., 2009, Cookson et al., 2017). In this thesis, I consider an approach to economic evaluation in which we measure outcomes in QALYs only but treat this as a measure of efficiency, recognising that the decisions we are trying to inform are not solely about maximising QALYs but making trade-offs between efficiency and equity.

I should note that there is nothing new *per se* about regarding the opportunity cost of funding new healthcare interventions as health forgone. What is relatively novel with what I discuss here is moving beyond monetary costs per QALY gained. Arguably, economic evaluation has been unable to do so due to a lack of empirical evidence to support quantification of health forgone. Recently, the feasibility of explicitly considering opportunity costs has improved with a surge in empirical research attempting to estimate the marginal cost at which health systems could produce a QALY (e.g., Claxton et al., 2015a, Vallejo-Torres et al., 2018, Edney et al., 2018). Here, I do my part in contributing to this research for a Swedish context. Additionally, I examine the conceptual basis for treating prioritisation decisions as health versus health trade-offs. Opportunity cost as health forgone has typically been motivated by assumptions about a fixed budget and a healthcare sector perspective (Claxton et al., 2010, Brouwer et al., 2019). I take a different starting point, to see where it leads me.

Thesis aim

The overarching aim of this thesis is to provide a more robust conceptual and empirical foundation for contemplating trade-offs in healthcare priority setting in terms of health gained and health forgone. This aim has three parts. First, it is to improve our understanding of the meaning of opportunity cost in economic evaluation. Second, it is to provide empirical evidence on the opportunity cost of funding new medical treatments in Swedish health care. Finally, it is to demonstrate how this evidence can be used to interpret cost-effectiveness evidence and guide healthcare priority setting.

Outline and papers

Chronologically, I first set out to provide an empirical estimate of the marginal cost of producing a life year and QALY in Sweden (Paper I), which could function as a measure of the opportunity cost of funding new medical treatments. I then wrote an extensive discussion on the definition of opportunity cost in economic evaluation, which turned into a short paper on the interpretation of the "the cost-effectiveness bookshelf" (Paper II). In Paper III, I illustrated how marginal cost per QALY estimates could be used to inform decision making in Sweden, particularly in the context of making trade-offs between efficiency and equity. Finally, I attempted to estimate the opportunity cost of reducing hospital bed capacity (Paper IV) to provide more credible evidence on the link between healthcare resource use and health outcomes.

The remainder of this thesis is organised with one chapter devoted to each of its three aims. Chapter II begins by discussing the definition of opportunity cost, containing much of what was left out of the published version of Paper II. In Chapter III, I discuss the methods and results of empirical work on the link between healthcare resource use and health outcomes, including my own (Papers I and IV). Chapter IV deals with the role of this empirical evidence in healthcare priority setting, including, but not restricted to the arguments explored in Paper III. Finally, Chapter V summarises my conclusions.

A small glossary

In the interest of clarity, it could be worthwhile to spell out what I mean by two words that I use continuously throughout this thesis. First, I use *healthcare priority setting* to describe decision making – or the principles underlying it – about providing or not providing health care at a group level. A typical example that I use is a decision about whether or not a new drug should be provided. Second, my use of *intervention* is most often interchangeable with treatment, health technology, or programme of care. In cases where I discuss provision of care in a sense that is more vague, e.g., increases in staffing levels compared to providing a drug, I use *healthcare services*.

II. OPPORTUNITY COST

Opportunity cost is a consequence of resource scarcity. The use of a resource for one purpose precludes its use for other purposes. It is commonly asserted that opportunity cost is a fundamental concept in economics (Ferraro and Taylor, 2005). However, closer inspection reveals its exact definition to be arbitrary and open to various interpretations (Potter and Sanders, 2012, O'Donnell, 2009). I would like to maintain that opportunity cost is a fundamental, or at least a very useful concept for economic evaluation. I therefore begin by discussing its definition.

Best alternative use

Economics text books tend to define the opportunity cost of an alternative as the value or benefit of the next best alternative (Ferraro and Taylor, 2005). Let me exemplify:

> You have won free tickets to see an Eric Clapton concert and a Bob Dylan concert, neither of which have any resale value. Sadly, both are performing on the same night. On any given day, you would be willing to pay up to $50 to see Dylan. Assume there are no other costs of seeing either performer. Based on this information, what is the opportunity cost of seeing Eric Clapton?

If you decide to see Clapton you forgo the opportunity to see Dylan, which you value at $50. Consequently, your opportunity cost of seeing Clapton is $50. Here, there is not much ambiguity. Now, consider the original question (Ferraro and Taylor, 2005) from which the previous one was adapted, in which respondents were asked to circle the correct answer (A-D)

> You won a free ticket to see an Eric Clapton concert (which has no resale value). Bob Dylan is performing on the same night and is your next-best alternative activity. Tickets to see Dylan cost $40. On any given day, you would be willing to pay up to $50 to see Dylan. Assume there are no other costs of seeing either performer. Based on this information, what is the opportunity cost of seeing Eric Clapton?
>
> A. $0 B. $10 C. $40 D$50

According to Ferraro and Taylor (2005) the correct answer is B ($10), because you forgo the value of seeing Dylan ($50) but would have to incur the cost of the ticket ($40) to choose this alternative. Surprisingly, this was the least popular answer in a survey posed to 199 highly trained economists in

the United States. We could take this result to mean that many economists have a poor understanding of the meaning of opportunity cost, but another interpretation is that opportunity cost simply is not a well-defined concept (Potter and Sanders, 2012). While there are several potential ambiguities in this question[1], the most important one for the present discussion is that we could view not having to pay for the Dylan ticket as a benefit of seeing Clapton, which would mean that the opportunity cost of seeing Clapton is $50 regardless of the price of a Dylan ticket.

We can understand this ambiguity by differentiating between the opportunity cost of using a resource and the opportunity cost of choosing an option. In my first, modified question there is a single resource (your time) which has an alternative use. Therefore, there is little room for confusion by conflating the option of seeing Clapton and the use of your time to do so. In the original question, there are two resources in play (your time and $40 cash). Seeing Clapton means forgoing the use of your time to see Dylan, but it also means gaining some use of $40. The opportunity cost of using your time to see Clapton is unambiguously $50, but it is not obvious whether we should equate this with the opportunity cost of seeing Clapton or if that should be defined as the net value of seeing Dylan[2].

This shows in what way opportunity cost is not a fundamental concept to economics. If we know the value of all (in this case both) options available to us, we need not define opportunity cost to maximise value (O'Donnell, 2009). Assume that your willingness to pay (WTP) for Clapton is $60. It is essentially arbitrary whether we define your benefit of seeing Clapton as $60+$40 and your opportunity cost as $50, or your benefit as $60 and your opportunity cost as $50-$40.

Even if opportunity cost is arbitrary and non-fundamental to economics in this sense, it could still be quite useful to economic evaluation. To make this case, however, there are several aspects of the common text-book definition that need unpacking. First, the assumption that the alternative under evaluation (e.g., the Clapton concert) is the best one and the implication that the opportunity cost is determined by the next-best alternative is peculiar from the point of view of economic evaluation. Whether the evaluated alternative is the best is, presumably, what we would like to shed

[1] Answers A and C could be defensible because we might reason that your willingness to pay for Dylan cannot exceed $40, otherwise you would have gone to the Dylan concert (Potter and Sanders, 2012). However, this has to do with the meaning of "willingness-to-pay" and the logical consistency of the information in the question rather than the definition of opportunity cost.

[2] The opportunity cost of seeing Dylan is, unambiguously, the value of the best alternative use of your time (the value of seeing Clapton) and the value of the best alternative use of $40 cash. Because the alternative use of $40 is a benefit forgone by seeing Dylan, there is symmetry argument to be made in favour of treating $40 as a benefit of seeing Clapton; i.e. if it is a cost of seeing Dylan, it is a benefit of seeing Clapton.

some light on. Boardman et al. (2014) provide a definition that seems more suitable for economic evaluation:

> The opportunity cost of using an input to implement a policy is its value in its best alternative use.

This is better than the standard microeconomics text-book definition for the two reasons discussed above. That is, it allows for the evaluated use of a resource (input) to be inferior to (less valued than) the alternative use of the resource and it assigns an opportunity cost to the use of a resource rather than to the choice of an option. This does not mean that we cannot, or that I shall not, sometimes refer to the opportunity cost of choosing an option. That would, naturally, be the sum of opportunity costs of resources used, but with the arbitrary element of whether we add the value of alternative uses gained (e.g., some use of $40) by choosing the option to the benefits side or subtract them from the opportunity-cost side. It simply bears mentioning that the meaning of the opportunity cost of using a resource is more exact in this respect.

We must also be clear about opportunity cost to whom (Culyer, 2018). In economic evaluation, we are not going to be interested in households or individuals choosing between different consumption bundles (or concerts). Consider a decision maker managing a clinic and contemplating whether to hire a new nurse. What is the opportunity cost of this? To the clinic, it is the value of the best alternative use of the wage it must pay, as well as that of resources devoted to hiring and potential training. To the nurse, it is the value of the forgone opportunity to work somewhere else. To health care or society at large it is the value of the best alternative use of the nurse's labour. In most cases of economic evaluation, the relevant opportunity cost will be that to society in some sense.[3] We wish to know whether the benefits of a medical intervention to patients, their families, and fellow citizens outweigh the benefits that are forgone, by the same people, by devoting resources to the provision of this intervention.

A final aspect of the definition of opportunity cost that needs addressing is that of value. In the choice between the Clapton and the Dylan concert, it is assumed that value is measured in terms of WTP. This might, arguably, be reasonable to keep as a tacit assumption from the perspective of the individual or household, but not in the context of decisions about public spending on health. Clearly, we cannot determine the best alternative use of a resource without first specifying what makes one use better than another (Culyer, 2018, Booth, 2019).

[3] However, I should add that the relevance of societal opportunity cost ought not to be confused with the tenets of, what is often called, a "societal perspective" in economic evaluation (Jönsson, 2009).

QALYs and WTP

By the nature of the interventions that health economists evaluate, the most salient and, often, most important benefits are improvements in patients' health. The resources used to provide an intervention will typically include the time of healthcare personnel, pharmaceuticals, and medical or surgical devices. Resource use often lends itself to be measured in money since the healthcare sector pays for it in that way, e.g., it pays wages to its staff and purchases drugs from pharmaceutical companies. Therefore, it seems quite natural that we should end up measuring benefits in terms of health gains and costs in terms of money.

However, we should not confuse money with opportunity cost. The opportunity cost of using a resource, as defined in the previous section, is the value of the best alternative use of the resource. This implies, at least for opportunity cost to have any meaning, that it must be possible to express benefits and opportunity costs using the same *numeraire*. If the goodness of a particular use (its value) depends on the size of the health improvement it could generate, opportunity cost is appropriately measured in terms of health forgone. The best alternative use is the one in which the resource could generate the greatest health improvement. On the other hand, if better uses are distinguished by their higher monetary value, then we need instead be able to somehow monetise health improvements. There are, conceivably, other ways in which one could measure value, but I shall restrict my discussion to health and money since they are the two most apparent and practically relevant alternatives.

From the present discussion, it appears that we must first be able to render different kinds of health improvements comparable. For this purpose, health improvements can usefully be categorised as improvements in either quality of life or length of life. There is no objective way in which to compare quality-of-life gains to increases in longevity, but economic evaluations will typically combine them into a single metric by assigning quality-of-life weights to different health states. Most often, a weight of zero will correspond to being dead and a weight of 1 to a state of perfect health. From here, it is possible to calculate the number of QALYs that are gained with an intervention. For example, we can gain 3 QALYs by extending life in perfect health by 3 years, extending life in fair health (here taken to mean that the weight assigned to this health state is 0.5) by 6 years, or improving health from fair to perfect for 6 years.

QALYs are useful by virtue of being a generic measure which makes it possible to compare different kinds of health gains across different kinds of disease areas. Also, even if one measures value in terms of money, QALYs could be a reasonable starting point for determining the WTP for health improvements. That is, start by calculating the number of QALYs gained

with a treatment and then determine the WTP for those QALYs. Although, it is possible to imagine scenarios in which the WTP for a treatment is radically different from the WTP for the QALYs generated by the same treatment (Johnson, 2012).

A key message of this thesis is that there are many advantages of *not* valuing health in terms of money. Still, at this point, it seems appropriate to say something about the merits of doing so. Such a discussion takes its starting point in *welfare economics*. Orthodox welfare economics is concerned with the maximisation of social welfare but simultaneously holds that there is no way to measure it. This is because it is assumes that individuals' own preferences are the basis for social welfare and that the aggregation of individual preferences (interpersonal comparisons of utility) is not within the scope of a positive science (e.g., Robbins, 1932: 120-5). Indeed, without making any value judgements this seems an inescapable conclusion.[4] According to such a restrictive view, it is very hard to say whether a healthcare intervention increases or decreases social welfare. It is only possible to know that an intervention increases social welfare if all individuals either prefer the state of the world brought about by the intervention compared to the reference case, or are indifferent between them; i.e., when the change constitutes a Pareto improvement.

However, if those that prefer the new state of the world were able to compensate those that did not, leaving themselves still preferring the change and the others indifferent, the change constitutes a *potential* Pareto improvement. An appealing theoretical justification for measuring benefits and opportunity costs in terms of WTP is that this can be construed as a form of "compensation test" (Kaldor, 1939, Hicks, 1939). If the aggregate WTP for what is gained with an intervention exceeds the aggregate WTP for the alternative use of resources devoted to providing it, the state of the world brought about by it would, in a sense, be more efficient than the reference case. It would, in principle, be possible to compensate the losers so that some people were made better off without anyone being made worse off.

Being able to say something (almost) objective about allocative efficiency is, indeed, quite appealing but it is important to realise that such a compensation test does not prove that social welfare would increase. Merely that it would if we were to also carry out the hypothetical compensation. Further, the fact that the money value of benefits exceeds the money value of forgone resource use does not prove that compensation actually could be carried out (Boadway, 1974). To use WTP as a measure of value in

[4] A value judgement has of course already been made when assuming that social welfare depends on individual preferences. However, this seems a weaker value judgement than a system for making interpersonal comparisons of utility (Boadway and Bruce, 1984).

economic evaluation, we must in practice therefore make the leap of faith that aggregate WTP is a form of direct measure of social welfare. That is not to say that WTP cannot be informative about the relative goodness of different uses of resources, but if one accepts these limitations, it loses its objective appeal. The merits of using WTP to measure value lies, rather, in the fact that money is a natural numeraire. The benefits (or disbenefits) of a medical intervention may not be restricted to length of life or quality of life. For instance, it could affect patients' ability to work and therefore lead to increased or decreased private production and consumption. Private consumption has a value that is easily measured in money, but which is not commensurable with QALYs.

In this thesis, I take the view that value is measured in QALYs. Consequently, opportunity cost must also be measured in QALYs and the best alternative use of a resource will be that which produces the greatest number of QALYs. Any effect associated with the use of a resource that is not measurable in QALYs must be disregarded, because accounting for it implies that we change our measure of value and the way in which we identify a best use. This should not be seen as a strong value judgement, because it is motivated by the realisation that it is virtually impossible to define a measure of value that healthcare decision makers are seeking to maximise. By some form of consensus, we have arrived at QALYs as a reasonable measure of one aspect of value. So while we can think of efficiency in terms of QALY or health maximisation (Wagstaff, 1991), this is highly unlikely to be the sole criterion used to set priorities. QALY maximisation could be motivated from an extra-welfarist perspective (Brouwer et al., 2008, Brouwer and Koopmanschap, 2000) but I am not concerned with the theoretical foundation for why healthcare priority setting should want to maximise QALYs (in particular) either. The aim that I pursue here is to clarify how one should think about opportunity cost with respect to this arbitrary efficiency objective, keeping in mind that the results of an economic evaluation will be used to make trade-offs between this measure of efficiency and other objectives. Still, it should be noted that theoretical and empirical considerations about what a QALY represents have implications for which effects are measurable in QALYs (Johannesson and O'Conor, 1997, Johannesson, 1995b, Nyman, 2004, Lundin and Ramsberg, 2008, Nyman, 2006, Nyman, 2011, Meltzer, 1997). I will assume that QALYs are unaffected by changes in non-medical consumption.

The bookshelf and the meaning of "displacement"

To clarify the concept of opportunity cost in the context of economic evaluation, I use the cost-effectiveness bookshelf (Culyer, 2016, Paulden et al., 2017). This graphical model assumes that health care can be divided into a number of independent interventions and illustrates these as books on a shelf. In Figure 1, I draw a bookshelf for a health system that could provide a maximum of nine interventions. The width of a book is the dollar cost of providing it and the constraint on how many books could be provided comes in the form of a bookend, which is a representation of the health system's budget. Books that fit on the shelf to the left of the bookend are provided, while those that are placed to its right are not. To provide one of the unfunded books, we would have to either replace one or several of the books on the left end of the shelf with it or move the bookend farther to the right. Assuming we are constrained by the current position of the bookend, we know from the height of a book whether it would be optimal to arrange it to the left or to the right of the bookend. The height of a book is the number of QALYs it produces per dollar. We can understand the role of economic evaluation as determining both the height and width of the books on the shelf, since this would allow a decision maker to pick books by height until no more would fit on the left end of the shelf. It seems intuitive that this maximises the number of QALYs produced. Because the area of a book's spine is the total number of QALYs it produces, we should pick the tallest books first if this is our objective. Here, we would pick A, F, K, B, and M, in that order, at which point no more books would fit and the budget would be exhausted. We refer to M as the *marginal intervention* because it was the last to be funded and should be first to be displaced if we want to release funding for some other purpose.

Weinstein and Zeckhauser (1973) show that the problem of maximising an outcome subject to a budget constraint is equivalent to deciding whether the cost per unit of effectiveness of an intervention is below a critical ratio. This cut-off has become known as the *cost-effectiveness threshold*. The bookshelf illustrates, much less technically, that the threshold is determined by the cost effectiveness of the marginal intervention (Paulden et al., 2017, McCabe et al., 2008, Phelps and Mushlin, 1991, Johannesson and Weinstein, 1993, Culyer et al., 2007). That is, when new books are introduced to the shelf, they must be taller than M to be considered for adoption. Or in terms of opportunity cost, if we consider directing funds toward the provision of a new intervention, the best alternative use of those funds is to keep providing M.

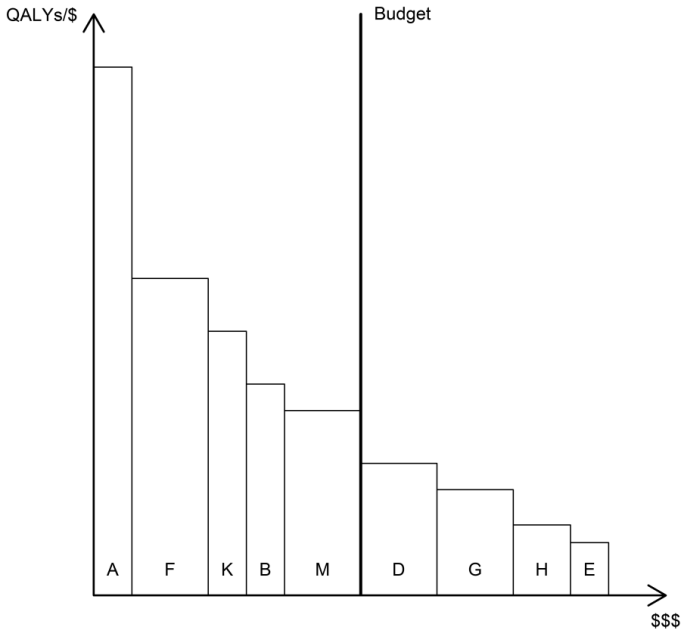


Figure 1 Bookshelf illustration of a health system which could provide a maximum of nine independent interventions

The merits of this cute metaphor are that it illustrates why the opportunity cost of funding a new intervention would be health forgone and the rationale for interpreting cost-effectiveness evidence with the help of a threshold. However, the notion that the threshold is determined by the least cost-effective currently funded intervention predates the bookshelf. I will now use the bookshelf to explore some issues with how this idea translates from theory to reality. First, assume that the books on the shelf are not independent interventions. Instead, there are three patient groups and intervention A, B, D, or E could be provided to group 1, intervention F, G, or H to group 2, and intervention K or M to group 3. That is, we can only provide one intervention to each group. The alphabetical order of the interventions within a group determines how costly and effective they are. B is more costly and more effective than A, D is more costly and more effective than B, and so on. In the absence of a budget constraint, we would always choose to provide interventions E, H, and M. Assuming there are no extendedly dominated interventions (see Paper II and Karlsson and Johannesson (1996), from which this example is borrowed), the alphabetical order also signals which intervention within a group has the highest

cost per QALY gained. In group 1, A has the lowest cost per QALY gained and E has the highest cost per QALY gained.

In a choice involving mutually exclusive alternatives, the cost per QALY gained has to be calculated as an incremental cost-effectiveness ratio (ICER) in relation to the appropriate comparator to be informative for the objective of health maximisation (Johannesson and Weinstein, 1993). For example, the ICER for B must be calculated as the incremental cost of B relative to A divided by the incremental QALY gain of B relative to A. Figure 2a provides a bookshelf illustration where the height and width of a book represents its incremental cost-effectiveness and incremental cost, respectively. To maximise health, we would first pick interventions A, F, and K, then replace A with B and K with M, before running out of funds. While this illustration is identical to Figure 1, arranging five books to the left of the bookend is no longer the same as providing five interventions. Instead, some books belong together as different volumes of the same collection. For example, intervention B is represented by books (or volumes) A and B. This is important for how we think about the marginal intervention and the opportunity cost of funding a new one. Figure 2b illustrates what would happen if we were to fund intervention D. To afford D, we would have to displace book M, but this is not the same as displacing intervention M. Displacing book M means that we release funds by providing intervention K to patient group 3 instead of intervention M.

A less abstract example may serve to clarify this interpretation. Dual antiplatelet therapy with acetylsalicylic acid (ASA), often lifelong, together with 12-month treatment of a P2Y12-inhibitor is a common treatment regimen in the prevention of further cardiovascular events in post myocardial infarction (MI) patients. When antiplatelet therapy with ASA was added to standard of care, it could be considered the first volume of antiplatelet secondary prevention. Next, a volume for the P2Y12-inhibitor clopidogrel was added as dual antiplatelet therapy improved health outcomes compared with ASA alone (Weintraub et al., 2005). Further developments in cardiovascular medicine then showed that treatment with another P2Y12-inhibitor, ticagrelor, may improve health outcomes further by replacing clopidogrel (Nikolic et al., 2013). Thus, antiplatelet secondary prevention for MI consists, crudely, of three volumes reflecting a first treatment, the addition of a new pharmaceutical to a previous one, and the replacement of an old one with a new one. Keeping the volume for clopidogrel reflects that it was used as the comparator for ticagrelor and that, at least in theory, we could reduce spending by reverting to this treatment.

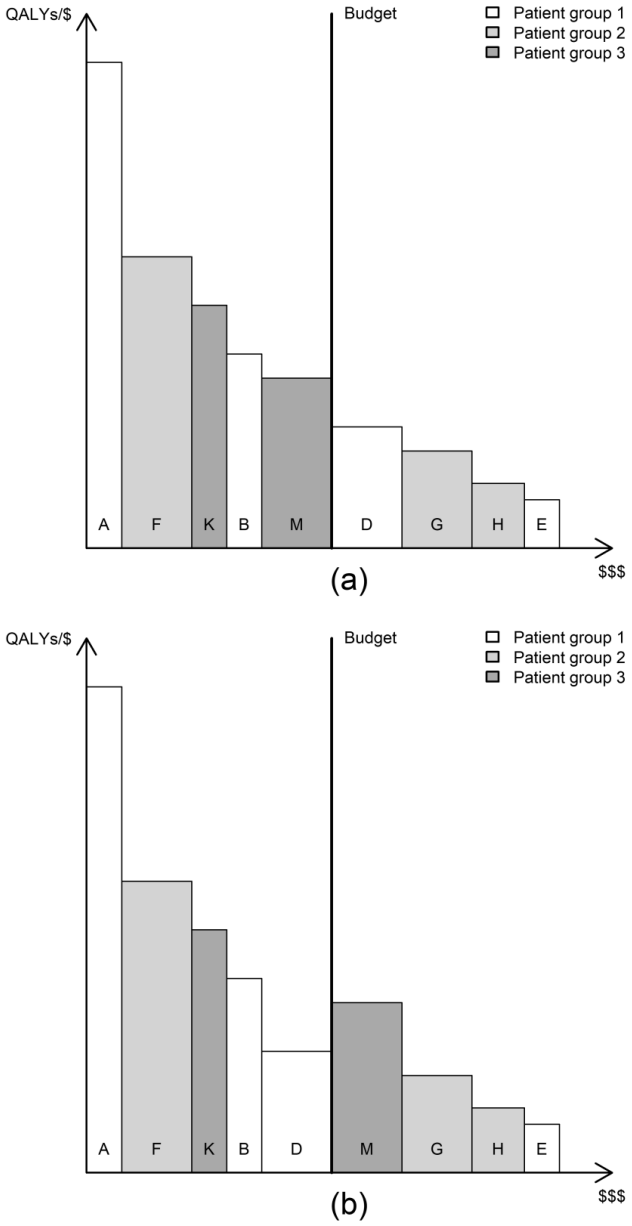


Figure 2 Bookshelf illustration of a health system which could provide a maximum of one intervention to each of three different patient groups, with mutually exclusive alternatives for treatment within each group. (a) The current system. (b) The system after adopting intervention D by reverting to intervention K for patient group 3

Another example may help illustrate that interventions are not necessarily well defined and could be divided into an arbitrary number of volumes. A study of the optimal starting age and frequency of population screening for atrial fibrillation evaluates over 2 billion different screening designs (Aronsson et al., 2017). The design with the lowest ICER is a single screening occasion at the age of 75, but the intervention could be expanded with repeated screening and initiation at a younger age, where each expansion comes with a higher ICER. Since the variations in design of a screening programme are practically infinite it should, in principle, be possible to define volumes (designs) that correspond to arbitrary increments in cost. Similarly, ASA, clopidogrel, and ticagrelor could be split into different volumes for different subsets of patients.

It is also worth noting how difficult it is to split health care into individual interventions, programmes, technologies or treatments. A hospital department, for instance, could be thought of as a collection of many volumes, where each additional volume reflects more staff, more beds, new equipment, et cetera, but ultimately, it is hard to define precisely where one collection ends and another begins. Perhaps it would be possible to push the bookshelf metaphor further to accommodate this aspect, but that is not my goal here. My intention is to show that individual books may not represent well-defined, standalone interventions. When books are included or displaced in the bookshelf, it could reflect the addition of new pharmaceuticals or the displacement of old ones, but it could also mean that already funded "interventions" are expanded or reduced in some fashion (e.g., a change in the amount of labour used by a care unit, the waiting list for a surgery, or the frequency of a screening programme).

My discussion has important implications for how we interpret the theoretical construct of the cost-effectiveness threshold as the ICER of the least cost-effective currently funded intervention. Studies attempting to identify displaced interventions have found that few treatments or healthcare services are discontinued to finance new ones (Appleby et al., 2009, Wammes et al., 2020, Karlsberg Schaffer et al., 2016, Karlsberg Schaffer et al., 2015). A possible explanation for this finding is that healthcare decision makers are reluctant to ration care explicitly and may be more inclined to dilute rather than to deny care (Klein, 2010). If this is correct, we should not think of the marginal interventions as a well-defined treatment akin to the one we are considering for adoption. Translating from theory to reality, it seems more appropriate to interpret displacement of the marginal intervention as the slight reduction of many different healthcare services (i.e., many thin books immediately to the left of the bookend).

Another explanation for the difficulty of identifying displaced interventions is that healthcare budgets tend to increase over time, which means

that there is a certain scope to avoid retracting services that are already funded. Paulden et al. (2017) discuss the possibility of new interventions imposing an opportunity cost on the health system in the long run, even when the budget is expanding. I illustrate the implications of such a dynamic budget constraint in Figure 3. Let G be a new intervention considered for adoption. Since the budget expands, funding G does not lead to any displacement in the usual sense (Figure 3b). However, if we were to not fund G, the increased budget would allow us to adopt D (Figure 3a). Thus, D constitutes the health forgone, i.e., the opportunity cost of reimbursing G. When the budget is not held fixed, the marginal intervention might be one that is currently not included in the system. It is of course possible to view D as, e.g., a pharmaceutical just waiting to be funded, but it seems more plausible to interpret it as an expansion of current healthcare services. A more general point to be made is that the opportunity cost of funding the new intervention is identified by comparing the two hypothetical health systems that would come about with and without it. This example demonstrates that the hypothetical health system in the reference scenario (Figure 3a) may differ from the current system (Figure 1). The distinction between the current system and the reference system was obscured in the previous example because the budget was held fixed.

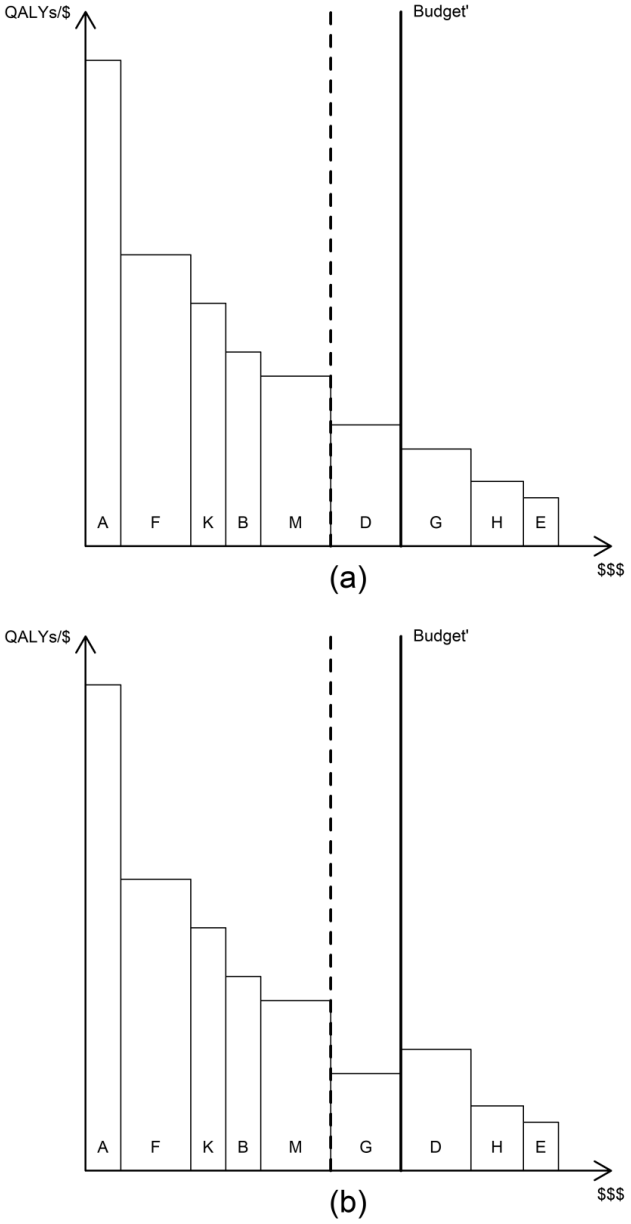


Figure 3 Bookshelf illustration of a health system where the budget is expanding. (a) The reference system. (b) The system if intervention G is adopted

Implications of inefficiency

The previous section explains the appropriate interpretation of new interventions displacing old ones, but opportunity cost as displacement has also been criticised on the grounds that displaced intervention may not constitute the best alternative use of resources (Pekarsky, 2015, Eckermann and Pekarsky, 2014). Up to this point, I have assumed that the books are ordered by height to begin with. That is, resources are already efficiently allocated. Now, consider the implications of a disorganised bookshelf for the identity of opportunity cost. In the current example, illustrated in Figure 4, the health system could provide a maximum of eight interventions (A through G and X), where X is a new intervention considered for adoption. For convenience, I assume that all interventions have the same budget impact, which will not affect our conclusions as long as we interpret "intervention" appropriately. The bookshelf in Figure 4a shows that interventions A through D are those currently funded, but that this allocation is not the most efficient. Figure 4b illustrates the composition of the healthcare system after investing in X by displacing the marginal intervention (D). According to the logic applied previously, we would identify the opportunity cost of investing in X by evaluating the scenario in 4b using 4a as a reference scenario, which would reveal the area of D to be the health forgone and the area of X to be the health gained (for convenience, the letters are henceforth used to refer both to the interventions and to the number of QALYs they produce). However, the best alternative use of the funds released by disinvesting in D is not X. Figure 5a depicts the system if the released funds were put to their best alternative use, which is the appropriate reference scenario with respect to the definition of opportunity cost as best alternative use. The scenario in Figure 5b is the same as in 4b, but here the opportunity cost is identified in comparison with 5a. This illustrates the essence of Eckermann and Pekarsky (2014)'s critique of opportunity cost as displaced healthcare services.

However, there is an additional layer to their critique. Figure 6 illustrates the case of inefficient displacement, where intervention B is displaced instead of D. The references scenarios are the same as before (4a and 5a) and the comparison of scenario 6a to 4a would reveal the opportunity cost of investing in X to be the health forgone from the displaced intervention (B) (Culyer, 2016). Eckermann and Pekarsky (2014) regard the true opportunity cost as that of the joint actions of adoption and financing. This is the health forgone by not investing in the best alternative (E) and the net health forgone from inefficient displacement (B-D). Figure 6b clarifies the rationale for this claim by highlighting that, in comparison with 5a, we gain both D (by not displacing it) and X, while E+B is the health forgone.

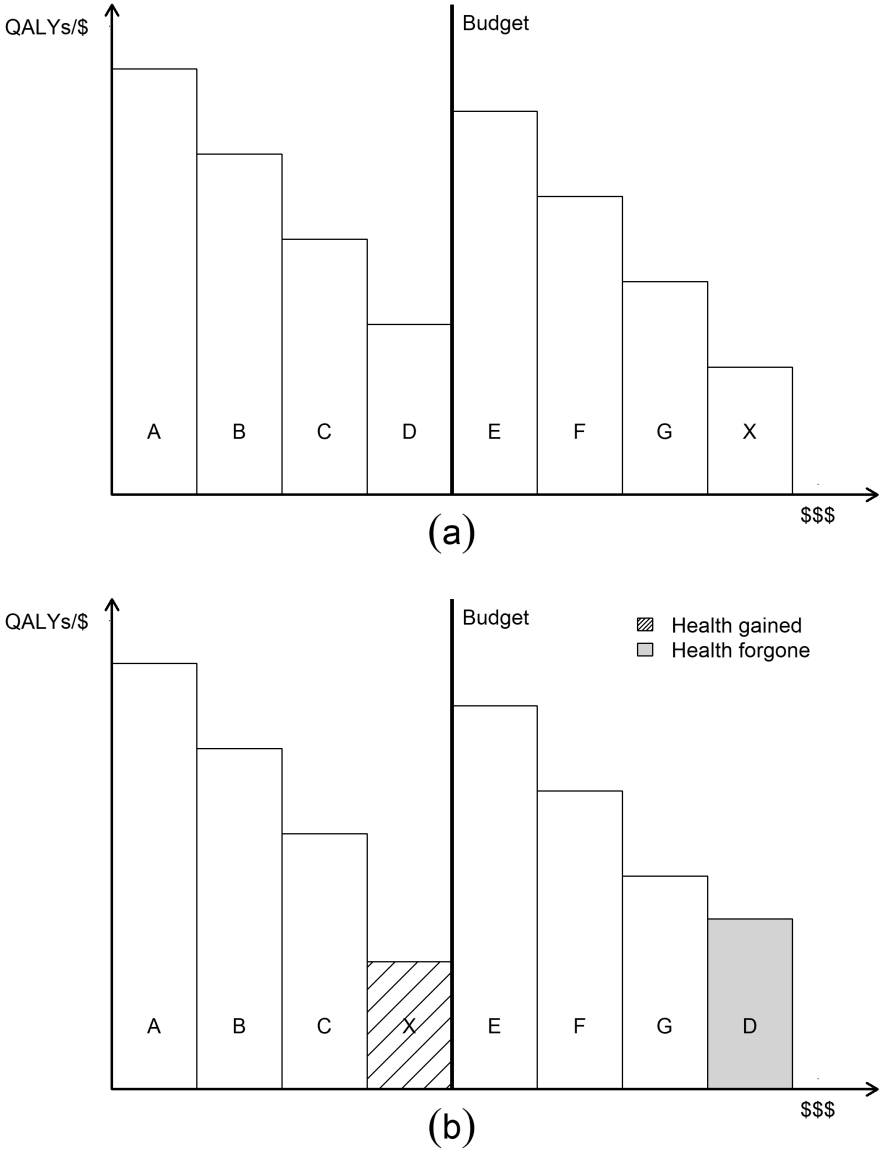
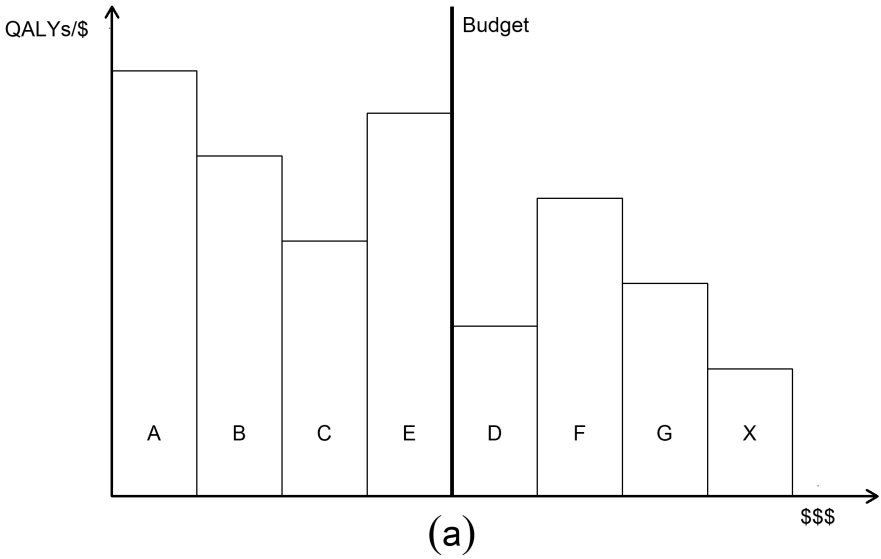


Figure 4 Bookshelf illustration of a health system adopting a new intervention, X. (a) The current system and the reference system. (b) The system if adopting X, with benefits and opportunity cost in relation to the reference system



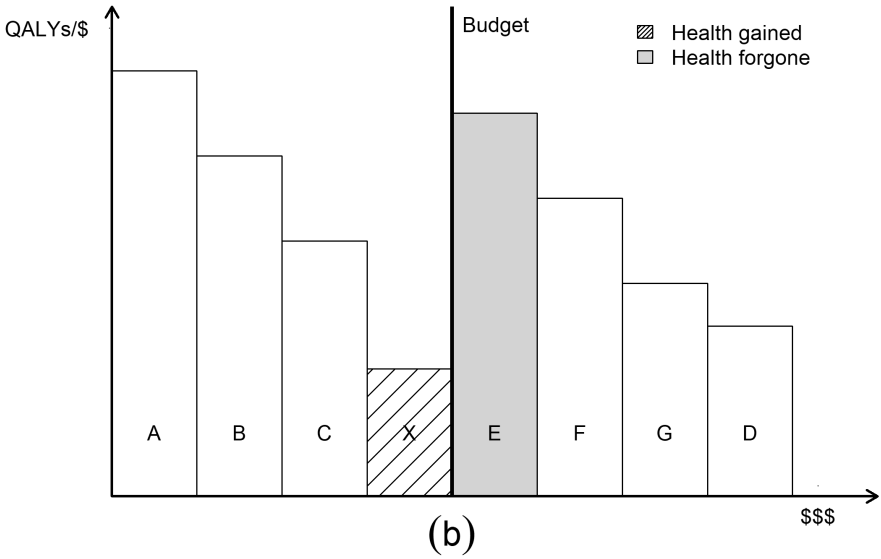


Figure 5 Bookshelf illustration of a health system adopting a new intervention, X, with best alternative use as the reference case. (a) The reference system. (b) The system if adopting X, with benefits and opportunity cost in relation to the reference system

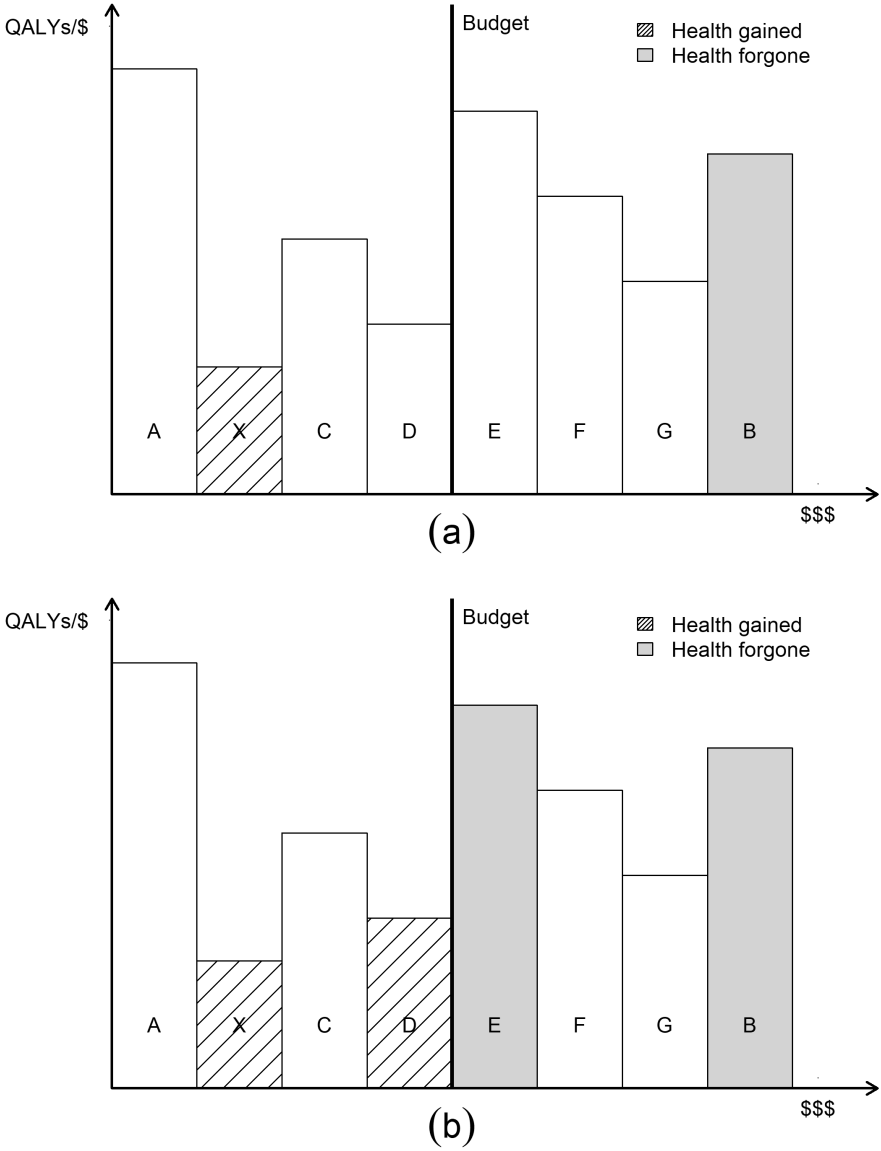


Figure 6 Bookshelf illustration of a health system adopting a new intervention, X. Benefits and opportunity cost with (a) the current system as the reference case (Fig. 4a), (b) best alternative use and efficient displacement as the reference case (Fig. 5a)

To understand why we would end up with an opportunity cost of E+B-D, recall Ferraro and Taylor (2005)'s question about Clapton and Dylan. Treating the benefit of keeping D as a negative cost is just an arbitrary accounting practice, but one that does make sense in this situation if we want to balance the value of X against *its* opportunity cost. Eckermann and Pekarsky (2014)'s point is undoubtedly correct in one sense. The best alternative to invest in is E and the best alternative to disinvest in is D. However, in the sense that it is correct, it is no longer obvious what the opportunity cost of X or that of investing in X would mean, because the act of investing has been tied up with the act of disinvesting. It may be helpful to view the choice from yet another perspective in which we spell out the options available to us. We can:

a) invest in X by disinvesting in B,
b) invest in X by disinvesting in D,
c) invest in E by disinvesting in B,
d) invest in E by disinvesting in D,
e) do nothing.

Out of these five options, (d) is the best (it produces the most health). Consequently, the opportunity cost of choosing option (a) must be the value of option (d). Option (a), then, has benefits equal to X-B and an opportunity cost equal to E-D. This is superficially different from the choice between scenario 6b and 5a, where I characterised the benefit as X+D and the cost as E+B, but in both cases, we can rearrange to get X on the benefits side and E+B-D on the cost side. This shows in what way opportunity cost is a non-fundamental and arbitrary concept (O'Donnell, 2009). When presented with the information contained in any one of the bookshelves in this chapter, we can solve the optimisation problem without a working definition of opportunity cost. The most efficient thing, obviously, is to sort the books by height (c.f., Culyer et al., 2007). Figure 7 illustrates the opportunity cost of investing in X by displacing D if we take the liberty of expanding the list of options available to us even further. Here, we do not just consider the health forgone by disinvesting in B instead of D, but also that of not disinvesting in C to finance F. After rearranging, we end up with X on the benefits side and E+B+F-C-D on the cost side.

There is nothing incorrect about this calculation either. The question is what decision it, or the one according to the definition proposed by Eckermann and Pekarsky (2014), is relevant for informing? (Culyer, 2016, Culyer, 2018).

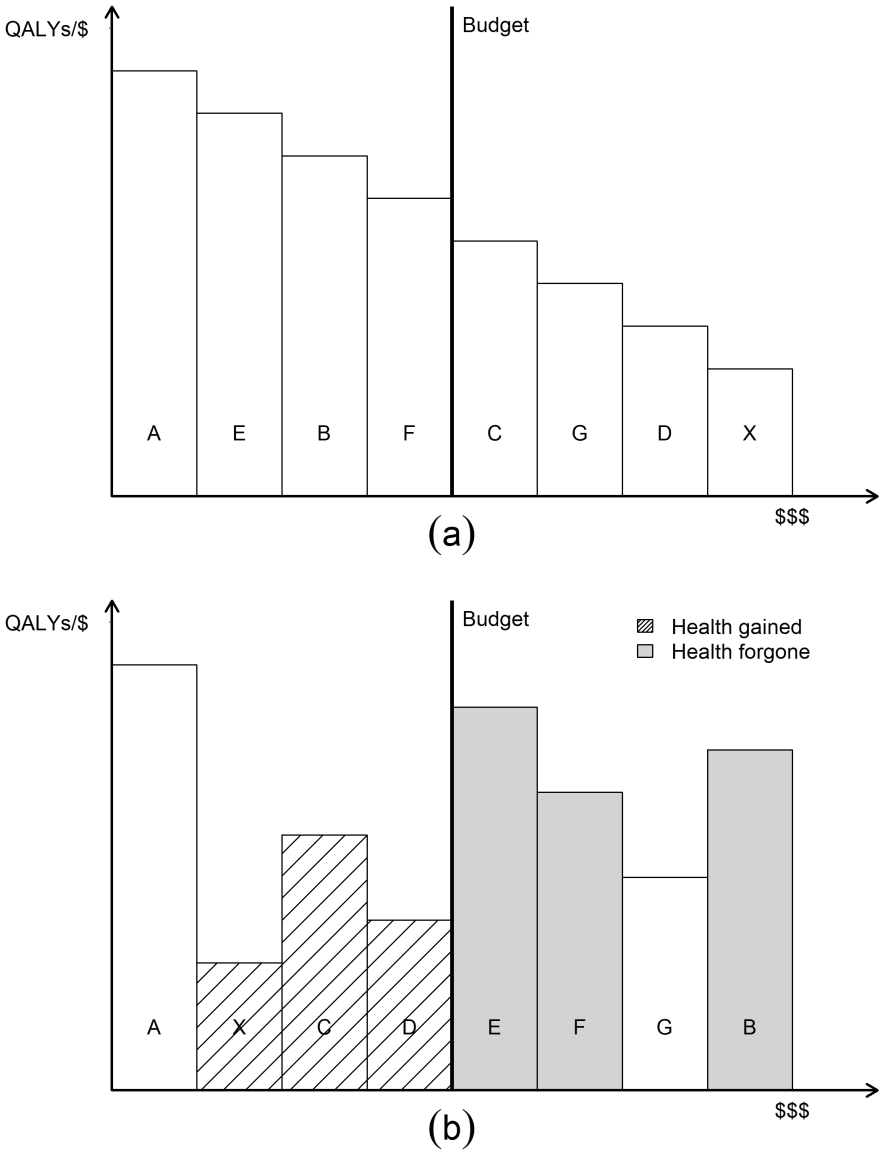
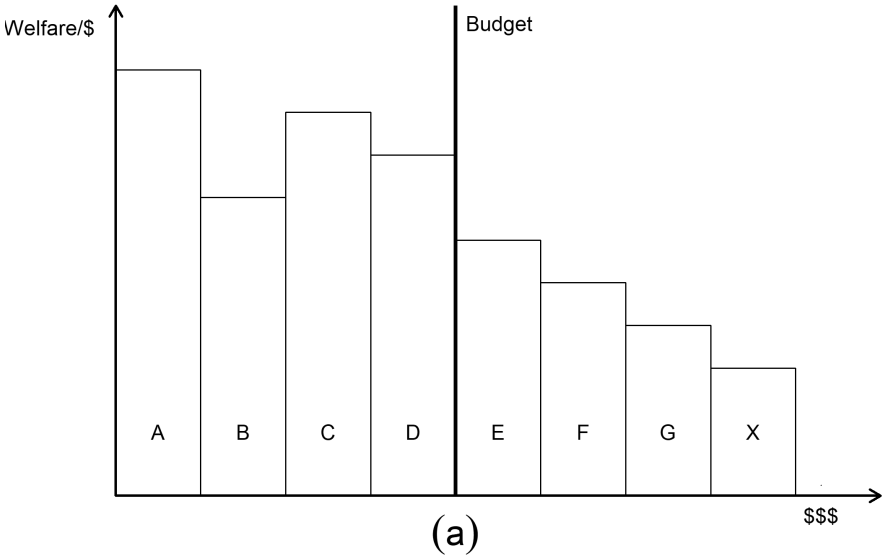


Figure 7 Bookshelf illustration of a health system adopting a new intervention, X, with full efficiency as the reference case. (a) The reference system. (b) The system if adopting X, with benefits and opportunity cost in relation to the reference system

Eichler et al. (2004) comment that decision-makers are not faced with a clean slate and that, in reality, the cost-effectiveness of new interventions is assessed on a one-by-one basis. The decision is perhaps limited to simply saying yes or no to X. In such a case, the scenario in Figure 4a is the one that *would* come about if we chose not to invest in X, and arguably therefore the correct reference scenario for determining the opportunity cost of saying yes to X. If decision-makers were able to choose between different investment options (if they could choose to invest in X or E), they might still have no control over what was displaced and the opportunity cost of investing in X would be E, not E+B-D.

Being able to quantify the opportunity cost of investing in X or using a resource to implement X seems very useful, even fundamental to an economic evaluation seeking to inform a yes or no decision about investing in X. When the scope of the decision expands to become a choice about which two of the interventions B, D, E, and X should be funded, we no longer need to know or define such a quantity (assuming we know the benefits associated with each intervention). Because opportunity cost appears to be a useful concept precisely when we are uncertain about the alternative use that we forgo, it now seems rather odd that we should define it as the value of best alternative use. Before providing what I hope can be a resolution to this oddity, I would note that a potential problem was only made apparent when considering an inefficient health system. You will find that the problem becomes even more apparent when we consider that a reason for why we have inefficiencies (the books are not ordered by height) may be that decision makers are not trying to maximise QALYs. Figure 8a illustrates a welfare maximising decision maker's view of the bookshelf. On this shelf, we have some measure of welfare (according to the decision maker) per dollar on the y-axis and the books are almost ordered by height, with the exception of intervention B. In Figure 8b, we have a QALY maximiser's view of the same bookshelf. If the decision maker displaces efficiently according to her own perspective, she would indeed displace B to release funds for new investments. According to a QALY-maximising perspective it would be better to displace D, but to insist that this is the opportunity cost seems rather pig headed if we anticipate that the decision maker will in fact displace B.

The apparent solution is to use the decision maker's definition of welfare as our outcome instead of QALYs. If that were possible, I can think of no argument against it. My assumption in this thesis is that there is no such measure of value available to us. If we recognise that we are dealing not just with a second-best state of the world but also a second-best analysis (c.f., Culyer, 2016) where the measure of value does not align perfectly with the decision maker's objective, we may have to rethink our definition of opportunity cost.



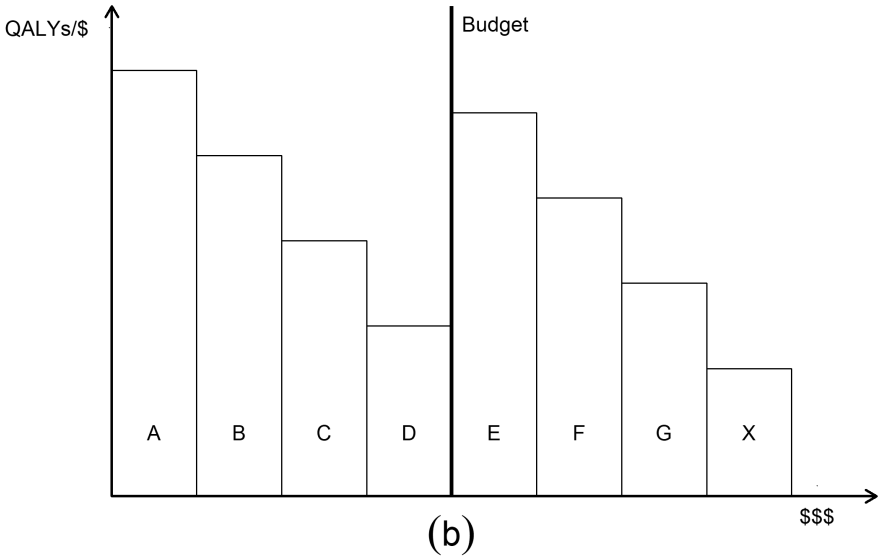


Figure 8 Bookshelf illustration of a health system from the perspective of (a) welfare maximisation. B is the least prioritised, currently funded intervention; (b) QALY maximisation. D is the least cost-effective, currently funded intervention

Expected alternative use

Instead of the value of best alternative use, we should consider that opportunity cost could also be defined as the value of *expected* alternative use. In practice, this may actually be a more commonly used definition in economic evaluation (Culyer, 1978), even if it is not an explicit one. Let us use the following example to make the distinction between best and expected:

> You are a healthcare decision maker considering the adoption of intervention A. If you were to not adopt A, you are almost certain that the funds *would* be used for intervention B. However, you also know that the funds *could* be used for intervention C and that C produces better outcomes than B.

Here, C is the best alternative use and B is the expected alternative use. There are several reasons why we would be more interested in the expected alternative use. First, it aligns with the common-sense argument that we should not let the perfect become the enemy of the good. If A produces more health than B, it is peculiar to conclude that the adoption of A leads to a net health loss. It does compared to C, but we do not regard the adoption of C as a relevant reference case in this example. Second, and on a related note, denying funds for a new intervention on the grounds that they, hypothetically, *could* be used for something else may seem unreasonable to many decision makers (Paulden et al., 2014). That something more productive is *expected* to be forgone provides a far more compelling argument for saying no. We might say that the moral weight of this argument is greater.

I can think of one reason why we would still prefer to consider best alternative use. Decisions based on expected use could lead to sub optimisation by reinforcing an inefficient use of resources, or accepting an inefficient decision-making process, and thereby preventing more efficient investments in the long run. In other words, our expectation on alternative use is incorrect or even self-fulfilling. However, there is a counter argument which I have already visited. Intervention B could be preferred to C on the grounds of some aspect of value not reflected by QALYs. An economic evaluation measuring value in terms of QALYs but trying to inform decision making that is not concerned with QALY maximisation should avoid imposing its own definition of "best" on the decision maker. Another favourable aspect of avoiding best is that opportunity cost becomes an empirical matter. We can, in principle, identify the expected alternative use of a resource without making value judgements about what makes one use better than another. This makes opportunity cost measurable in physical quantities forgone instead of value forgone (c.f., Parkin, 2016). At some point, however, we do need to make a leap from quantities to value by deciding which effects are relevant to the decision we are trying to inform.

In the past, discussion on the appropriate basis of a cost-effectiveness threshold has focused on the role of perspective and budget restrictions. The assumption that opportunity cost is health forgone has been motivated by a fixed budget and a healthcare sector perspective (Claxton et al., 2010, Brouwer et al., 2019). From my discussion it should be clear that if the objective is to maximise QALYs, then the best alternative is always the one that would produce the most QALYs, regardless of whether the budget is fixed or flexible. The WTP for a QALY is irrelevant to the objective of QALY maximisation. If the objective is to maximise monetary value, we do need to know WTP for a QALY (if we measure health effects in QALYs), but the best alternative use of a resource is similarly independent of the budget constraint. A funny and, I believe, overlooked consequence of opportunity cost as best alternative use is that the best use could be found in a public sector other than the one an evaluation is focused on. For instance, when conducting a cost-benefit analysis of infrastructure investments, the appropriate shadow price of a resource could be the monetary value of health forgone by not devoting that resource to health care; not the WTP for the resource. That it is unusual to reason in this way goes to show that budget restrictions do matter. We typically care about expected alternative use.

Summary and implications

To define opportunity cost as value in best alternative use, we must have a measure of value in mind, because without a measure of value, we cannot identify a best use. If value is measured in QALYs, then the opportunity cost of using a resource must always be QALYs forgone. Contrary to past criticism of a QALY-maximising perspective, it appears irrelevant whether healthcare budgets are fixed or flexible or whether there actually is a well-defined budget subject to which QALYs are maximised (Johannesson, 1995a). However, I have argued that value in expected alternative use is a more appropriate definition of opportunity cost in economic evaluation. The best alternative use of a resource could be a very unlikely one. That treatment should be denied because the funds could be used for something better does not seem like a compelling argument when they almost certainly would not.

The discussion leaves us, hopefully, a little bit wiser, but does not provide the solid foundation for always conceiving of trade-offs as health gained versus health forgone that I had hoped for. In some situations, we will be reasonably confident that the adoption of a new intervention crowds out other uses in health care. For example, if a hospital introduces a new routine among its nursing staff which takes up some of their time, then the most reasonable expectation is that we forgo some other use of that time (e.g., looking after patients). In other instances, it is less clear whether we

should expect the opportunity cost to be borne by health care, taxpayers or other public sectors. Most notably, prescription pharmaceuticals in Sweden are financed by the central government and not by the regional councils responsible for most other healthcare services. Because the regions receive a lump-sum payment which is not ring-fenced for drug spending and due to the mysteries of internal budgeting, there is reason to believe that the adoption of a new drug will crowd out other health spending in the short run, but in the long run, it is hard to know what to expect. This is a matter which should be investigated empirically, but one which I do not address in this thesis. In the very long run, drug spending could not grow unhindered at the expense of other public sectors. Ultimately, it would have to crowd out other health spending. This provides one justification for assuming that the opportunity cost of healthcare resource use is always health forgone. However, I am not yet willing to commit to this as a well-founded expectation. I have resisted use of the term *health opportunity cost* (Sculpher et al., 2017), because opportunity cost should be either expected alternative use (which might not be health) or best alternative use. Alas, I am forced to conclude that this is where I end up. If we want to always measure costs as health forgone, then currently, the best argument is that resources would *or at least* could have been used for other healthcare services.

The most important practical implication of the discussion in this chapter is that we should not look primarily to well-defined, standalone interventions when we try to identify opportunity costs. There are some exceptions. For example, Chen et al. (2018) identify procedures with long waiting lists in Ireland, which would or could be eliminated instead of funding new interventions. However, there is also a need to generate cost-effectiveness evidence on non-patented technologies (Pekarsky, 2015) which we might be forgoing. In the next chapter, I will focus on the use of regression to estimate the marginal health benefits of healthcare resources and healthcare spending, which is in line with a less literal interpretation of what the marginal intervention is.

Although it has served mainly as an analytic tool for my discussion, I would like to conclude this chapter by highlighting some limitations of the cost-effectiveness bookshelf. In Paper II, I state that the bookshelf could be a reasonable characterisation of a health system, contrary to what one might think from its simplified presentation. This is true in the sense that concerns about constant returns to scale or the divisibility of interventions (Birch and Gafni, 1993) appear irrelevant when we give “intervention” a more flexible interpretation. However, there are aspects of a real-world health system which we are not able to capture in the bookshelf. First, and most obviously, we cannot illustrate cost-decreasing interventions; they would have negative width. More important, and less obviously, we cannot

illustrate dominated interventions for the same reason, or extendedly dominated ones if the height of a book is to maintain its relevance to the objective of health maximisation. This means that it is not possible to illustrate the type of inefficiencies that would result from adopting interventions that should never be adopted from a health-maximising perspective. A health system which has nevertheless adopted these types of interventions could not be represented by a bookshelf, even in principle. Despite these limitations, the bookshelf provides an intuition for the theoretical construct of the marginal intervention, and in this chapter, I have used it to explain how this construct might be observable in the real world.

III. EMPIRICAL ESTIMATES

Methods for estimating the marginal cost of producing health

The problems of estimating the effect of healthcare spending on health outcomes are well known (Gravelle and Backhouse, 1987). I briefly explain the essence of these problems using data from Paper I on life expectancy and healthcare spending for 20 Swedish regions. If we plot the relationship between life expectancy and spending in 2015, we see that regions that spend more per capita on health care have a lower life expectancy on average (Figure 9a). This should obviously not be taken at face value as evidence of health spending being harmful. A region may choose (or be forced) to spend more on health care than another, precisely because of the relatively poor health of its population. Figure 9b instead plots the relationship between national average life expectancy and spending over time for 2003 to 2015. Here, we find a positive association between spending and life expectancy. This is more in line with (my) prior beliefs about a ceteris paribus effect of spending more on healthcare, but it would be a mistake to take this association at face value as well. For example, increases in spending could be driven by increases in income, which either cause or coincide with other causes of longevity.

These are both standard examples of confounding (alternatively omitted variable bias, selection bias, or spurious correlation), which means that there is a third, unmeasured variable causing both life expectancy and healthcare spending. It is this variable (the confounder) rather than healthcare spending which explains the association we observe in the data. Although these examples suggest that we would be concerned about underestimating the effect of spending when making comparisons between regions and overestimating it when making comparisons over time within the same region or country, confounding factors may not work in the way that I have theorised. If regions with relatively poor population health are also relatively poorer income-wise, they may be unable to spend more even if they should like to do so. Further, less public provision of health care could be offset by more private spending or informal care.

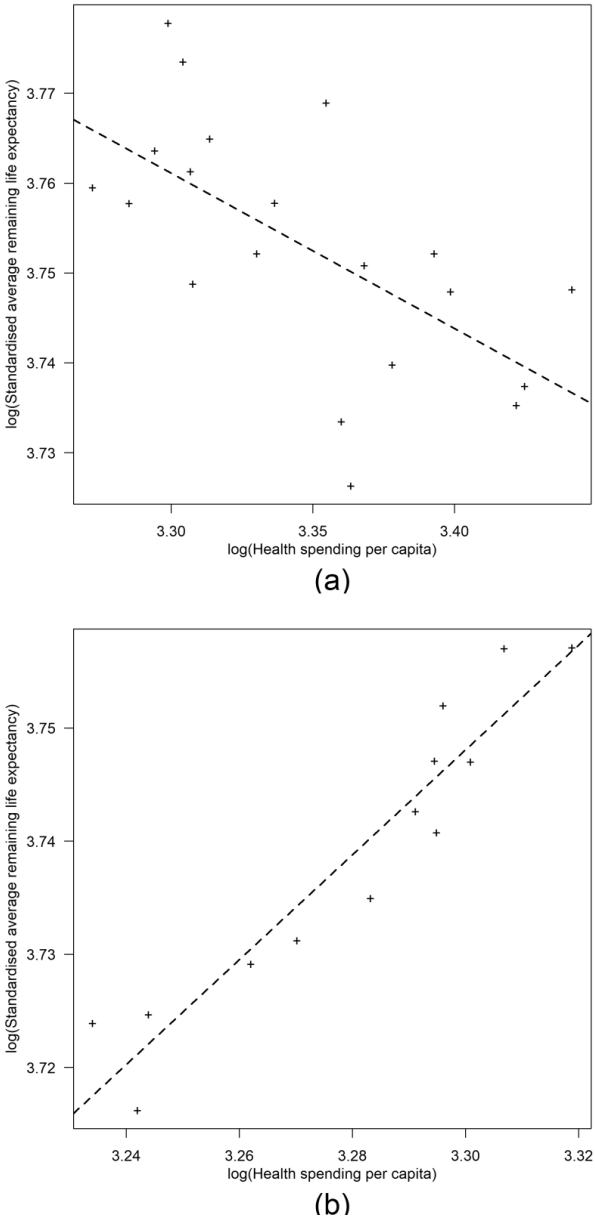


Figure 9 Association between health spending per capita and age-sex-standardised average remaining life expectancy, in logarithms. (a) 20 Swedish regions in 2015. (b) Swedish national averages, 2003-2015. Dashed lines are conditional expectations estimated by OLS

The point is that cofounding makes it tricky to say anything about the effect of spending more on health care based on the observed association between spending and health outcomes. Ideally, for the association to have an unambiguous causal interpretation, we would like to randomly assign different levels of spending to different regions. Because it is not within my power to do so, I will have to resign myself to deal with confounding in some other way.

I will consider two ways of dealing with confounding in observational data. The first is to use a *selection-on-observables* strategy in which we assume that it is possible to measure and control for all confounding factors. By holding fixed all factors causing both spending and health outcomes we can, in principle, recover the ceteris paribus effect of spending. The obvious challenge with this approach is to provide an exhaustive and convincing list of potential confounders. In doing so we, must also make sure not to control for the wrong factors. For example, it seems that we should like to control for healthcare need (alternatively poor health or morbidity) because this might explain why regions with worse outcomes are spending more on health care. However, if we were to measure need as the number of people hospitalised for a certain medical condition, we would be controlling for one of the ways in which more spending can be expected to improve health. If spending goes towards the provision of, e.g., more hospital beds, which leads to more hospitalisations, then this is not a factor that we would like to be held fixed. Additionally, adjusting for factors that are outcomes of spending could introduce new confounders that were not previously a problem (Angrist and Pischke, 2009).

Instead of removing variation in health outcomes and healthcare spending that confounds the relationship between the variables, the second approach is to identify one or several specific causes of health spending that are not confounders. Assuming these factors are as good as random, i.e., there are no confounders in the relationship between these factors and the outcome, they can be used as *instrumental variables*. In principle, the ceteris paribus effect of spending on the health outcome can be derived by scaling the association between an instrument and the outcome by the association between an instrument and spending. For example, if a unit-change in an instrument is associated with $2 million more in spending and 1% lower mortality, then it implies that the 1% reduction in mortality was due to those additional millions. In practice, identifying factors that satisfy the criteria of a good instrument is hard to say the least. However, we could hope to find determinants of spending which are plausibly as good as random once we control for other factors. Instrumental variable strategies are typically described as natural (or quasi-) experiments by virtue of utilising (as good as) random variation in the treatment variable, but if we rely on adjustment for covariates to make the case that an instrument is as

good as random, the strategy essentially reduces to selection on observables. Still, it could produce somewhat more credible results by isolating a specific source of variation in spending for which confounding appears to be less of a problem.

Many papers have investigated the relationship between mortality and health spending (Gallet and Doucouliagos, 2017), but few have done so with a causal interpretation in mind with the purpose of informing a measure of opportunity cost. To my knowledge, the study by Martin et al. (2008) constitutes the first attempt in this this respect. Using cross-sectional data on mortality (years of life lost, YLL) and the proportion of lone pensioner households and the proportion of the population providing unpaid care as instruments for spending, they find that the costs per life year gained from spending more on cancer and circulator disease are about £13,000 and £8,000, respectively. This analysis is updated for 10 programmes of care by Martin et al. (2012), and later by Claxton et al. (2015a) to provide an overall cost of £13,000 per QALY for the English National Health Service (NHS) across 23 programmes of care in 2008. Lomas et al. (2019) conduct similar analyses showing that this method produces estimates between £5,000 and £15,000 per QALY when applied to ten different cross sections between 2003 and 2012. Subsequent English work has used a different set of instruments for spending based on the funding formula used to allocate resources between English primary care trusts, with estimates of the marginal cost of producing a QALY between £5,000 and £10,000 (Martin et al., 2021). These instruments are indices used by the department of health to calculate compensation for population age, healthcare input prices, and gradual adaption to changes in allocation according to the funding formula.

Following Claxton et al. (2015a), Edney et al. (2018) also use cross-sectional data and the proportion of the population providing unpaid care as an instrument for spending to estimate the cost of producing a QALY in the Australian healthcare system at AUD 28,033. Other studies have instead used panel data to account for time-invariant, unobserved confounders. van Baal et al. (2019) use a panel on cardiovascular mortality and cardiovascular hospital spending in different age-sex groups in the Netherlands between 2003 and 2010 to estimate the effect of spending on mortality in first differences, with fixed effects for groups and years. They derive a marginal cost per QALY of €41,000 from cardiovascular spending. Vallejo-Torres et al. (2018) use a panel on healthcare spending and quality-adjusted life expectancy in 17 Spanish regions between 2009 and 2013, with fixed effects for years and regions to estimate a cost per QALY of about €25,000.

The first step in the studies that I have briefly reviewed is to estimate an effect of healthcare spending on mortality, and in a second step to

incorporate quality-of-life effects.[5] This involves weighting life years gained by appropriate quality-of-life norms, but also making an approximation of pure quality-of-life effects (Edney et al., 2022). In this thesis, I focus primarily on the mortality effect of spending, which forms the backbone of this type of study. I now move on to my own attempt to estimate the marginal cost of a QALY in Sweden.

Evidence from Sweden

To provide an estimate of the Swedish healthcare system's marginal cost of producing health, I use a panel on average remaining life expectancy (ARLE) and public healthcare spending per capita for 20 Swedish regions between 2003 and 2015. ARLE is constructed by averaging 202 age-sex-specific life expectancies (male aged 0-100+, female aged 0-100+) for each region and year using constant national population weights based on the year 2016. This means that variation in ARLE reflects variation in age-sex-specific mortality but not variation in the composition of the population. Spending per capita is transformed into fixed prices using a healthcare price index (see Paper I), thus reflecting a measure of resource use rather than nominal spending. I use a log-log specification for ARLE as a function of spending per capita, which makes the coefficient on spending interpretable in terms of an elasticity.

As discussed in the previous section, there are reasons to expect both positive and negative bias from confounders in the relationship between spending and life expectancy. Regions with high socioeconomic status and a healthier demographic profile will presumably have higher income and be able to afford more healthcare. We can expect this to be offset to at least some extent by Sweden's regional redistribution system, which redistributes funds between regions based on income and transfers additional funds from the central government to the regions based on socioeconomic and demographic factors (Swedish Agency for Public Management, 2014). Further, the same factors are likely to be important determinants of the population's need for healthcare and high-need regions could opt for a higher tax rate which would allow them to spend more on health care.

My strategy for tackling the problem of confounding is to use potentially exogenous sources of variation in nurse labour supply as instruments for spending. Figure 10 provides a stylised illustration of the rationale for this strategy. If we assume that a region acts to maximise some output related to life expectancy using only nurses and one other input, then a region with a lower cost of hiring nurses would, all other things being equal,

[5] With the exception of Vallejo-Torres et al. (2018) who weight their life-expectancy outcome using quality-of-life norms

produce at a higher level of output using more nurses; potentially, it would also use more or less of the other input depending on the size of the income and substitution effects (in Figure 10, these effects exactly offset each other). In fixed prices, this should be reflected by higher spending in the region using more nurses. I attempt to capture such variation using the number of newly graduated nurses per capita and the proportion of nurses nearing retirement (adjusted for the proportion of the working-age population nearing retirement) as instruments for health spending.

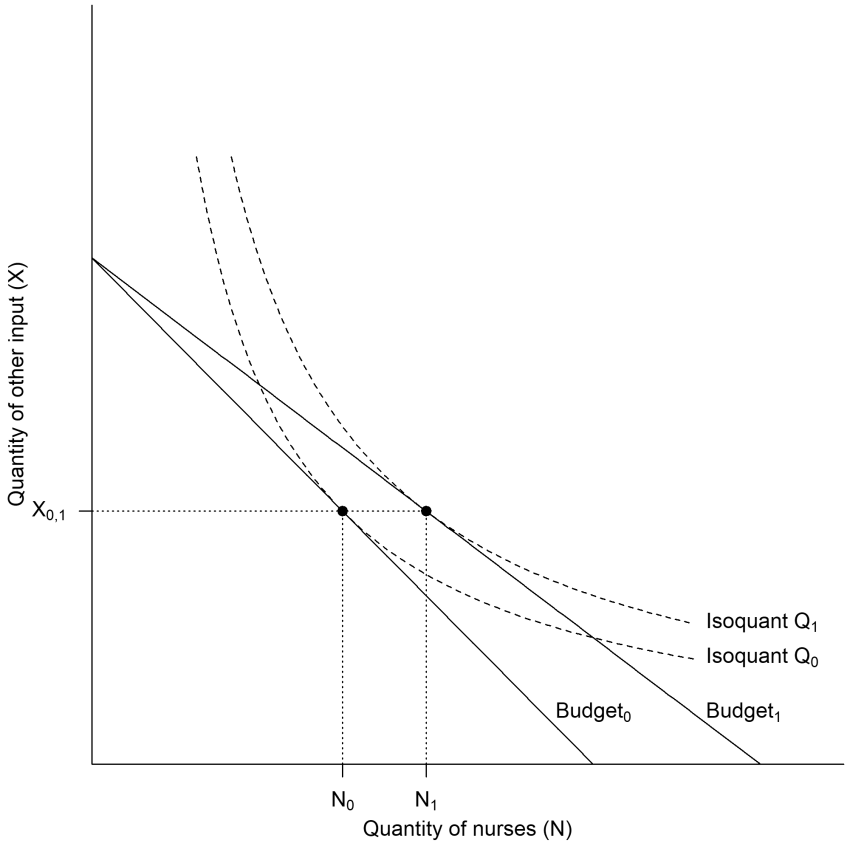


Figure 10 Stylised illustration of the effect of nurse labour supply on real health spending. An increase in the supply of nurses lowers the cost of employing nurses; it is possible to employ more nurses (the budget line pivots). A region seeking to maximise output (Q) chooses to produce at a higher level of output (Q_1) by employing more nurses (N_1). Nominal spending is unchanged, real spending increases

An advantage of using panel data is that we can adjust for unobserved confounders using fixed effects. Ideally, we would like to use fixed effects for both regions and years (two-way fixed effects, TWFE), which would remove all time-invariant differences between regions and account for national trends in the instruments and life expectancy. However, this would also discard potentially useful variation. For example, dummies for regions and years explain 94.5% of the variation in log ARLE. Instead, I use fixed effects for years only, which means that I rely on differences *between* regions for estimation. This maintains more variation in the data but at the risk of including differences between regions in nurse labour supply and ARLE that do not reflect a causal effect of the former on the latter.

Table 1, column 1 reports my regression results without adjusting for any covariates. Both the ordinary least squares (OLS) and two-stage least squares (2SLS) estimates of the spending elasticity of life expectancy are negative. However, crude differences between regions in nurse labour supply are probably not as good as random. To make such an assumption slightly more plausible, I begin by controlling for the population's education level (the proportion of the population with at least three years secondary education), the presence of a teaching hospital, and the logarithm of population density. Column 2 shows that this produces a positive and significant 2SLS estimate of the elasticity (0.162, s.e. = 0.058). Further, we find that spending and life expectancy increase with graduated nurses and decrease with nurses nearing retirement, which is in line with the rationale for the use of these variables as instruments. The regression reported in Column 3 adds a dummy for the Norrland regions and the proportion of the population that is employed, which leads to a larger elasticity (0.207, s.e. = 0.08). Column 4 adjusts for covariates measuring healthcare need (or population morbidity). This results in a much smaller but still significantly positive elasticity (0.085, s.e. = 0.037). The decrease in the coefficient on spending could indicate that bias from need leads us to overestimate the effect of spending on life expectancy. However, an alternative interpretation is that these variables are bad controls (Angrist and Pischke, 2009) in the sense that they measure the causal mechanism by which health spending affects life expectancy. I measure need using the number of patients hospitalised for first-time myocardial infarction, lung cancer, injuries, and alcohol-related diagnoses. The number of patients hospitalised per year is probably less sensitive to the amount of healthcare provided compared to the number of hospitalisations, but it is possible that regions that spend more also hospitalise more patients as a consequence. Finally, column 5 adds controls for the age and sex of the population. Although I control for variation in mortality from age and sex through standardisation of the outcome, there might be an indirect influence via need. For example, a larger share of elderly people would not mechanically result in lower ARLE, but it

could mean a greater need for health care. If two regions have the same level of spending, but one has a greater need for healthcare than the other, then we may expect that it will have worse health outcomes. Adjusting for age (and sex) is complicated by the fact that the age composition of the population must be determined by its mortality rates. To avoid problems with endogenous regressors, I use changes in mean age and the proportion of males from migration between regions. The addition of these covariates has a negligible impact on the results.

Table 1 OLS and 2SLS regressions of log average remaining life expectancy (ARLE) on log healthcare expenditure per capita for 20 Swedish regions, 2003-2015 (n=260). Log graduated nurses per capita and the proportion of nurses nearing retirement are used as instruments for expenditure.

	(1)	(2)	(3)	(4)	(5)
OLS					
log(Expenditure)	−0.122	−0.097	−0.050	0.001	0.000
	(0.010)	(0.007)	(0.009)	(0.011)	(0.010)
Reduced Form					
log(Graduated nurses)	−0.001	0.003	0.003	0.002	0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
pr(Nurses, age 60-69)	0.024	−0.027	−0.024	−0.012	−0.011
	(0.007)	(0.014)	(0.018)	(0.011)	(0.013)
First Stage					
log(Graduated nurses)	0.050	0.029	0.026	0.029	0.031
	(0.011)	(0.007)	(0.008)	(0.007)	(0.007)
pr(Nurses, age 60-69)	0.007	−0.006	−0.019	−0.073	−0.059
	(0.055)	(0.023)	(0.024)	(0.045)	(0.039)
2SLS					
log(Expenditure)	−0.068	0.162	0.207	0.085	0.080
	(0.038)	(0.058)	(0.080)	(0.037)	(0.037)
Reduced Form R^2	0.017	0.376	0.704	0.787	0.789
First Stage R^2	0.067	0.475	0.522	0.704	0.710
First Stage F-value	20.641	13.486	10.828	16.311	17.934
Sargan p-value	0.056	0.149	0.187	0.528	0.432

Notes: The reduced form is a regression of log ARLE on the instruments. The first stage is a regression of log expenditure on the instruments. Regressions reported in column (1) include no covariates. The following covariates are added incrementally: (2) pr. at least 3-year secondary education, log population density, teaching hospital dummy, (3) pr. employed, Norrland dummy, (4) patients with first-time myocardial infarction, new lung cancer, alcohol related diagnoses, and injuries per capita, (5) change in mean age an pr. male from inter-regional migration; 'pr' = proportion. All regressions include fixed effects for years. Heteroskedasticity, autocorrelation and cross-sectional correlation consistent standard errors (Driscoll and Cray, 1998) in parentheses. R^2 is calculated after residualising with respect to the fixed effects.

Assuming there is no residual confounding and none of the controls are bad, we can use the estimated elasticity to derive the marginal cost of producing a life year. For a permanent increases in spending of 1 SEK during remaining life years, the cost per life year gained is the spending per capita divided by the elasticity (see Paper I). For an elasticity of 0.076 and SEK 27,827 per capita (national average spending in 2016), the marginal cost per life year is about SEK 370,000.[6] I attempt to translate this to the marginal cost of producing a QALY in two steps. First, based on health-related quality-of-life norms (Burström et al., 2014) by age and sex applied to the Swedish population in 2016, we would gain 0.855 QALYs per life year. The cost per mortality-related QALY would be SEK 430,000. Second, because I have no means to estimate any pure quality-of-life effect of spending, I use the results from Claxton et al. (2015a) as a surrogate. They similarly first derive a cost per life year (£25,214) and mortality-related QALY (£30,270), implying 0.833 QALYs gained per life year. Including pure quality-of-life effects brought their estimate to £12,936 per QALY. Assuming that my cost per mortality-related QALY would scale similarly (£12,936/£30,270=0.427) makes the cost per QALY SEK 180,000.

Table 1 also reports some diagnostics on the 2SLS results, which can be used to assess their validity. The F-values from Wald tests on the exclusion of the instruments from the first stage regressions (i.e., the regressions of spending on the instruments) fall between 10 and 20, which is above the typical rule-of-thumb threshold for weak instruments (Stock and Yogo, 2005, Staiger and Stock, 1997). Further, it is not possible to reject the null of valid instruments in a Sargan test of over-identifying restrictions in any of the regressions including covariates, which lends some support to the validity of the instruments across all these specifications. We can also make a crude assessment of the plausibility of the first stage. Regressing the log of nurses per capita on the instruments shows that more graduated nurses and fewer nurses nearing retirement are indeed associated with more nurses employed. Regressing spending on nurses in log-log by 2SLS suggests that a 1% increase in nurses leads to a 0.25% increase in spending at the margin, or SEK 760,000 per nurse at the sample mean (8.76 nurses per 1,000 people). Assuming the instruments only affect the use of nurses and no other inputs, this would imply nurses being paid about SEK 40,000 per month at the margin, which does not seem too far off the mark (in 2016, the average monthly wage for a non-specialist nurse was SEK 33,800).

Even if these analyses support the validity of my results to some extent, it is hard to explain why the coefficient on spending would be sensitive to

[6] 0.076 is the elasticity reported as the central estimate in Paper I, which is from the same regression as in Table 2, column 5 but dropping the covariates which are not significant. Using 0.08, reported here, would result in a cost of about SEK 350,000 per life year.

adjustment for covariates if the instruments were good. That the Sargan test fails to reject despite large coefficient movements raises some doubts about relying on this diagnostics test as an indication of credible results. I now turn to how may think about the issue of credibility, how this study falls short of providing compelling evidence on causality, and how these shortcomings can be addressed.

Credibility and experimental evidence

A good instrumental variable should be as good as randomly assigned, but as is clear from my exercise in the previous section, it is hard to argue that this assumption holds without relying on arcane diagnostics testing. In fact, no study to date on the marginal cost of producing a QALY has presented a convincing case of a good instrument. Instead, instruments have been assumed to be as good as randomly assigned when adjusting for a number of covariates, but unless it is clear which covariates that are necessary to adjust for, this essentially just shifts the problem of unobserved confounding to another level. There are many more covariates that I could have adjusted for, and this might have reduced the coefficient on spending further (which is the major concern if we are being conservative) but might also have made it larger. This problem does not only apply to studies using instrumental variables but is an issue with the credibility of selection-on-observables strategies in general (Altonji et al., 2005). The words of Douglas Adams at the beginning of this thesis capture my frustration with trying to convince myself and a sceptic reader that SEK 180,000 per QALY can be given a causal interpretation (coupled with the hope that it can). The following passage expresses a similar sentiment, but speaks more closely to the matter at hand:

> The econometric art as it is practiced at the computer terminal involves fitting many, perhaps thousands, of statistical models. One or several that the researcher finds pleasing are selected for reporting purposes. This searching for a model is often well intentioned, but there can be no doubt that such a specification search invalidates the traditional theories of inference. The concepts of unbiasedness, consistency, efficiency, maximum-likelihood estimation, in fact, all the concepts of traditional theory, utterly lose their meaning by the time an applied researcher pulls from the bramble of computer output the one thorn of a model he likes best, the one he chooses to portray as a rose (Leamer, 1983).

Considering that the only (almost) fail-safe way of avoiding confounding is random assignment, it seems appropriate to briefly survey the experimental evidence on the link between health spending and health outcomes. To my knowledge, there are two relevant pieces of randomised evidence,

both from the United States in terms of health insurance coverage. The RAND health insurance experiment (RAND HIE) randomised about 4,000 individuals representative of the abled-bodied population aged 14 to 61 to different insurance plans. Those assigned to the plan with full coverage utilised more care but saw on average no improvement in health status (Brook et al., 1983). The more recent Oregon health insurance experiment (Oregon HIE) randomised offers of health insurance to about 30,000 households who signed up for the state's waiting list for spots in its public insurance scheme for low-income adults aged 19 to 64. Using random assignment by lottery as an instrument for insurance coverage, the study found that insurance lead to increased healthcare utilisation and modest improvements in self-reported health status (Finkelstein et al., 2012). Neither the RAND HIE nor the Oregon HIE found any significant effect of insurance coverage on mortality.

We could take these findings to mean that we should not expect to find any effect of spending on mortality, but I can think of two problems with such an interpretation. First, these studies may have been way too small to detect a significant difference from a fairly low baseline mortality rate (Chandra et al., 2021). Both the RAND HIE and the Oregon HIE did observe a lower mortality among those (fully) insured. Second, the care that people choose to (or are forced to) forgo due to having to pay for it themselves may be different from that which people are forced to forgo in a publicly funded health system such as the one in Sweden. Less spending in Sweden could imply fewer nurses and physicians, fewer staffed hospital beds, longer waiting lists and certain therapies or drugs not offered to anyone that seeks care. However, randomised trials are not the only source of experimental evidence in a broader sense. For example, Miller et al. (2021) study the effect of states expanding their public health insurance for low-income adults in 2014 on mortality in a sample of 566,000 individuals aged 55 to 64 that were likely to be affected by the expansion. By comparing individuals in expansion and non-expansions states before and after expansion, they find that expansion-state individuals were more likely to be insured and had a lower mortality after expansion, equivalent to a 0.35 percentage-point reduction per year of insurance coverage.

Although this study is not actually randomised, it scores over the more pedestrian work on the marginal cost of producing a QALY in two important respects. First, it identifies an explicit source of variation in insurance coverage that is plausibly as good as random. Expansion states could be different from non-expansion states, but it is at least prima facie plausible that insurance coverage and mortality in expansion states would have evolved in the same way as in non-expansion states, had they not expanded their insurance schemes. Second, and more importantly, it assesses threats to the validity of this design. Part of this is adjusting for covariates, but

crucially, this is done to confirm that results are insensitive to adjustment (I could have cast my use of variation in nurse labour supply in terms of a quasi-experiment, but from such a perspective my results would have revealed it to be a failed one). Additionally, the study assesses validity in several other ways. For example, using another sample of individuals aged 65 and above who should be unaffected by expansion, it shows that there was no significant difference in mortality or insurance-coverage trends between expansion and non-expansion states in these individuals. This provides support for the assumption that trends in non-expansion states are valid counterfactuals for trends in expansion states in the absence of expansion.

Vanness et al. (2021) use evidence from a similar study on health insurance expansion in the United States (Sommers, 2017) to simulate that a $10 million increase in health spending would result in a loss of 96 QALYs from individuals dropping their insurance due to higher premiums. This implies one QALY lost from covering $104,000 of spending on new interventions. I imagine that this estimate of opportunity cost may be perceived as more believable, by virtue of being based on better evidence on the causal link between spending and mortality. Even if evidence from the United States is unlikely to generalise to other countries due to the idiosyncrasies of its health system (the main benefit of being insured may well be avoiding medical debt; Angrist and Pischke (2015)), we can still learn from U.S. studies on health insurance.

The main problem with empirical work to date on the marginal cost of producing a QALY could be described as a lack of credible research designs (Angrist and Pischke, 2010). For example, it is hard to provide a convincing explanation of why the proportion of the population providing unpaid care would only affect mortality via health spending (Claxton et al., 2015a), why it would have opposite effects on spending in England and Australia[7] (Edney et al., 2018), or why the number of graduated nurses would reflect exogenous variation in labour supply only when adjusting specifically for population density, education level, and the presence of a teaching hospital. To produce credible evidence on the opportunity cost of healthcare resource use, we should look to quasi-experimental studies for inspiration and try to identify sources of as good as random variation in resource use with high face validity. Further, we should strive to, as far as possible, support claims about variation in resource use being as good as random. Had I been able to show that there were no differences between regions in which many and few nurses graduate except in terms of mortality and how many nurses they employed, my results would have been much more convincing.

[7] Edney et al. (2018) do provide an explanation and a detailed report of threats to the validity of their instrument. It is unconvincing in the sense that the provision of unpaid care is not an obvious or an unambiguous cause of increased healthcare resource use.

That being said, weak evidence may be better than no evidence, but our scientific inquiry should not stop here. With the purpose of informing policy and decision making it seems desirable to generate results that would rank higher in a hierarchy of evidence.

In next section, I present a second empirical analysis of my own. I shall not pretend that this analysis has a strong, ex-ante case for as good as random assignment on a par with a quasi-experiment, but although it does not provide clean identification, it takes threats to validity and coefficient stability (Altonji et al., 2005, Oster, 2019) more seriously than prior studies.

More evidence from Sweden

Sweden has one of the lowest numbers of hospital beds per capita in the OECD (2.07 beds per 1,000 people in 2019) and has made large bed reductions in recent decades. Because bed shortages are being discussed as a major problem in Swedish health care (e.g., Vårdfokus, 2021), it could be relevant to consider the provision of hospital beds as an alternative use of resources when funding new, cost-increasing interventions. Here, I attempt to estimate the mortality effects of reductions in hospital bed capacity, or conversely, the benefits of providing more beds, using a panel on standardised mortality (deaths per 1,000 people) and hospital bed capacity (staffed beds per 1,000 people) for all 21 regions between 2001 and 2019.

This study takes its starting point in the observation that regions that made smaller reductions in hospital beds between 2001 and 2019, saw larger decreases in mortality during this period on average. Figure 11 illustrates this association, including an OLS estimate of 0.38 (s.e. = 0.21) fewer deaths per bed. Although only marginally significant (which may be what is reasonable to hope for with 21 observations), this provides a preliminary reason to suspect that bed reductions have led to higher mortality. Because the association is between within-region changes in mortality and bed capacity, it cannot be explained by unobserved differences *between* regions (as in Paper I). Still, there may be changes in unobserved determinants of mortality that coincide with or cause changes in bed capacity. The potential sources of confounding are similar to those discussed previously, with the difference that the provision of beds is one of many uses of healthcare resources. By omitting the provision of other care, we might overstate the effect of bed capacity on mortality if changes in the resource constraints of one region allows it to provide both more beds and more of other healthcare services. Assuming fixed resource constraints, we would instead expect more beds to go along with less of other care, which could lead us to underestimate the effect of providing more beds.

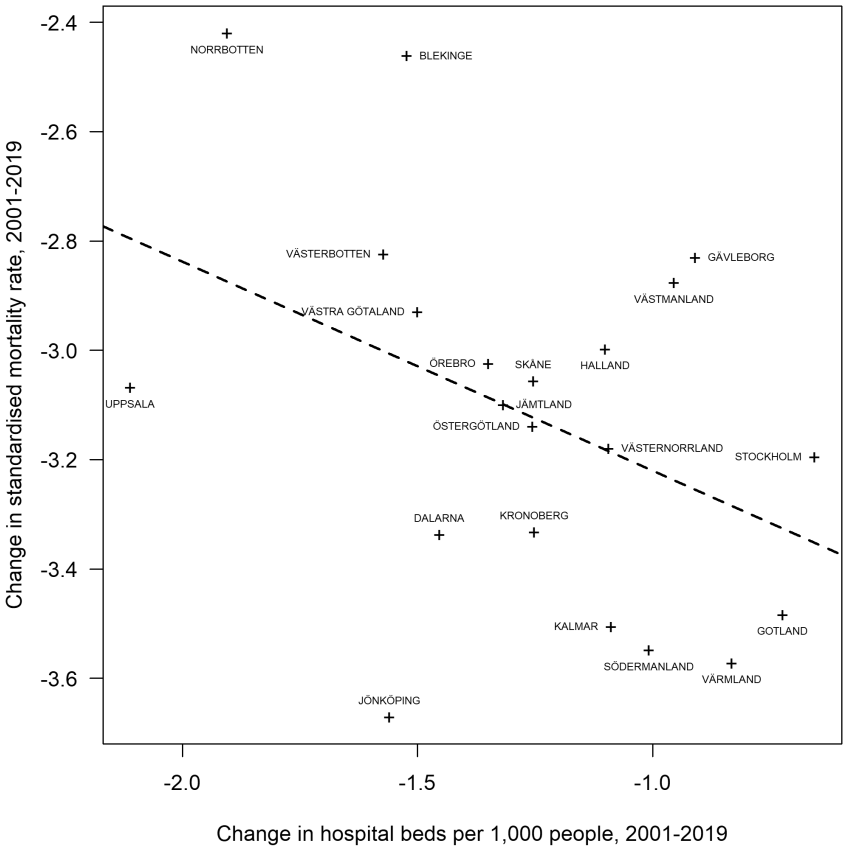


Figure 11 Association between change in standardised mortality (deaths per 1,000 people) and change in hospital beds per 1,000 people, 21 Swedish regions 2001-2019. The dashed line is the conditional expectation estimated by OLS, equivalent to 0.38 deaths averted per additional bed

The purpose of the first-difference estimate in Figure 11 is only to provide an accessible illustration of what kind of variation I draw on in my analysis. For estimation, I exploit this variation using OLS with TWFE. Table 2 reports the main results from this analysis. I find that one more bed is associated with 0.29 (s.e. = 0.08) fewer deaths per year. This coefficient is stable to covariate adjustment, which rules out (or at least makes less likely) several alternative explanations for why we would observe a negative association between bed capacity and mortality. These alternative explanations include changes in income and the socioeconomic or demographic

composition of the population, which could both reduce healthcare need and increase the resources available to a region, and the provision of other care as measured by nurses, physicians, and spending per capita in areas outside of inpatient care. To the extent that the estimate is sensitive to covariate adjustment, the results indicate that 0.29 fewer deaths per bed may understate the effect of bed capacity on mortality, assuming selection on observed factors is informative about selection on unobserved factors (Altonji et al., 2005, Oster, 2019). Adjusting for all covariates, I find that one more bed is associated with 0.40 (s.e. = 0.10) fewer deaths per year. Because data are not available on all covariates for the full sample, I also use a subsample on 20 regions between 2005 and 2019, which yields similar results in the regressions that are possible to estimate with both samples.

Table 2 Regressions of standardised mortality (deaths per 1,000 people) on bed capacity (hospital beds per 1,000 people), adjusting for different sets of covariates. Results are reported for the full sample (21 regions, 2001-2019) and a subsample (excluding 2001-2004 and the region Gotland) based on data availability

	Full sample, 2001-2019 n = 399			*Subsample, 2005-2019 n = 300*		
	coef.	s.e.	R^2	coef.	s.e.	R^2
No controls	−0.29	(0.08)	0.03	−0.30	(0.08)	0.04
Age and sex	−0.32	(0.08)	0.06	−0.36	(0.06)	0.07
Income	−0.29	(0.08)	0.03	−0.30	(0.07)	0.04
Lagged income	−0.29	(0.08)	0.04	−0.33	(0.07)	0.05
Education	−0.30	(0.10)	0.05	−0.29	(0.08)	0.06
Other socioeconomics	−0.30	(0.08)	0.04	−0.31	(0.07)	0.05
All socioeconomics	−0.35	(0.10)	0.07	−0.35	(0.08)	0.09
Age, sex, and socioeconomics	−0.40	(0.11)	0.11	−0.39	(0.08)	0.12
Outpatient nurses and physicians				−0.31	(0.07)	0.07
Outpatient spending				−0.31	(0.11)	0.06
Day surgeries				−0.29	(0.08)	0.04
All outpatient controls				−0.30	(0.08)	0.09
Age, sex, socioeconomics and outpatient				−0.40	(0.10)	0.16
Inpatient nurses, physicians, and spending				−0.31	(0.08)	0.06
All controls				−0.40	(0.10)	0.17

Notes: All regressions include fixed effects for regions and years. coef. is the coefficient on bed capacity. Robust standard errors within parentheses, clustered on region. R^2 is calculated after residualising with respect to the fixed effects. See Paper IV, Table 1 for a full list of covariates.

Table 3 presents the results from additional robustness checks. This includes investigations of a log-log or quadratic specification, the use of more detailed covariates, and population weights, neither of which has any sizeable impact on my main results. I also find that a regression of crude (non-

standardised) mortality on bed capacity with population weights, which can be thought of as mimicking the underlying individual-level data (Angrist and Pischke, 2009), produces comparable but slightly larger estimates (-0.39, s.e. = 0.19 without covariates; -0.47, s.e. = 0.11 with all covariates).

Table 3 Alternate specifications of regressions reported in Table 2

	Full sample, 2001-2019 *n = 399, 147*, 126†*			*Subsample, 2005-2019* *n = 300, 120*, 100†*		
	coef.	s.e.	R^2	coef.	s.e.	R^2
Log-log specification						
No controls	−0.07	(0.02)	0.02	−0.06	(0.02)	0.02
Age, sex, and socioeconomics	−0.10	(0.03)	0.12	−0.09	(0.02)	0.12
All controls				−0.08	(0.03)	0.16
Quadratic specification, linear term						
No controls	−0.27	(0.09)	0.03	−0.24	(0.10)	0.04
Age, sex, and socioeconomics	−0.40	(0.11)	0.11	−0.34	(0.08)	0.12
All controls				−0.33	(0.12)	0.17
Quadratic specification, quadratic term						
No controls	−0.03	(0.05)	0.03	−0.11	(0.08)	0.04
Age, sex, and socioeconomics	0.01	(0.07)	0.11	−0.09	(0.14)	0.12
All controls				−0.12	(0.12)	0.17
Age-sex-specific covariates						
No controls	−0.29	(0.08)	0.03	−0.30	(0.08)	0.04
Age, sex, and socioeconomics	−0.30	(0.10)	0.19	−0.38	(0.12)	0.19
All controls				−0.46	(0.13)	0.26
Population weights						
No controls	−0.24	(0.07)	0.03	−0.30	(0.08)	0.05
Age, sex, and socioeconomics	−0.33	(0.09)	0.14	−0.40	(0.07)	0.17
All controls				−0.50	(0.11)	0.23
Population weights, crude mortality						
No controls	−0.39	(0.19)	0.05	−0.35	(0.17)	0.05
Age, sex, and socioeconomics	−0.33	(0.08)	0.56	−0.38	(0.07)	0.48
All controls				−0.47	(0.11)	0.51
Every third year*						
No controls	−0.33	(0.13)	0.05	−0.42	(0.14)	0.09
Age, sex, and socioeconomics	−0.39	(0.19)	0.27	−0.54	(0.19)	0.33
All controls				−0.59	(0.23)	0.48
Every third year, differences†						
No controls	−0.46	(0.24)	0.05	−0.35	(0.19)	0.03
Age, sex, and socioeconomics	−0.59	(0.23)	0.21	−0.58	(0.26)	0.21
All controls				−0.65	(0.30)	0.39
Somatic beds						
No controls	−0.29	(0.09)	0.05	−0.30	(0.10)	0.04
Age, sex, and socioeconomics	−0.40	(0.12)	0.12	−0.40	(0.09)	0.12
All controls				−0.40	(0.11)	0.17
Psychiatric beds						
No controls	−0.56	(0.30)	0.05	−0.34	(0.37)	0.04
Age, sex, and socioeconomics	−0.68	(0.25)	0.12	−0.27	(0.39)	0.12
All controls				−0.18	(0.40)	0.17

Notes: * and † indicate analyses using data for every third year (in the subsample, observations for 2005, 2007, 2010, 2013, 2016, and 2019). See Table 2 for further notes.

Finally, I am able to demonstrate that higher bed capacity is associated with more bed days (144 days per bed, s.e. = 42) and lower bed occupancy rates (-14.6 percentage points per bed per 1,000 people, s.e. = 3.7), which provides some insight on how bed capacity might have an effect on mortality (see supplement, Paper IV).

To translate these results to the number of life years and QALYs gained per bed, I define two subgroups and estimate the same regressions for the population aged 0 to 64 and the population aged 65 and above. I then multiply the mortality reduction for each age group by its average discounted (quality-adjusted) life-expectancy in 2019 and weight the estimates by the groups' population shares to produce estimates of deaths averted, life years gained, and QALYs gained for the whole population. I find a strong association in the older population (1.16 deaths averted per bed, s.e. = 0.35) and an essentially zero association in the younger population (0.06 deaths averted per bed, s.e. = 0.05), which translates to 2.9 QALYs gained per bed (s.e. =0.95) or 3.13 QALYs gained per bed (s.e. = 1.08) when adjusting for covariates. As an alternative approach, I use the elasticity of mortality with respect to bed capacity from my log-log specification and apply the effect of a permanent marginal increase in beds on age-sex-specific mortality rates in 2019. I extrapolate these shifts in mortality curves to derive an increase in the implied quality-adjusted life-expectancy of the population, equivalent to 2.66 QALYs gained per bed, or 3.22 QALYs gained per bed when adjusting for covariates.

Compared to the results in Paper I, these estimates can more credibly be given a causal interpretation, because they are largely insensitive to adjustment for potential confounders. Instead of relying on an arbitrary set of covariates to produce a final result, I here use covariates as a means to assess potential threats to validity. Had I found that covariate adjustment had a large impact on my results, I would have been in the same situation as before in terms of picking one specification "to portray as a rose", but from the finding that none of the factors that I adjust for matter, we can be more confident in the assumption that there is no confounding from unobserved factors. To the limited extent that it does matter, the results suggest that my estimates may understate the effectiveness of bed provision in reducing mortality. In this respect, this analysis is in the same vein as a quasi-experiment. I assume that variation in bed capacity with TWFE is as good as random, and report analyses to support this assumption. The crucial difference, and main limitation of this study compared to quasi-experimental work, is that I do not provide an explanation for the source of variation in bed capacity.[8]

[8] I explore the role of nursing staff shortages (The National Board of Health and Welfare, 2018) using graduated nurses per capita as an instrument for bed capacity, but am unable to provide compelling

Table 4 Regressions of standardised mortality (deaths per 1,000 people) on bed capacity (hospital beds per 1,000 people) by age group, with translation of effects to life years and quality-adjusted life years (QALYs) gained per bed.

	Age 0-64		*Age 65+*		*Overall*	
	coef.	s.e.	coef.	s.e.	coef.	s.e.
Full sample, no controls						
Deaths	−0.06	(0.05)	−1.16	(0.35)	−0.28	(0.08)
Life years gained	1.60	(1.16)	12.47	(3.78)	3.77	(1.20)
QALYs gained	1.32	(0.96)	9.20	(2.79)	2.90	(0.95)
Full sample, all controls						
Deaths	−0.07	(0.05)	−1.68	(0.43)	−0.39	(0.11)
Life years gained	1.66	(1.27)	17.97	(4.56)	4.92	(1.69)
QALYs gained	1.37	(1.05)	13.26	(3.37)	3.75	(1.33)
Subsample, all controls						
Deaths	−0.01	(0.05)	−1.90	(0.49)	−0.39	(0.10)
Life years gained	0.18	(1.23)	20.41	(5.22)	4.22	(1.39)
QALYs gained	0.15	(1.01)	15.05	(3.85)	3.13	(1.08)

Notes: Life years and QALYs gained are calculated by multiplying deaths averted by average remaining (quality-adjusted) life expectancy, discounted at a 3% rate. 25.4 years/20.1 QALYs for ages 0-64; 10.7 years/7.9 QALYs for ages 65 and above. Overall figures are calculated by weighting age-specific effects 80:20. Overall standard errors are derived by estimating the equations jointly. See Table 2 for further notes.

There is also a drawback of the analysis presented here, in that it does not directly address the link between healthcare spending and mortality. However, making the leap of faith that more beds could be provided with more staff (one nurse, one assistant nurse, and one physician to staff six beds) we can derive a cost per QALY gained of SEK 420,000 (see supplement, Paper IV). This seems a reasonable assumption, considering that Sweden has made large absolute reductions in bed capacity since 2001, which implies small capital requirements and physical space constraints to making at least marginal increases in the number of staffed beds. However, it might understate the cost by omitting the resources used for procedures and medication that would potentially go along with more (longer) hospitalisations. A further source of underestimation in the cost per QALY is that the application of quality-of-life norms and life expectancies for the general population to deaths averted could overstate the number of QALYs gained (Claxton et al., 2015a). It should be noted that the estimate might also

evidence on this link. Bed capacity is positively associated with graduated nurses (0.36 beds per graduated nurse, s.e. = 0.2) with marginal significance and leads to large but imprecise estimates of deaths averted per bed when used as an instrument (see supplement, Paper IV).

overstate the cost per QALY gained by including only effects on mortality, thus assuming that higher bed capacity provides no improvements in quality of life for patients.

I would like to share one final thought on this study in the context of estimating the marginal cost of producing a QALY. I find it hard to trust a regression of mortality on spending when it is not supported by auxiliary analyses demonstrating that spending goes towards the provision of some form of care. In this respect, QALYs gained per hospital bed seems a more credible piece of evidence on the value of alternative use of healthcare resources. If new interventions are financed by diluting rather than denying care, then a creeping reduction in hospital bed capacity is precisely the sort of politically expedient form of financing we would expect. The provision of hospital beds could therefore serve as a reasonable expectation on alternative use funding, even without evidence to support that more spending leads to more beds. The results of this study can also be seen as a piece of a larger puzzle in determining a credible measure of opportunity cost, in which evidence on the cost-effectiveness of other specific uses of resources may provide further pieces.

IV. ROLE IN PRIORITY SETTING

Cost-effectiveness evidence is typically reported as the monetary cost per QALY gained. As we saw in Chapter II, the theoretical role of such a ratio is that it can be used to maximise the number of QALYs produced subject to a budget constraint when choosing from a list of potential investments. Alternatively, and more realistically when decisions are made on a one-by-one basis, it can be judged against a cost-effectiveness threshold. To maximise QALYs, this threshold should represent the opportunity cost (i.e., QALYs forgone) associated with the monetary cost. Although health maximisation is often held up as the theoretical objective of cost-effectiveness analysis, explicit considerations of opportunity cost have been rare (Gafni and Birch, 2006). This is not surprising in the absence of evidence which would allow us to be explicit. In this chapter, I discuss how this might change with the emergence of empirical evidence on the (health-)value of alternative uses of healthcare resources. As I have indicated, more work is needed to provide credible estimates of the number QALYs that we can expect to forgo when funding new interventions, bur for the sake of presentation, I shall use SEK 180,000 per QALY for this quantity.

Equity-efficiency trade-offs

In the context of judging the monetary cost per QALY gained against a cost-effectiveness threshold, it seems natural that a measure of opportunity cost should be used to inform the value of the threshold. However, for a threshold to be applied as a strict decision rule we would have to assume that decision makers were concerned only with health maximization. In reality, this objective may not align with social values or the remit of healthcare decision makers (Coast, 2009, Cleemput et al., 2011). The overarching principles for priority setting in Swedish healthcare are that (1) care should be provided without discriminating based on a patient's personal characteristics or role in society, (2) resources should be allocated based on needs, and (3) there should be a reasonable relation between costs and effects, measured in terms of improved health and quality of life, when choosing between different services or interventions (SFS 2017:30, Prop. 1996/97:60). To discuss the role of cost-effectiveness evidence in setting priorities according to these principles, I focus on drug reimbursement decisions made by The Dental and Pharmaceutical Benefits Agency (TLV), for which the use of cost-effectiveness evidence is relatively formalised.

In Sweden, prescription drugs are subsidised through The Pharmaceutical Benefit and TLV is tasked with deciding which drugs should be included in it (SFS 2002:160). Striving for a balance between needs, non-discrimination, and cost-effectiveness, TLV has over time established a practice in which decisions are primarily based on an assessment of cost per QALY gained, disease severity, and rarity. TLV will generally accept a cost per QALY below SEK 1 million for drugs targeting very severe diseases, and may accept a cost of up to SEK 2 million per QALY for conditions that are both rare and very severe (TLV, 2016b, TLV, 2016a). Paying more per QALY gained in severe disease is motivated by allocating resources based on needs, and the premium on rarity by the notion that orphan drugs are inherently more expensive to develop per patient treated and that the size of a patient group is a grounds of discrimination (TLV, 2016b). In its guidelines for economic evaluation (TLV, 2017), TLV recommends that all relevant benefits and costs be included, regardless of to whom they accrue. However, there has been some recent changes to these recommendations in terms of what is deemed relevant. The guidelines used to recommend that non-medical consumption be included as a cost of added life years (TLV, 2003), but no longer do (c.f., TLV, 2015). Since 2018, TLV also interprets that accounting for production in added life years discriminates against patients who are not part of the labour force (TLV, 2018b).

Here, I characterise the objective of health maximisation as an efficiency criterion and TLV's approach to decision making as an equity-efficiency trade-off (Wagstaff, 1991). To illustrate how cost-effectiveness evidence can be used to inform equity-efficiency trade-offs, I use a relatively recent pharmaceutical reimbursement decision as a case. In 2018, TLV made a decision on a new drug, Orkambi, for the treatment of cystic fibrosis. This drug was expected to affect n = 245 patients with an incremental gain of 1.65 QALYs per patient ($\Delta Q_i/n$) and an incremental healthcare cost of 2.54 million per patient ($\Delta C_h/n$) compared to standard of care (TLV, 2018c), which resulted in a cost per QALY gained of

$$\frac{\Delta C_h}{\Delta Q_i} = \text{SEK 1.54 million per QALY.}$$

TLV decided to accept this cost and reimburse Orkambi with reference to the severity and rarity of cystic fibrosis (TLV, 2018a). Had the information that other patients could be expected to forgo one QALY per SEK 180,000 spent on Orkambi been available to TLV, they would have known that Orkambi was cost ineffective by comparing SEK 1.54 million per QALY to this "threshold". However, I will venture a guess that this was never in doubt. The question was, rather, what degree of cost-ineffectiveness was acceptable with respect to the equity arguments for reimbursement. If we let k

denote the marginal cost per QALY for the healthcare services that are forgone, we can quantify the degree of cost-ineffectiveness as

$$\text{NHB} = \Delta Q_i - \frac{\Delta C_h}{k} = \Delta Q_i - \Delta Q_j = -3{,}053 \text{ QALYs},$$

which is the expected net health loss from reimbursement where ΔQ_j denotes the health forgone by other patients. However, it is not immediately clear how this quantity could have informed the trade-off between equity and efficiency. Instead, it would have been more informative to translate the numerator in the ICER to an opportunity cost

$$\frac{\Delta C_h/k}{\Delta Q_i} = \frac{\Delta Q_j}{\Delta Q_i} = 8.6 \text{ QALYs forgone per QALY gained.}$$

This would have framed TLVs decision as a judgement on the relative importance of QALYs forgone and QALYs gained. If TLV's sole objective were health maximisation, they would have no reason to consider anything other than the NHB, but precisely because it is not, QALYs forgone per QALY gained could be a more useful metric. We can understand the role of this metric in terms of equity weighting (Round and Paulden, 2018), if we assume that TLV's objective is to maximise a net equity-weighted health benefit (NWHB)

$$\text{NWHB} = w_{ij}\Delta Q_i - \Delta Q_j,$$

where w_{ij} is the equity weight assigned to health gained (by patient group i) relative to health forgone (by [the unidentified] patient group j). TLV would decide to reimburse a new drug if

$$w_{ij}\Delta Q_i > \Delta Q_j,$$

since this would imply a positive NWHB. Equivalently, this condition could be stated as

$$\frac{\Delta Q_j}{\Delta Q_i} < w_{ij}.$$

That is, if the number of QALYs forgone per QALY gained was less than the equity weight assigned to QALYs gained. Because QALYs forgone are calculated as $\Delta Q_j = \Delta C_h/k$, we can also restate this rule as

$$\frac{\Delta C_h}{\Delta Q_i} < kw_{ij}$$

This shows that the decision rule for maximising NWHB is essentially no different from TLV's current practice of judging the monetary cost per QALY gained against an equity-adjusted monetary cost-per-QALY threshold (c.f., Bobinac et al., 2012). Being able to quantify health forgone would not have magically solved the problem of deciding whether SEK 1.54 million per QALY gained was an acceptable cost. The main advantage would have been to frame the decision as a trade-off between health gained and health forgone and make a judgement on whether the health gained by Orkambi patients was at least 8.6 times as valuable as the health forgone by other patients.

Technically, this approach is just comparing another kind of ratio to another kind of threshold. However, it seems preferable because it separates the efficiency criterion from the equity criteria and makes the trade-off more transparent. Incidentally, it also highlights a soft benefit of ratio metrics over net benefits (O'Mahony, 2020). It is unlikely that TLV would be able to define equity weights for different disease characteristics or patient characteristics in advance to explicitly calculate an NWHB, but by making decisions on QALYs forgone per QALY gained, equity weights could be established through practice; potentially with input from research eliciting social values on health-versus-health trade-offs (Lindholm et al., 1998).

If we transform TLV's SEK 1 million and SEK 2 million per QALY thresholds into equity-weight thresholds, the implication is that they are willing to sacrifice about 5.5 unidentified QALYs to gain one QALY for a patient with a very severe disease or about 11 QALYs if the disease is also rare. However, because these thresholds have evolved without information on opportunity cost, they may not reflect trade-offs that TLV would be comfortable making. In fact, there are two reasons to suspect that the thresholds would have been lower, had they instead evolved through decisions based on the number of QALYs forgone per QALY gained. First, there is a tendency to neglect opportunity cost when it is not made explicit, both in private (Frederick et al., 2009) and public decision making (Persson and Tinghög, 2020). The monetary cost per QALY gained is not explicit about the alternative use of resources and may increase the likelihood of opportunity cost neglect in healthcare priority setting, where those who stand to benefit from a new treatment are identified but those who bear the opportunity cost are not. Second, denying treatment with reference to the monetary cost per QALY being too high puts decision makers in the unenviable position of having to defend why they are not willing to pay more to save someone's life. This could add pressure to ignore the opportunity cost of reimbursement. Being explicit about opportunity cost is unlikely to make anyone happy about treatment being denied, but it might make such a decision, when warranted, easier to defend or explain.

Increased transparency about trade-offs, making opportunity costs more salient to decision makers, and making the reasons for negative decisions easier to communicate, all speak in favour of using evidence on the opportunity cost of healthcare resource use to explicitly report cost-effectiveness evidence as health gained and health forgone, instead of letting it inform a monetary cost-per-QALY threshold. Additionally, there is one more important reason which has been obscured by my choice of example. Healthcare interventions use a variety of resources and resources might not mean money. When an intervention is a patented drug, the money that the healthcare sector must pay is a resource with an alternative use, but changes in the utilisation of other care that results from drug treatment may, e.g., consume or free-up the time of healthcare personnel. Although health care pays for this time in money, the resource that is directly affected is the time itself and this might have a different expected alternative use than the funding for the drug. As we get more evidence on the QALY-value of alternative uses of resources, it seems to me that we should like to attach a shadow price to each resource rather than let them take a detour via money. This would more accurately reflect how many QALYs we could expect to forgo.

If the uses of different resources have different opportunity costs, a most likely implication is also that the patients who forgo health differ depending on which resources are used to provide a treatment. This implies, in turn, that it would be necessary to have information on the equity-bearing characteristics of the patients whose health is forgone to make a fully informed trade-off between equity and efficiency (Bobinac et al., 2012, Sculpher et al., 2017, Round and Paulden, 2018). In particular, it poses some problems for the approach of letting equity weights evolve through practice, for though it could be meaningful to make a trade-off between QALYs gained and unidentified QALYs forgone, we would have to assume that the equity-bearing characteristics of patients forgoing health did not change between decisions for the implied equity weights to be comparable (Paulden and McCabe, 2021). Estimating the marginal health benefits of specific uses of resources appears like a starting point for addressing this issue. For example, from my analysis of hospital bed reductions, we find that mortality effects are almost exclusive to the population above 65, which says at least something about who forgoes health.[9]

The lack of a ready solution to this problem should not obscure my key message. If the goal of economic evaluation is to support equity-efficiency trade-offs and efficiency is defined in terms of health maximisation, then cost-effectiveness evidence should not be reported as a monetary cost per

[9] Claxton et al. (2015b) take us further towards a solution by providing an estimate of the average absolute QALY shortfall for health forgone from marginal budget decreases in the English NHS.

QALY gained. When we are able to quantify opportunity costs, it is both natural and helpful to report them.

A societal perspective and non-health benefits

The main drawback of efficiency as health maximisations is that it requires us to omit benefits and costs that are not measurable in health. Or more accurately, such effects are not benefits and costs from the perspective of health maximisation. Here, I briefly consider how we could account for non-health effects in the approach to economic evaluation that I have discussed this far. Let us assume that we can measure the WTP for all these effects and let ΔB_c denote the WTP for non-health benefits and ΔC_c the WTP for non-health costs associated with an intervention in relation to a comparator. We can then summarise all non-health effects as a net non-health benefit (NCB)[10]

$$\text{NCB} = \Delta B_c - \Delta C_c.$$

With the NCB, we can at very least answer whether considering non-health benefits and costs would make it more or less desirable to fund the intervention, i.e., is the NCB positive or negative? To go beyond this simple inspection, we would require some means to render the NCB commensurable with QALYs. Having assumed that the NCB is a monetary value, the WTP for a QALY (v) – which has been estimated at about SEK 3 million in Sweden (Olofsson et al., 2019) – seems suitable. Let us consider the case of Orkambi again and focus on the productivity gains from reduced absence from work and increased survival. The productivity gains associated with Orkambi were not part of TLV's base case and the estimates from the company's analysis were not made publicly available. For the sake of illustration, I approximate the non-health benefit as SEK 550,000 per patient treated, of which SEK 380,000 are due to increased survival [11]

$$\Delta B_c = \text{SEK } 550{,}000 \times 245 = \text{SEK } 135 \text{ million}.$$

Assuming the QALYs forgone by funding Orkambi instead of other healthcare services ($\Delta Q_j = 3{,}457$) were associated with the same

[10] NCB for net consumption benefit, but non-health benefits need not be consumption.

[11] From TLV (2018c), we know that the actual figure was based on an employment rate of 0.65 for working age patients and 124 hours of work lost for every event of pulmonary exacerbation. My approximation is based on a patient with an average monthly (157 hours) wage of SEK 34,600 (Statistics Sweden), an assumption of 50% social fees paid by the employer, a 0.65 probability of being employed, and 0.48 events averted every year due to treatment (48 week rates were 1.14 in the control group and 0.70 in the treatment group) over the average treatment length (17 years) at 3% discount rate (TLV, 2018c), which has a present value of SEK 168,384. Additionally, assuming 1.7 life years gained per patient with the same employment rate and 5% absence from work, survival production is SEK 384,502 per patient.

productivity gain per QALY (SEK 550,000/1.65=SEK 333,333), reimbursement also meant forgoing

$$\Delta C_c = \frac{\Delta C_h}{k} \times \text{SEK } 333{,}333 = \text{SEK } 1.152 \text{ billion},$$

which would make the NCB = -SEK 1.017 billion. To consider the NCB jointly with QALYs gained and QALYs forgone, we could calculate its QALY-equivalent ($\text{NCB}/v = -339$ QALYs) and add it to the NHB

$$\text{NHB} + \frac{\text{NCB}}{v} = -3{,}392 \text{ QALYs},$$

or subtract it from the numerator in QALYs forgone per QALY gained

$$\frac{\Delta C_h/k \; - \text{NCB}/v}{\Delta Q_i} = 9.4 \text{ QALYs forgone per QALY gained.}$$

Although the WTP for a QALY translates non-health effects into QALYs, the theoretical role of these quantities is less obvious. We could motivate their use, either by conceiving of efficiency as maximisation of monetary value but with the expected alternative use of healthcare resources being other healthcare services (i.e., a fixed budget), or as a way of checking whether the local rationality of health maximisation aligns with societal optimality being defined more broadly (Brouwer et al., 2006, Brouwer et al., 2019). However, it should be clear from my example that accounting for non-health effects would tend to make cost-effective interventions more cost effective and cost-ineffective interventions even less cost effective. When we extend the types of benefits we consider it is necessary to also account for how much of these benefits we forgo (Sculpher et al., 2017, Walker et al., 2019), and under the assumption that health gained and health forgone are associated with the same productivity this result follows. Therefore, focusing strictly on health effects may be less of a problem than I indicated before. This could change when non-health benefits are extraordinary in some sense and particularly with information on who forgoes health, which makes it possible to judge what is and what is not extraordinary.

Sensitivity analysis

I have argued that QALYs forgone per QALY gained appears a more relevant metric than the NHB for informing an equity-efficiency trade-off. However, QALYs forgone is a form of ICER and is therefore associated with the same disadvantages in terms of its interpretation when performing sensitivity analyses or when considering mutually exclusive strategies

(Stinnett and Mullahy, 1998, Paulden, 2020). Here, I begin by extending Paulden (2020)'s framework for illustrating NHB in the cost-effectiveness plane to include equity weights.

Figure 12 illustrates six out of nine scenarios considered in TLV's sensitivity analysis of the cost-effectiveness evidence on Orkambi.[12] Figure 12a plots the incremental cost and incremental QALY gain in each scenario relative to standard of care. The slope of the dashed line equals the ICER of SEK 1.54 million per QALY and the dotted lines are the minimum and maximum ICERs in the different scenarios (B = SEK 1.41 million per QALY, E = SEK 1.87 million per QALY). Figure 12b translates the monetary costs to QALY forgone and shows that the number of QALYs forgone per QALY gained are still the same ICERs but expressed with another scale on the y-axis. In Figure 13a, I draw a line with a slope of unity through scenario A and the least and most cost-effective scenarios. This results in an intercept on the y-axis which corresponds to the net health loss (NHL) in these scenarios[13]. We see that the differences in ICERs or QALYs forgone per QALY gained between scenarios cannot be used to consistently say anything about whether the drug is more or less cost-effective in one scenario compared to another. For example, the ICER in scenario I is lower than in scenario A, but scenario I implies the largest NHL. Finally, in Figure 13b, the slope of the lines is equal to 11 QALYs forgone per QALY gained, which is the approximate equity weight implied by TLV's maximum threshold of SEK 2 million per QALY. The lines intersect the x-axis at the NWHB in each scenario. Assuming this indeed represents how TLV would like to make trade-offs between equity and efficiency, we see again that changes in the ICER between scenarios will not necessarily tell us whether the drug should be more or less preferred from an equity perspective. For example, the ICER in scenario F is lower than in scenario A, but with this equity weight it should be slightly less preferred.

[12] See TLV (2018c, Table 17). I have named the scenarios A through I, but omitted scenarios C, G, and H from Figures 12 and 13 to avoid cluttering.

[13] Although not illustrated, we could similarly identify the negative NHB in Figure 13a where the lines intersect the x-axis. NWHB and NWHL can also be identified in Figure 13b on the x-axis and y-axis, respectively, but with the equity weight applied to QALYs forgone in the former case and to QALYs gained in the latter.

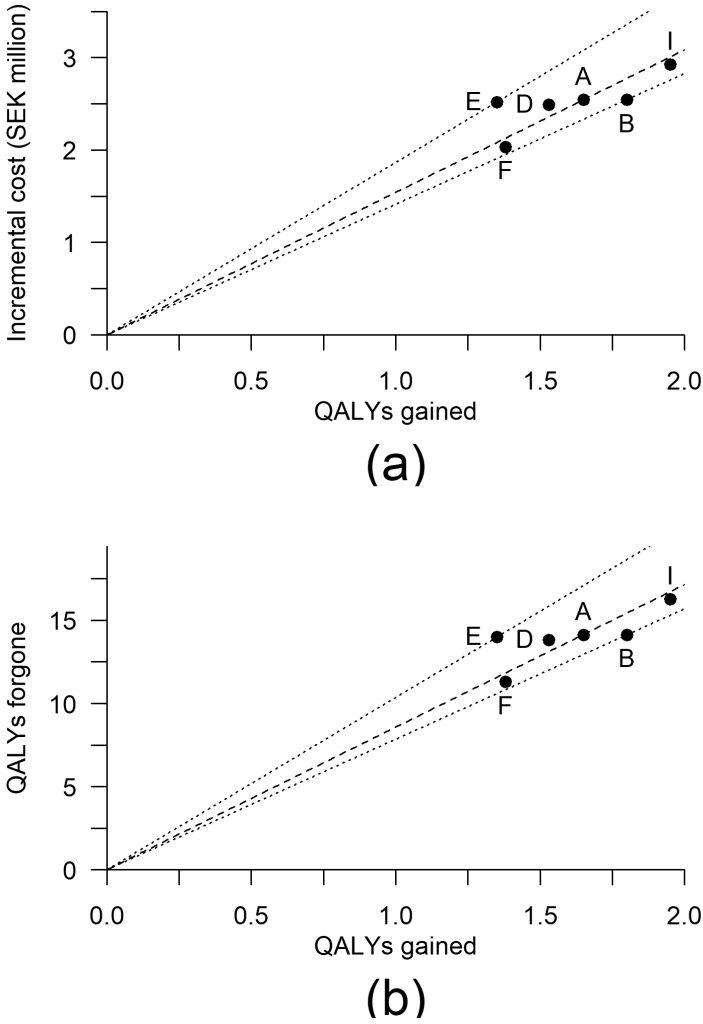


Figure 12 Sensitivity analysis for cost-effectiveness evidence on Orkambi, with six different scenarios. (a) Monetary cost on y-axis. (b) Health cost on y-axis. The slope of the dashed line is the ICER / QALYs forgone per QALY gained in the base case (A). Dotted lines represent minimum and maximum ICER / QALYs forgone per QALY gained across scenarios

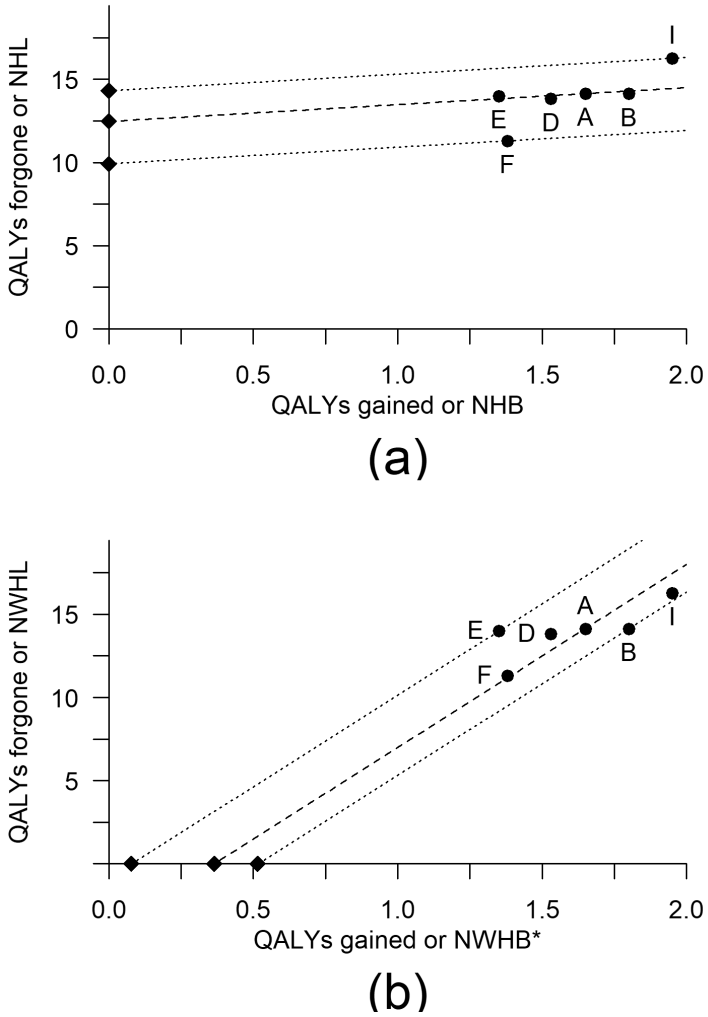


Figure 13 Sensitivity analysis for cost-effectiveness evidence on Orkambi, with six different scenarios. (a) Dashed / dotted lines with a slope of unity drawn through the base case (A) / least and most cost-effective scenarios intersect the y-axis at the net health loss (NHL) or the x-axis at the net health benefit (NHB). (b) Dashed / dotted lines with a slope equal to the equity weight drawn through base case / least and most prioritised scenarios intersect the y-axis at the net equity-weighted health loss (NWHL) or the x-axis at the net equity-weighted health benefit (NWHB*). In NWHB*, QALYs forgone are divided by the equity weight. In NWHL, QALYs gained are multiplied by the equity weight

The substance of this illustration is that sensitivity to changes in modelling assumptions neither in ICERs nor QALYs forgone per QALY gained are interpretable without defining an equity weight and explicitly calculating the NWHB; the difficulty of defining an equity weight in advance is quite easily overcome by doing the calculations over a range of weights. However, the objective of scenario analyses might not be to determine the relative cost-effectiveness or level of priority of different scenarios. If decision makers are purely interested in confirming that the number of QALYs forgone per QALY gained does not exceed a certain threshold (e.g., 11) in any scenario, then this is trivially achieved without further ado.

The drawback of ICERs is more pronounced in the context of probabilistic sensitivity analysis (Stinnett and Mullahy, 1998); here, we have to calculate the NWHB. I will now explore two implications of statistical uncertainty in this type of cost-effectiveness evidence. Partly because large uncertainty in empirical estimates of opportunity cost may seem like an obstacle to setting an explicit policy parameter. We begin by noting that ΔC_h, ΔQ_i, and k^{-1} (marginal productivity of funds in alternative use) are random variables, and that we estimate NWHB as a combination of their estimates

$$\widehat{\mathrm{NWHB}} = w_{ij}\widehat{\Delta Q}_i - \widehat{\Delta C}_h\,\widehat{k^{-1}}.$$

We see that if the estimator for k^{-1} is uncorrelated with the estimator for ΔC_h, then our estimator for NWHB is unbiased, assuming the estimators for ΔC_h, ΔQ_i, and k^{-1} are themselves unbiased.[14] One implication of statistical uncertainty in k^{-1} is therefore that its estimator must be uncorrelated with the estimator for ΔC_h if the estimator for NWHB (or NHB) is to be unbiased. It seems tricky to determine this correlation, because sampling error in the marginal productivity of alternative use is entirely hypothetical. If uncertainty about the cost of an intervention is largely specific to its trial, then it may be reasonable to assume no correlation. It is clear, however, that potential correlation between marginal productivity and model parameters in economic evaluation may impact the interpretation of results (Stinnett and Mullahy, 1998), and future work may want to address the potential source and sign of correlation to determine if this issue could be of practical importance.

The second implication that I explore relates to the familiar point that statistical uncertainty is only relevant to decision making for determining whether it is worthwhile acquiring more information (Claxton, 1999). Again, for the sake of illustration, I will make some arbitrary assumptions about the costs and benefits of Orkambi. I assume that the incremental health gain and cost per patient are normally distributed with standard

[14] This follows from $E\left[w\widehat{\Delta Q} - \widehat{\Delta C}\,\widehat{k^{-1}}\right] = wE\left[\widehat{\Delta Q}\right] - E\left[\widehat{\Delta C}\right]E\left[\widehat{k^{-1}}\right] - Cov\left(\widehat{\Delta C}, \widehat{k^{-1}}\right)$

deviations equal to a quarter of their means. For the marginal productivity of alternative use, I assume that the only source of uncertainty is the estimated elasticity (coef. = 0.076, s.e. = 0.032). Figure 14a shows the resulting distribution of NWHB. With $w_{ij} = 11$, the expected NWHB is positive and there is a 0.71 probability that saying yes is the correct decision. Figure 14b illustrates that this probability exceeds 0.5 at $w_{ij} = 8.1$ and that it is unlikely to be the right call to prioritise this drug if decision makers are only willing to forgo, say, 5 QALYs per QALY gained[15].

Knowing the opportunity cost of resource use with certainty would make us more confident that saying yes was indeed the correct decision. However, just like with uncertainty about resource use and health gains, uncertainty about opportunity cost does not let decision makers avoid making decisions; but they can postpone them. Figure 15 plots the expected value of perfect information (EVPI) on the marginal productivity of alternative use. This shows that at $w_{ij} = 11$, knowing opportunity cost with certainty would be worth about 200 QALYs, which indicates the potential benefit of postponing the decision to collect better information. Unlike uncertainty in (some) model parameters, uncertainty about opportunity cost is not something that drug companies seeking reimbursement can do anything about. The interpretation is, rather, that making close-call decisions with very uncertain information on the opportunity cost of spending comes at a very high opportunity cost in itself if it is possible to become more certain. Therefore, the main implication of statistical uncertainty in opportunity cost estimates is not that decision makers should be cautious about using them, but that it would be very valuable to have estimates that were more precise. Considering that the value of information on this parameter is not restricted to a single decision, but affects all decisions made by a reimbursement authority, it may be hard to understate this value.

[15] There is a difference from 8.6 QALYs forgone per QALY gained reported previously, partly because I use the exact value of k = SEK 183,538, partly because of a slight skew in the distribution of NWHB when k^{-1} is treated as a random variable. Expected NWHB is positive when w exceeds 8.4 QALYs forgone per QALY gained.

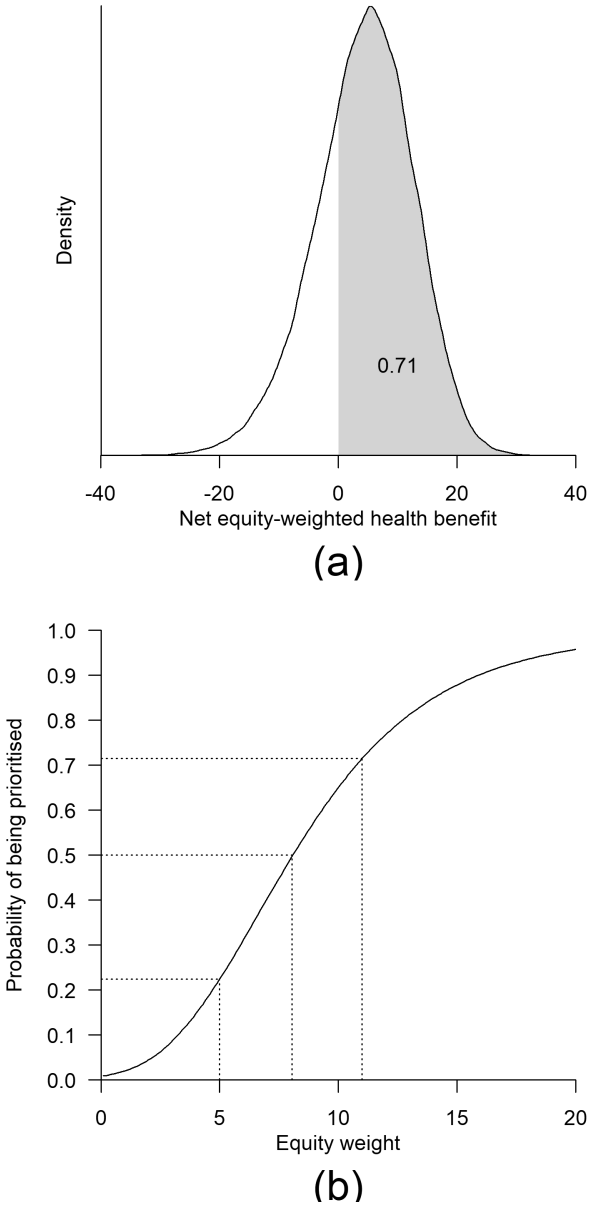


Figure 14 Fictional probabilistic sensitivity analysis of cost-effectiveness evidence on Orkambi. (a) Distribution of net equity-weighted health benefit with equity weight, w = 11 QALYs forgone per QALY gained. (b) Probability of being prioritised with w = 5, 8.1, and 11

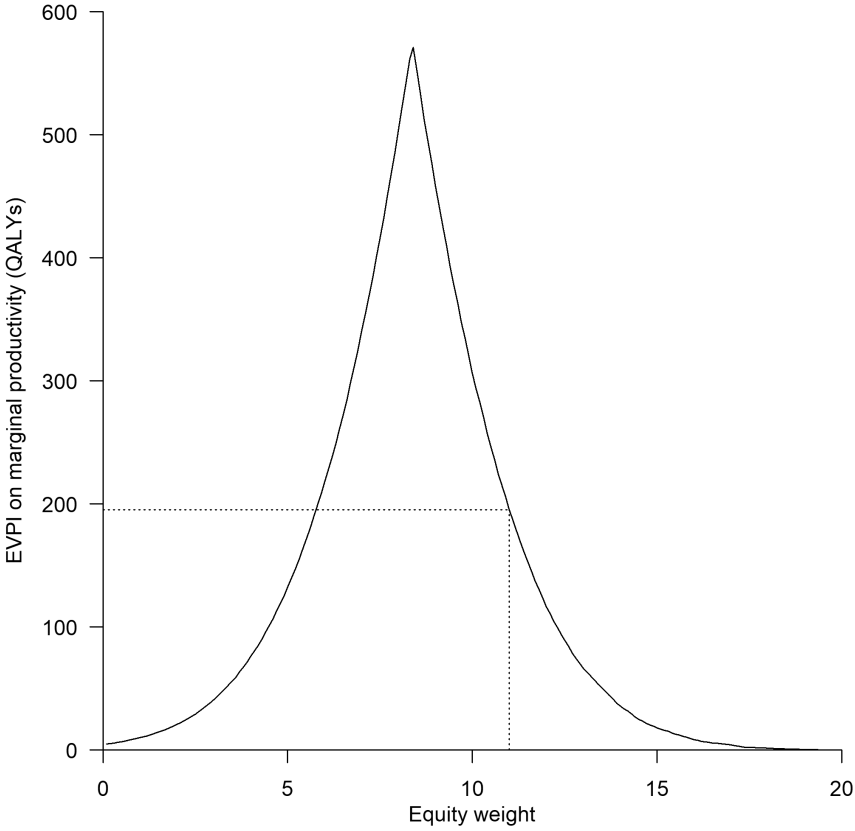


Figure 15 Expected value of perfect information (EVPI) on the marginal productivity of funds in alternative use, in the reimbursement decision on Orkambi

Do high prices imply high priority?

There is a difference between patented drugs and other intervention that bears mentioning. I have, this far, offered no opinion about how many QALYs it would be appropriate to sacrifice to gain 1 prioritised QALY. In part, because I believe that the role of economic evaluation is to describe this trade-off rather than to make it. However, I also believe it is appropriate to scrutinise whether paying more per QALY gained for severe and rare conditions is a sane operationalisation of ethical principles for priority setting in Swedish health care. Setting aside the controversy on rarity as a relevant criterion (Juth, 2017, Sandman and Gustavsson, 2017, Juth et al., 2021) there is the matter of strategic behaviour on the part of pharmaceutical companies (Pekarsky, 2015, Paulden and McCabe, 2021). Accepting higher costs per QALY gained for drugs targeting severe conditions is consistent with the principle of allocating resources according to need under the assumption that it is necessary to accept high prices for patients to gain access to these drugs. This assumption is not necessarily correct. We may be able to settle on a threshold (e.g., 11 QALYs forgone per QALY gained) which determines the maximum price that we are willing to pay, but there is also a minimum price that pharmaceutical companies would be willing to accept. It is only when we raise our maximum acceptable price from below to above a company's minimum acceptable price that we effectively give priority to patients. When we declare that we will accept prices up to SEK 1 million per QALY, we should expect companies to set their prices accordingly. It should be in the interest of all patients, even those who stand to benefit from these drugs, that we pay no more than necessary. In this respect, our current approach to priority setting for prescription drugs is not entirely sane. If high-need patients have unmet care needs that are not drug-related, we could conceivably allocate resources more appropriately according to need by being more restrictive with how much we pay for new drugs.

V. CONCLUSIONS

In this thesis, I have attempted to provide a more robust foundation for measuring the costs of new healthcare interventions as health forgone and for using this form of cost-effectiveness evidence in healthcare priority setting. I have done so by examining the definition of opportunity cost in the context of economic evaluation, by providing empirical estimates of the health that healthcare resources could produce in alternative use, and by illustrating how this evidence could have informed trade-offs between equity and efficiency in a case of pharmaceutical reimbursement.

If we define opportunity cost as value in best alternative use and value is measured in QALYs, the opportunity cost of using a healthcare resource must always be QALYs forgone. However, if healthcare priority setting is not solely about maximising QALYs, value in expected alternative use seems a more appropriate definition to use in economic evaluation. The best evidence is that Swedish health care could produce one additional QALY by spending between SEK 180,000 and SEK 420,000, which means that spending SEK 1 million on a new intervention translates to forgoing between 2.4 and 5.6 QALYs. I have identified and addressed some issues with the credibility of these estimates, but further work remains to be done in this respect. Future work should also explore in which situations, and to what extent, we can expect that a resource *would* have been used to provide other healthcare services. How many QALYs it *could* have produced, had it been devoted to, e.g., provision of more hospital beds may still be informative for priority setting, but less so from the perspective of expected alternative use. The practical relevance of my empirical work is that it allows cost-effectiveness evidence to be reported as QALYs forgone per QALY gained. This is more transparent about trade-offs between equity and efficiency than the monetary cost per QALY and could improve decision making by making opportunity costs more salient to decision makers and the reason for sometimes denying costly treatments easier to communicate.

ACKNOWLEDGEMENTS

Above all, I would like to thank my supervisor, Martin. I was very lucky to end up under his mentorship. He has been a great teacher, but more importantly, a fantastic coach. It's his encouragement, praise, and good advice that has helped me see this through. He's also been a good friend, and we've had a lot of fun (and pints) together over the past few years. By extension, I am grateful to Gustav for introducing us (additionally, I acknowledge that he's a generally fun guy).

I would like to thank Gazi, who first encouraged me to pursue a PhD and Pierre, whom I worked and spent most of my time with during my master's. I would probably not have ended up in research if it weren't for them.

Thanks to Mikael for good suggestions and critique during my half-time seminar. Thanks go also to the people at York who invited me to present my first paper and inspired the direction that I took with this thesis.

I am grateful to Magnus, for sharing his expertise about medicine and health care, and Erik for teaching me about research design and causal inference (sorry for the lack of clean identification).

Additionally, I would like to thank a lot of people that I've had the pleasure to get to know at HSA and NEK. Kasper, whom I have enjoyed working and traveling with, and Martina, Mattias, and Therese who were also around when I first got here and helped shape my CMT experience. David, for fun discussions about econometrics. Lasse, for a constant and friendly presence. Lars-Åke, for throwing good parties. Liam, for tolerating that the desk next to his is very, very untidy. Lena, Karin, Linnea, and Jenny for always being helpful and friendly when it comes to administrative stuff. And a whole lot of other good people.

Finally, I thank my family. Malina, Bob, and Harry for love and meaning and my parents, whose support makes my life a whole lot easier.

Jonathan Siverskog
Late at night, 2022

REFERENCES

Altonji, J.G., Elder, T.E. & Taber, C.R. (2005). Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *J Polit Econ,* 113(1), 151-184.

Angrist, J.D. & Pischke, J.S. (2009). *Mostly harmless econometrics : An empiricist's companion,* Princeton: Princeton University Press.

Angrist, J.D. & Pischke, J.S. (2015). *Mastering 'metrics : The path from cause to effect,* Princeton: Princeton University Press.

Angrist, J.D. & Pischke, J.S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *J Econ Perspect,* 24(2), 3-30.

Appleby, J., Devlin, N., Parkin, D., Buxton, M. & Chalkidou, K. (2009). Searching for cost effectiveness thresholds in the NHS. *Health Policy,* 91(3), 239-45.

Aronsson, M., Svennberg, E., Rosenqvist, M., Engdahl, J., Al-Khalili, F., Friberg, L., Frykman, V. & Levin, L.A. (2017). Designing an optimal screening program for unknown atrial fibrillation: a cost-effectiveness analysis. *Europace,* 19(10), 1650-1656.

Birch, S. & Gafni, A. (1993). Changing the problem to fit the solution: Johannesson and Weinstein's (mis) application of economics to real world problems. *J Health Econ,* 12(4), 469-476.

Boadway, R.W. (1974). The Welfare Foundations of Cost-Benefit Analysis. *Econ J,* 84(336), 926-939.

Boadway, R.W. & Bruce, N. (1984). *Welfare economics,* Oxford: Basil Blackwell.

Boardman, A.E., Greenberg, D.H., Vining, A.R. & Weimer, D.L. 2014. *Cost-benefit analysis: Concepts and practice,* Harlow: Pearson Education Limited.

Bobinac, A., van Exel, N.J.A., Rutten, F.F.H. & Brouwer, W.B.F. (2012). Inquiry into the relationship between equity weights and the value of the QALY. *Value Health,* 15(8), 1119-1126.

Booth, N. (2019). On value frameworks and opportunity costs in health technology assessment. *Int J Technol Assess,* 35(5), 367-372.

Brook, R.H., Ware, J.E., Rogers, W.H., Keeler, E.B., Davies, A.R., Donald, C.A., Goldberg, G.A., Lohr, K.N., Masthay, P.C. & Newhouse, J.P. (1983). Does free care improve adults health - results from a randomized controlled trial. *New Engl J Med,* 309(23), 1426-1434.

Brouwer, W., van Baal, P., van Exel, J. & Versteegh, M. (2019). When is it too expensive? Cost-effectiveness thresholds and health care decision-making. *Eur J Health Econ,* 20(2), 175-180.

Brouwer, W.B. & Koopmanschap, M.A. (2000). On the economic foundations of CEA. Ladies and gentlemen, take your positions! *J Health Econ,* 19(4), 439-59.

Brouwer, W.B.F., Culyer, A.J., van Exel, N.J.A. & Rutten, F.F.H. (2008). Welfarism vs. extra-welfarism. *J Health Econ,* 27(2), 325-338.

Brouwer, W.B.F., van Exel, N.J.A., Baltussen, R.M.P.M. & Rutten, F.F.H. (2006). A dollar is a dollar is a dollar - or is it? *Value Health,* 9(5), 341-347.

Burström, K., Sun, S., Gerdtham, U.G., Henriksson, M., Johannesson, M., Levin, L.A. & Zethraeus, N. (2014). Swedish experience-based value sets for EQ-5D health states. *Qual Life Res,* 23(2), 431-42.

Chandra, A., Flack, E. & Obermeyer, Z. (2021). The health costs of cost-sharing. NBER Working Paper No. 28439.

Chen, T.C., Wanniarachige, D., Murphy, S., Lockhart, K. & O'Mahony, J. (2018). Surveying the cost-cffectiveness of the 20 procedures with the largest public health services waiting lists in Ireland: Implications for Ireland's cost-effectiveness threshold. *Value Health,* 21(8), 897-904.

Claxton, K. (1999). The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *J Health Econ,* 18(3), 341-64.

Claxton, K., Martin, S., Soares, M., Rice, N., Spackman, E., Hinde, S., Devlin, N., Smith, P.C. & Sculpher, M. (2015a). Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. *Health Technol Assess,* 19(14), 1-503.

Claxton, K., Sculpher, M., Palmer, S. & Culyer, A.J. (2015b). Causes for concern: is NICE failing to uphold its responsibilities to all NHS patients? *Health Econ,* 24(1), 1-7.

Claxton, K., Walker, S., Palmer, S. & Schulpher, M. (2010). Appropriate perspectives for health care decisions. CHE Research Paper 54.

Cleemput, I., Neyt, M., Thiry, N., De Laet, C. & Leys, M. (2011). Using threshold values for cost per quality-adjusted life-year gained in healthcare decisions. *Int J Technol Assess Health Care,* 27(1), 71-6.

Coast, J. (2009). Maximisation in extra-welfarism: A critique of the current position in health economics. *Soc Sci Med,* 69(5), 786-792.

Cookson, R., Drummond, M. & Weatherly, H. (2009). Explicit incorporation of equity considerations into economic evaluation of public health interventions. *Health Econ Policy Law,* 4(2), 231-245.

Cookson, R., Mirelman, A.J., Griffin, S., Asaria, M., Dawkins, B., Norheim, O.F., Verguet, S. & A, J.C. (2017). Using cost-effectiveness analysis to address health equity concerns. *Value Health,* 20(2), 206-212.

Culyer, A., McCabe, C., Briggs, A., Claxton, K., Buxton, M., Akehurst, R., Sculpher, M. & Brazier, J. (2007). Searching for a threshold, not setting one: the role of the National Institute for Health and Clinical Excellence. *J Health Serv Res Policy,* 12(1), 56-8.

Culyer, A.J. (1978). *Measuring health: Lessons for Ontario,* Toronto: University of Toronto Press.

Culyer, A.J. (2016). Cost-effectiveness thresholds in health care: a bookshelf guide to their meaning and use. *Health Econ Policy Law,* 11(4), 415-32.

Culyer, A.J. (2018). Cost, context, and decisions in health economics and health technology assessment. *Int J Technol Assess,* 34(5), 434-441.

Driscoll, J.C. & Cray, A.C. (1998). Consistent covariance matrix estimation with spatially dependent panel data. *Rev Econ Stat,* 80(4), 549-560.

Eckermann, S. & Pekarsky, B. (2014). Can the real opportunity cost stand up: Displaced services, the straw man outside the room. *PharmacoEconomics,* 32(4), 319-325.

Edney, L.C., Afzali, H.H.A., Cheng, T.C. & Karnon, J. (2018). Estimating the reference incremental cost-effectiveness ratio for the Australian health system. *PharmacoEconomics,* 36(2), 239-252.

Edney, L.C., Lomas, J., Karnon, J., Vallejo-Torres, L., Stadhouders, N., Siverskog, J., Paulden, M., Edoka, I.P. & Ochalek, J. (2022). Empirical estimates of the marginal cost of health produced by a healthcare system: Methodological considerations from country-level estimates. *PharmacoEconomics*, 40(1), 31-43.

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

Ferraro, P.J. & Taylor, L.O. (2005). Do economists recognize an opportunity cost when they see one? A dismal performance from the dismal science. *Contrib Econ Anal Policy,* 4(1).

Finkelstein, A., Taubman, S., Wright, B., Bernstein, M., Gruber, J., Newhouse, J.P., Allen, H., Baicker, K. & Oregon Health Study Group. (2012). The Oregon health insurance experiment: evidence from the first year. *Q J Econ,* 127(3), 1057-1106.

Frederick, S., Novemsky, N., Wang, J., Dhar, R. & Nowlis, S. (2009). Opportunity cost neglect. *J Consum Res,* 36(4), 553-561.

Gafni, A. & Birch, S. (2006). Incremental cost-effectiveness ratios (ICERs): The silence of the lambda. *Soc Sci Med,* 62(9), 2091-2100.

Gallet, C.A. & Doucouliagos, H. (2017). The impact of healthcare spending on health outcomes: A meta-regression analysis. *Soc Sci Med,* 179, 9-17.

Gravelle, H.S.E. & Backhouse, M.E. (1987). International cross-section analysis of the determination of mortality. *Soc Sci Med,* 25(5), 427-441.

Guindo, L.A., Wagner, M., Baltussen, R., Rindress, D., van Til, J., Kind, P. & Goetghebeur, M.M. (2012). From efficacy to equity: Literature review of decision criteria for resource allocation and healthcare decisionmaking. *Cost Eff Resour Alloc,* 10(1), E9.

Hicks, J.R. (1939). The foundations of welfare economics. *Econ J,* 49(196), 696-712.

Jagger, M. & Richards, K. (1969). You can't always get what you want. *Let It Bleed.* London: Decca Records.

Johannesson, M. (1995a). A note on the depreciation of the societal perspective in economic evaluation of health care. *Health Policy,* 33(1), 59-66.

Johannesson, M. (1995b). The relationship between cost-effectiveness analysis and cost-benefit analysis. *Soc Sci Med,* 41(4), 483-9.

Johannesson, M. & O'Conor, R.M. (1997). Cost-utility analysis from a societal perspective. *Health Policy,* 39(3), 241-53.

Johannesson, M. & Weinstein, M.C. (1993). On the decision rules of cost-effectiveness analysis. *J Health Econ,* 12(4), 459-67.

Johnson, F.R. (2012). Why not real economics? *PharmacoEconomics,* 30(2), 127-131.

Juth, N. (2017). For the sake of justice: Should we wrioritize rare diseases? *Health Care Anal,* 25(1), 1-20.

Juth, N., Henriksson, M., Gustavsson, E. & Sandman, L. (2021). Should we accept a higher cost per health improvement for orphan drugs? A review and analysis of egalitarian arguments. *Bioethics,* 35(4), 307-314.

Jönsson, B. (2009). Ten arguments for a societal perspective in the economic evaluation of medical innovations. *Eur J Health Econ*, 10(4), 357-359.

Kaldor, N. (1939). Welfare propositions of economics and interpersonal comparisons of utility. *Econ J,* 49(195), 549-552.

Karlsberg Schaffer, S., Sussex, J., Devlin, N. & Walker, A. (2015). Local health care expenditure plans and their opportunity costs. *Health Policy,* 119(9), 1237-1244.

Karlsberg Schaffer, S., Sussex, J., Hughes, D. & Devlin, N. (2016). Opportunity costs and local health service spending decisions: a qualitative study from Wales. *BMC Health Serv Res,* 16(1), 103.

Karlsson, G. & Johannesson, M. (1996). The decision rules of cost-effectiveness analysis. *PharmacoEconomics,* 9(2), 113-20.

Klein, R. (2010). Rationing in the fiscal ice age. *Health Econ Policy Law,* 5(4), 389-96.

Leamer, E.E. (1983). Lets take the con out of econometrics. *Am Econ Rev,* 73(1), 31-43.

Lindholm, L., Rosen, M. & Emmelin, M. (1998). How many lives is equity worth? A proposal for equity adjusted years of life saved. *J Epidemiol Community Health,* 52(12), 808-811.

Lomas, J., Martin, S. & Claxton, K. (2019). Estimating the marginal productivity of the English National Health Service from 2003 to 2012. *Value Health,* 22(9), 995-1002.

Lundin, D. & Ramsberg, J. (2008). On survival consumption costs: a reply to Nyman. *Health Econ,* 17(2), 293-7.

Martin, S., Lomas, J., Claxton, K. & Longo, F. (2021). How effective is marginal healthcare expenditure? New evidence from England for 2003/04 to 2012/13. *Appl Health Econ Health Policy*, 19(6), 885-903.

Martin, S., Rice, N. & Smith, P.C. (2008). Does health care spending improve health outcomes? Evidence from English programme budgeting data. *J Health Econ,* 27(4), 826-842.

Martin, S., Rice, N. & Smith, P.C. (2012). Comparing costs and outcomes across programmes of health care. *Health Econ,* 21(3), 316-37.

McCabe, C., Claxton, K. & Culyer, A.J. (2008). The NICE cost-effectiveness threshold: what it is and what that means. *PharmacoEconomics,* 26(9), 733-44.

Meltzer, D. (1997). Accounting for future costs in medical cost-effectiveness analysis. *J Health Econ,* 16(1), 33-64.

Miller, S., Johnson, N. & Wherry, L.R. (2021). Medicaid and mortality: New evidence from linked survey and administrative data. *Q J Econ,* 136(3), 1783-1829.

Nikolic, E., Janzon, M., Hauch, O., Wallentin, L. & Henriksson, M. (2013). Cost-effectiveness of treating acute coronary syndrome patients with ticagrelor for 12 months: results from the PLATO study. *Eur Heart J,* 34(3), 220-8.

Nyman, J.A. (2004). Should the consumption of survivors be included as a cost in cost-utility analysis? *Health Econ,* 13(5), 417-427.

Nyman, J.A. (2006). More on survival consumption costs in cost-utility analysis. *Health Econ,* 15(3), 319-322.

Nyman, J.A. (2011). Measurement of QALYs and the welfare implications of survivor consumption and leisure forgone. *Health Econ,* 20(1), 56-67.

O'Donnell, R. (2009). The concept of opportunity cost: Is it simple, fundamental or necessary? *Australas J Econ Educ,* 6(1), 21-37.

O'Mahony, J.F. (2020). Does cost-effectiveness analysis really need to abandon the incremental cost-effectiveness ratio to embrace net benefit? *PharmacoEconomics*, 38(8), 777-779.

Olofsson, S., Gerdtham, U.G., Hultkrantz, L. & Persson, U. (2019). Value of a QALY and VSI estimated with the chained approach. *Eur J Health Econ*, 20(7), 1063-1077.

Oster, E. (2019). Unobservable Selection and Coefficient Stability: Theory and Evidence. *J Bus Econ Stat,* 37(2), 187-204.

Parkin, M. (2016). Opportunity cost: A reexamination. *J Econ Educ,* 47(1), 12-22.

Paulden, M. (2020). Calculating and Interpreting ICERs and Net Benefit. *Pharmacoeconomics*, 38(8), 785-807.

Paulden, M. & McCabe, C. (2021). Modifying NICE's approach to equity weighting. *Pharmacoeconomics,* 39(2), 147-160.

Paulden, M., McCabe, C. & Karnon, J. (2014). Achieving allocative efficiency in healthcare: nice in theory, not so NICE in Practice? *PharmacoEconomics,* 32(4), 315-8.

Paulden, M., O'Mahony, J. & McCabe, C. (2017). Determinants of change in the cost-effectiveness threshold. *Med Decis Making,* 37(2), 264-276.

Pekarsky, B. 2015. *The new drug reimbursement game: A regulator's guide to playing and winning,* Cham: Adis.

Persson, E. & Tinghög, G. (2020). Opportunity cost neglect in public policy. *J Econ Behav Organ,* 170(February 2020), 301-312.

Phelps, C.E. & Mushlin, A.I. (1991). On the (near) equivalence of cost-effectiveness and cost-benefit analyses. *Int J Technol Assess Health Care,* 7(1), 12-21.

Potter, J. & Sanders, S. (2012). Do economists recognize an opportunity cost when they see one? A dismal performance or an arbitrary concept? *South Econ J,* 79(2), 248-256.

Prop. 1996/97:60. Prioriteringar inom hälso- och sjukvården. Stockholm: Socialdepartementet.

Robbins, L. 1932. *An essay on the nature and significance of economic science,* London: Macmillan & Co., Ltd.

Rosling, H. & Hägersten, F. (2017). *Hur jag lärde mig förstå världen,* Stockholm: Natur & Kultur.

Round, J. & Paulden, M. (2018). Incorporating equity in economic evaluations: a multi-attribute equity state approach. *Eur J Health Econ,* 19(4), 489-498.

Sandman, L. & Gustavsson, E. (2017). The (ir)relevance of group size in health care priority setting: A reply to Juth. *Health Care Anal,* 25(1), 21-33.

Sculpher, M., Claxton, K. & Pearson, S.D. (2017). Developing a value framework: The need to reflect the opportunity costs of funding decisions. *Value Health,* 20(2), 234-239.

SFS 2002:160. Lag om läkemedelsförmåner m.m. Stockholm: Socialdepartementet.

SFS 2017:30. Hälso- och sjukvårdslag. Stockholm: Socialdepartementet.

Sommers, B.D. (2017). State medicaid expansions and mortality, revisited: A cost-benefit analysis. *Am J Health Econ,* 3(3), 392-421.

Staiger, D. & Stock, J.H. (1997). Instrumental variables regression with weak instruments. *Econometrica,* 65(3), 557-586.

Stinnett, A.A. & Mullahy, J. (1998). Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Med Decis Making,* 18(2 Suppl), S68-80.

Stock, J.H. & Yogo, M. (2005). Testing for weak intruments in linear IV regression. *In:* Andrews, D. W. K. & Stock, J. H. (eds.) *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg*. New York: Cambridge University Press.

Swedish Agency for Public Management (2014). *Det kommunala utjämningssystemet - en beskrivning av systemet från 2014* (2014:2).

The National Board of Health and Welfare (2018). *Kapacitet i sjukvården - att hantera kritiska lägen.*

TLV (2003). *Läkemedelsförmånsnämndens allmänna råd om ekonomiska utvärderingar* (LFNAR 2003:2).

TLV (2015). *Ändring i Läkemedelsförmånsnämndens allmänna råd (LFNAR 2003:2) om ekonomiska utvärderingar* (TLVAR 2015:1).

TLV (2016a). *TLV accepterar högre kostnad för läkemedel mot sällsynt sjukdom* [Internet]. Available from: https://www.tlv.se/om-oss/press/nyheter/arkiv/2016-12-20-tlv-accepterar-hogre-kostnad-for-lakemedel-mot-sallsynt-sjukdom.html [Accessed 23 March 2020].

TLV (2016b). *Underlag för beslut om subvention - omprövning: Cerezyme (imiglukeras) och Vpriv (velaglukeras alfa)* (Dnr 1961/2015).

TLV (2017). *Ändring i Tandvårds- och läkemedelsförmånsverkets allmänna råd (TLVAR 2003:2) om ekonomiska utvärderingar* (TLVAR 2017:1).

TLV (2018a). *Beslut* (Dnr 45/2018).

TLV (2018b). *Ny tillämpning av etiska plattformen* [Internet]. Available from: https://www.tlv.se/lakemedel/halsoekonomi/ny-tillampning-av-etiska-plattformen.html [Accessed 6 March 2020].

TLV (2018c). *Underlag för beslut om subvention - Nyansökan: Orkambi (lumakaftor + ivakaftor)* (Dnr 45/2018).

Vallejo-Torres, L., Garcia-Lorenzo, B. & Serrano-Aguilar, P. (2018). Estimating a cost-effectiveness threshold for the Spanish NHS. *Health Econ,* 27(4), 746-761.

van Baal, P., Perry-Duxbury, M., Bakx, P., Versteegh, M., van Doorslaer, E. & Brouwer, W. (2019). A cost-effectiveness threshold based on the marginal returns of cardiovascular hospital spending. *Health Econ,* 28(1), 87-100.

Vanness, D.J., Lomas, J. & Ahn, H. (2021). A health opportunity cost threshold for cost-effectiveness analysis in the United States. *Ann Intern Med,* 174(1), 25-32.

Vårdfokus (2021). *Bristen på vårdplatser – en fara för både personal och patienter* (Reporter: Westin, J.). Available from: https://www.vardfokus.se/nyheter/bristen-pa-vardplatser-en-fara-for-bade-personal-och-patienter/ [Accessed 28 February 2022].

Wagstaff, A. (1991). QALYs and the equity-efficiency trade-off. *J Health Econ,* 10(1), 21-41.

Walker, S., Griffin, S., Asaria, M., Tsuchiya, A. & Sculpher, M. (2019). Striving for a societal perspective: A framework for economic evaluations when costs and effects fall on multiple sectors and decision makers. *Appl Health Econ Health Policy*, 17(5), 577-590.

Wammes, J.J.G., Frederix, G., Govaert, P., Determann, D., Evers, S., Paulus, A., Stadhouders, N., Jeurissen, P., Oortwijn, W. & Adang, E.M.M. (2020). Case-studies of displacement effects in Dutch hospital care. *BMC Health Serv Res,* 20(1), E263.

Weinstein, M. & Zeckhauser, R. (1973). Critical ratios and efficient allocation. *J Public Econ,* 2(2), 147-157.

Weintraub, W.S., Mahoney, E.M., Lamy, A., Culler, S., Yuan, Y., Caro, J., Gabriel, S., Yusuf, S. & Investigators, C.S. (2005). Long-term cost-effectiveness of clopidogrel given for up to one year in patients with

acute coronary syndromes without ST-segment elevation. *J Am Coll Cardiol,* 45(6), 838-845.

Williams, A.H. & Cookson, R.A. (2006). Equity-efficiency trade-offs in health technology assessment. *Int J Technol Assess Health Care,* 22(1), 1-9.

Papers

The papers associated with this thesis have been removed for copyright reasons. For more details about these see:

https://doi.org/10.3384/9789179291365

FACULTY OF MEDICINE AND HEALTH SCIENCES

Linköping University Medical Dissertation No. 1795, 2022
Department of Health, Medicine, and Caring Sciences

Linköping University
SE-581 83 Linköping, Sweden

www.liu.se

LINKÖPING
UNIVERSITY