Public preference or patient experience?

Deciding on a metric for healthcare allocation

Micah Kaats
Applied Ethics MSc Thesis
Utrecht University | June 2020

Table of contents

Abstract

This thesis addresses the question of cost-effectiveness in healthcare. The discussion primarily focuses on metrics of distribution. Two main candidates are considered: public preferences and patient experiences. While I assume that allocating medical resources to promote positive health outcomes and minimize negative health outcomes is justified, I provide strong reasons to doubt the normative legitimacy of prevailing approaches that rely on public preference metrics to do so. The discussion proceeds in two parts. In the first part, I provide an overview of current practices to measure cost-effectiveness in healthcare. I highlight key discrepancies between public and patient valuations of disease and disability. I then discuss three of the most powerful arguments ushered in defense of public preference metrics of allocation in the health sector. These arguments cluster around three main themes: democracy, impartiality, and adaptation. Ultimately, I argue that all three cannot withstand normative scrutiny. Potential appeals to expert preferences and patient preferences are also considered and rejected. In the second part, I highlight key normative lessons that can be learned from this analysis and use them to inform the design and development of a new metric of cost-effectiveness rooted in patient experience. I argue for an approach that relies on subjective life evaluations of patients and healthy counterparts to estimate actual and potential wellbeing. Several distributive implications are discussed. The deficiencies of prevailing approaches to medical resource allocation in the health sector demand creative solutions to improve the metrics upon which they rely. This thesis attempts to provide one.

Introduction

How should we allocate medical resources to those who stand to benefit from them?[1]

The health industry is one of the largest and fastest growing sectors in the world. In 2016, more than 7.5 trillion US dollars was spent on healthcare, almost 10% of global GDP, and in many countries substantially more.[2] For the last two decades, health-related expenditures have grown at a rate of 4% per year, even faster than the 2.8% annual growth rate of the global economy.[3] Nevertheless, demand for medical services has consistently outstripped supply. Despite enormous investments, public and private insurance providers, hospitals, and pharmaceutical manufacturers have limited resources, limited funding, and limited capacity. Allocation of treatments, interventions, technologies, services, and medications is inevitable. Faced with mounting responsibilities and expanding budgets, stakeholders in both the public and private sector have increasingly relied on economic evaluations to make resource allocation and investment decisions.

For decades, economists and philosophers have recognized the normative significance of rationing in the health sector.[4] As a starting point, random allocation has seemed entirely unacceptable. Most observers would reject the idea that decisions to treat children with brain cancer or adults with back pain ought to be left up to chance. We need to have some systematic way of evaluating and comparing health outcomes in order to direct resources to those in need. In this context, the criterion of efficiency has been largely persuasive.[5] Given limited supply, there seems to be an ethical imperative to ensure that allocated medical resources are put to the best use possible.

The intuitive plausibility of the efficiency requirement has motivated an extensive literature on rationing strategies based on cost-effectiveness analysis (CEA).[6] If limited medical resources are considered as costs and positive health outcomes as benefits, then presumably we should seek to get as many benefits from as few costs as possible.[7] This strand of thinking has its intellectual roots in utilitarianism, which is itself a species of consequentialism.[8] While consequentialists understand the moral worth of an action in terms of the outcomes they produce, utilitarians take the additional step of defining moral actions as those that produce the greatest good for the greatest number. In the present context, the criterion of cost-effectiveness in healthcare stipulates that medical resources should be allocated to promote positive health outcomes and minimize negative health outcomes.

However, this presents a number of substantial methodological challenges. While longevity and death are typically assumed to represent endpoints on the continuum of good and bad health, differentiating what lies between them can be exceedingly difficult. Evaluating cost-effectiveness in healthcare requires putting seemingly incommensurate outcomes on the same scale – e.g. the relative benefits of reducing chronic pain, delaying early onset dementia, preventing depressive episodes, improving mobility, preventing malaria, and so on. Doing so requires the identification of some underlying "latent trait" common to all possible health improvements that is at least roughly ordinal, measurable, and comparable.[9] Put simply, we need a metric of cost-effectiveness.

---

[1] Many of the ideas in this thesis were originally presented in Happiness Research Institute and Leaps by Bayer (2020). I am grateful to Michael Birkjær, Alejandro Rubio, and Meik Wiking for their patient guidance and to my advisor, Siba Harb, for her thoughtful comments and support.

[2] World Bank data on current health expenditure (% of GDP). Retrieved from: https://data.worldbank.org/indicator/SH.XPD.CHEX.GD.ZS?locations=1W-XD-XM-XP

[3] Xu et al. (2018).

[4] Ubel (2001).

[5] Bickenbach (2016); Neumann et al (2016).

[6] Neumann et al (2016).

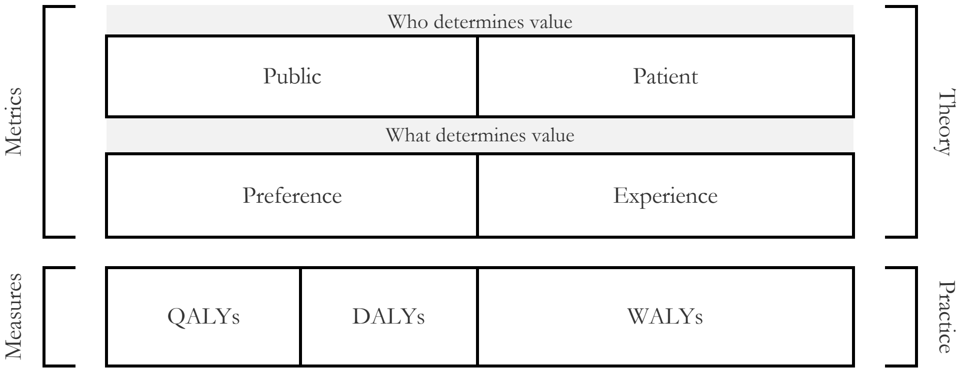
[7] Bickenbach (2016).

[8] Marseille & Kahn (2019).

[9] Bickenbach (2016).

This thesis offers an investigation into two leading candidates. Broadly speaking, treatments and interventions in the health sector can be deemed cost-effective in terms of their ability to satisfy public preferences or improve patient experiences.[10] In this analysis, I will use the term "metrics" to refer to these as theoretical constructs of value and "measures" to refer to the real-world instruments designed to capture them (Table 1). In practice, while public preference measures aim to provide insight into public opinion regarding various health conditions, patient experience measures aim to shed light on how patients experiencing those conditions judge their own quality of life for themselves.

Table 1: Metrics and measures of health



Public preference metrics – the incumbent candidate – have been widely implemented and play an important role in determining the global distribution of medical resources. The two most influential measures in this respect are Quality Adjusted Life Years (QALYs) and Disability Adjusted Life Years (DALYs). Both base their assessments of quality of life on stated public preferences. To calculate QALYs and DALYs, members of the public are asked to judge the severity and disability levels of various health states in an attempt to determine which diseases are generally considered to be better and worse than others.

Patient experience metrics – the challenger – have gained significant traction in psychology and economics, but remain mostly underdeveloped in the arena of healthcare decision-making. In practice, researchers typically measure subjective experience in terms of two dimensions: affect and evaluation. While measures of affect capture the quality of emotional experience – e.g. joy, sadness, anxiety, anger – measures of evaluation capture overall assessments of life as a whole – e.g. life satisfaction. In the health sector, an approach to cost-effectiveness analysis rooted in subjective experience would therefore seek to direct energy and investment towards those conditions and circumstances that are responsible for the greatest sources of patient suffering, as evaluated by patients themselves.

Despite their popularity, in this thesis I will argue that public preference metrics fail to provide an ethically justifiable means by which to allocate resources in the health sector. In doing so, I will take onboard two fundamental assumptions. First, that cost-effectiveness analysis is an appropriate means by which to allocate medical resources. Second, that the value of health can be assessed in terms of its impact on quality of life, or wellbeing. However, if the objective of cost-effectiveness analysis is to promote positive health outcomes and minimize negative health outcomes, rationing strategies based on satisfying public preferences cannot succeed. Ultimately, I hope to demonstrate that public preference metrics fail to live up to the normative and descriptive goals they set out for themselves.

---

[10] Three notes on terminology. First, throughout this thesis I will use the term "patients" to refer broadly to those experiencing all forms of disease and disability, and not necessarily those under the care or supervision of a physician. While this wording is admittedly unfortunate and bound to be controversial, I have chosen my language to be consistent with the relevant literature in health economics and philosophy. Second, the boundaries surrounding health, disease, and disability are porous and different opinions regarding their constitutive elements have been widely debated (Kovács, 1998). For the purposes of this discussion, I will broadly adopt the definition of health provided in the preamble to the constitution of the World Health Organization: "Health is a state of complete physical, mental and social wellbeing and not merely the absence of disease or infirmity" (Grad, 2002). Third, I will also use the terms "wellbeing" and "quality of life" interchangeably. While I take the idea that health can be measured in terms of quality of life to be widespread and foundational to standard practices of cost-effectiveness analysis in healthcare, it is not entirely uncontroversial. For an alternative approach to valuing health beyond its contribution to wellbeing, see Hausman (2015).

Given their widespread implementation, my arguments will mostly focus on metrics and measures that rely on the preferences of the general public to value health. However, I will also discuss the potential for expert preferences as well as patient preferences to serve as metrics of allocation in the health sector. Ultimately, I intend to argue that all three are ethically unfit for the aims of cost-effectiveness analysis in healthcare. Instead, I will argue for an approach to healthcare decision-making that is rooted in metrics of patient experience. Since no unified measure of patient experience in healthcare currently exists, in the second half of the thesis I will develop a practical and normative framework upon which one might stand, what I refer to as Wellbeing Adjusted Life Years (WALYs). While this discussion will primarily focus on metrics of allocation, I also recognize that these can have important distributive implications. In the concluding section, I will therefore highlight two key distributive issues that arise from the analysis presented in this thesis and comment on how they may be informed and developed in future research.

The thesis proceeds as follows. In Part I, I will introduce and evaluate medical resource allocation strategies that employ public preferences as the metric of value. This discussion will proceed from the bottom-up, from practice to theory. To better understand their normative significance, I will first provide a detailed introduction to the methodology underpinning QALYs and DALYs (Section 1.1). While both measures represent important steps to capture nuanced understandings of quality of life, in practice they often prove to be misaligned with patient experience (Section 1.2). How the public imagines disease and how patients experience disease are often at odds. This divergence raises the stakes in deciding which measure to prioritize. In fact, it begs the question why we would be justified in jettisoning the self-reports of patients in favor of public preferences in the first place. At this point I will zoom out to consider the strongest theoretical arguments put forth in defense of public preferences metrics. These cluster around three main themes, what I refer to as the democratic argument (Section 1.3), the impartiality argument (Section 1.4), and the adaptation argument (Section 1.5). By evaluating each individually, I hope to demonstrate that all three fail to provide strong reasons to favor public preference valuations of health. However, this discussion is not merely intended to be destructive. I will also highlight key lessons that can be learned from the failings of public preferences to inform the design and development of a new metric of cost-effectiveness in healthcare.

In Part II, I will attempt to lay the groundwork for a new metric rooted in patient experience. This discussion will begin by reiterating the lessons learned in the first half of the thesis. For any metric of cost-effectiveness to succeed, there are several criteria it should seek to meet. The metric should offer room for public involvement, harness collective knowledge, limit the potential for motivated reasoning and cognitive bias, rely on reliable sources, and be practically and meaningfully feasible. Once these standards are set, I will proceed again from the bottom-up by surveying the measures of subjective experience currently on offer in economics and psychology and evaluating their potential to inform healthcare decision-making (Section 2.1). Ultimately, I will argue that life evaluations – self-reported global assessments of life as a whole – provide the most suitable foundation upon which to build a new approach to resource allocation in the health sector. I will then sketch out how this approach could be realistically applied in practice using a new measure of patient experience: Wellbeing Adjusted Life Years (WALYs). I evaluate how this approach performs against the established criteria that emerged in the first half of the discussion (Section 2.2). Finally, I will close by commenting on two important distributive implications that remain unaddressed and highlight key areas for future research (Section 2.3).

Part I: Imagine all the health states

For the better part of the last century, economists have tended to think about wellbeing in terms of welfare, or utility. Wellbeing by this account consists in the fulfilment of desires or preferences. The more preferences a person is able to fulfill, the better off they become. This idea is motivated by the assumption that, given the right conditions, people will always act rationally to satisfy whichever preferences maximize their wellbeing. If competitive markets are left to their own devices, consumers can then achieve efficient allocations of resources simply by buying whichever goods make them better off. Individual preferences are revealed by purchase behavior, the market adapts to fill those preferences, profit is maximized, and efficiency is achieved.

Whatever one thinks of this general story, it becomes considerably warped in the health sector. From the perspective of efficiency as financial maximization, health benefits delivered to the most-well off in society quickly become more cost-effective than those delivered to the most disadvantaged. Willingness-to-pay collapses into ability-to-pay.[11] Adopting an allocation strategy of medical resources that relies so heavily on wealth and income has generally seemed ethically indefensible.[12] However, this presents a problem. If traditional market mechanisms cannot be relied upon to inform the allocation of resources, what should be?

In response to these challenges, health economists have increasingly relied on metrics of cost-effectiveness informed by stated public preferences. In practice, to evaluate efficiency in healthcare, representative samples of the general public are asked to evaluate and compare different health states and diseases to determine which has a higher level of disability. Medical resources can then be efficiently allocated towards whichever conditions the public generally considers to be worse than others. This approach has become widely implemented around the world and provides the foundation for conducting cost-effectiveness analysis using Quality Adjusted Life Years (QALYs) and Disability Adjusted Life Years (DALYs). In what follows, I will describe the methodology underpinning QALYs and DALYs in greater detail and discuss their most important limitations.

This discussion will proceed in the three steps. First, I will present an overview of how QALYs and DALYs are calculated in practice and used to inform healthcare decision-making. Second, I will highlight key discrepancies between public preference valuations of health and patient subjective experiences. In practice, the diseases and disabilities the public generally considers to be worse than others are not always the ones that seem to have the largest effects on self-reported patient quality of life. As a result, when considering which to prioritize, public preferences or patient experiences, the stakes are raised considerably. At this point, I will address the three most important arguments ushered in defense of public preference metrics in healthcare. I intend to argue that none can withstand normative scrutiny. Given their influence, this discussion with mostly focus on measures and metrics informed by the stated preferences of the general public. However, I will also discuss the potential implications of relying on expert preferences and even patient preferences instead as metrics of cost-effectiveness analysis in healthcare. Ultimately, while these may have comparative advantages, the enterprise of valuing health with preferences at all emerges as fundamentally flawed. I will conclude with several key lessons that can be taken away from this analysis and used to inform the design of new metrics of medical resource allocation.

1.1 QALYs and DALYs

Quality Adjusted Life Years (QALYs) are one of the most commonly used tools to evaluate costs and benefits in the health sector. They provide a common currency by which to compare a wide variety of medical interventions in terms of their impacts on longevity and quality of life. Longevity is measured in terms of the number of potentially added life years from a given intervention. Quality of life is then estimated using two inputs. First, patients are asked to report their subjective health status using questionnaires such as the EQ-5D. Second, representative samples of the general public are asked to judge the disability level of each health state using stated preference techniques.

---

[11] Dolan & Tsuchiya (2006).

[12] Hausman (2015, p. 190).

Developed in the late 1980s by an international panel of health experts and medical professionals, the EQ-5D has become the most widely used instrument to measure subjective health.[13] It asks respondents to evaluate their health in terms of five dimensions: (1) mobility, (2) self-care, (3) usual activities, (4) pain/discomfort, and (5) anxiety/depression. Each dimension has three levels: (1) no problems, (2) some problems, and (3) severe problems. Responses to these questions can then be used to generate summary scores representing one of 243 possible health states. For example, those with some problems walking about (mobility = 2), some problems washing or dressing (self-care = 2), no problems performing usual activities (usual activities = 1), no problems with pain or discomfort (pain/discomfort = 1), and some problems with anxiety or depression (anxiety/depression = 2) are given a health summary score of 22112. In clinical trials, these scores are collected regularly to track patient progress over time. Ultimately, they allow medical professionals to make determinations regarding the effects of interventions in fields ranging from oncology to dermatology.[14]

However, the responsibility of judging which health states are actually better or worse than others is then conferred upon the general public. To elicit these judgements, researchers typically present representative samples with ten to fifteen health states elicited from the EQ-5D. Participants are asked to respond to the following scenario for each health state using the Time Trade Off (TTO) method: "Imagine you have ten years left to live. You can either live in [health state] for all ten years or choose to give up some years to live for a shorter period of time in full health. Indicate the number of years in full health that you think is of equal value to ten years in [health state]."[15] Other valuation techniques used to calculate QALYs include the Standard Gamble (SG) and Visual Analogue Scale (VAS). Like the TTO, the SG also asks participants to choose between two hypothetical scenarios – e.g. living 20 more years with blindness in one eye or living 17 more years with perfect health – while the VAS asks respondents to rank health states on a visual scale from zero to 100, with zero indicating death and 100 indicating perfect health.[16]

One widely cited study using the TTO method in the United States found for example that most people would be willing to give up three full years of healthy life to avoid living ten years in the health state 22112 described above.[17] Quality weights for each health state are then calculated by dividing the total number of years participants would rather live in full health by ten years. The quality weight for the health state 22112 is then 0.7 (seven years divided by ten years), implying that one year in full health is equivalent to 0.7 QALYs in this state. QALYs range from 0 indicating death to 1 indicating perfect health. Studies conducted along these lines have been carried out throughout Western Europe, the Americas, Asia, and Africa.[18]

Once public preferences are collected and aggregated, they are then used to infer which health conditions are generally considered to be more or less desirable than others.[19] Armed with this knowledge, policymakers and healthcare professionals can then go about allocating medical resources towards whichever interventions and treatments efficiently address conditions the public considers to be more dire. Cost-effectiveness for new technologies is assessed in terms of the amount of QALYs they produce. In the United Kingdom for instance, the National Institute of Care and Excellence (NICE) recommends a cost-effectiveness threshold of £20,000 to £30,000 per QALY for new medical treatments.[20] Any new treatment or intervention that NICE evaluates to be capable of producing one extra QALY at less than this cost threshold is deemed to be cost-effective and formally recommended to the National Health Service (NHS). This process of conducting healthcare appraisals based on quality weights elicited from public preferences has also become standard practice in Australia, Canada, China, France, the Netherlands, South Africa, and Sweden, among others.[21] In practice, these

---

[13] This original version, the EQ-5D-3L, was later expanded to five levels in the EQ-5D-5L, although the former remains much more widely used. Other instruments including the SF-36 and HUI can also been used to calculate QALYs. Both rely on preference weights to value health states. For a thorough discussion, see Whitehead & Ali (2010).

[14] Rabin & Charro (2001).

[15] Attema et al. (2013).

[16] Alex & Wyrwich (2003); Craig et al. (2009).

[17] Shaw et al. (2005).

[18] Xie et al. (2014).

[19] For a detailed methodology, see Shaw et al. (2005).

[20] Recent estimates have suggested that the actual figure is closer to £13,000 in practice. See Claxton et al. (2013).

[21] Brazier et al. (2019); Rowen et al. (2017); Zhao et al. (2018).

methods are most often used to evaluate the cost-effectiveness of pharmaceuticals, followed by surgical procedures, non-diagnostic procedures, and screening processes.[22]

In the developing world, Disability Adjusted Life Years (DALYs) are more often used to asses comparative health burdens and impacts. DALYs can be considered as the inverse of QALYs. While QALYs measure health states on a scale from 0 indicating death to 1 indicating perfect health, DALYs measure health on a scale from 0 indicating perfect health to 1 indicating death. One DALY is therefore equal to one year of healthy life lost. In practice, cost-effectiveness is then measured in terms of Disability Adjusted Life Years averted as opposed to Quality Adjusted Life Years gained.

Like QALYs, DALYs also assign "disability weights" to various diseases based on elicited health preferences of the general population. However, instead of relying on health statuses obtained from the EQ-5D, or other patient questionnaires, disability weights are typically elicited by presenting public samples with pairwise comparisons of diseases using lay descriptions.[23] For example, in the most recent analysis of disability weights for the Global Burden of Disease study, representative samples of European countries were asked to judge which hypothetical person was healthier: someone who "drinks a lot of alcohol and sometimes has difficulty controlling the urge to drink and while intoxicated has difficulty performing daily activities" – indicating mild alcohol use disorder – or someone who "has sudden, prolonged seizures once a week, with violent muscle contractions and stiffness, loss of consciousness, loss of urine or stool control and between seizures has drowsiness, memory loss, difficulty concentrating and anxiety" – indicating severe epilepsy.[24] In the published study, this process was repeated for 255 diseases. Researchers then analyzed response patterns to estimate which conditions the public generally considers to be more burdensome than others.

In recent years, prominent international organizations including the World Health Organization (WHO) and the Bill and Melinda Gates Foundation have adopted an "average cost per DALY averted" metric to guide their global health investments. DALYs are generally preferred for the purposes of global health as their design allows them to be much broader in scope. Weights are assigned to diseases as a whole as opposed to specific health states. The resulting DALY estimations can then be more easily employed to evaluate differences in overall health status across populations. This level of abstraction may be appropriate when the objective is to compare macro level health conditions between or within countries over time. However, for the more targeted purpose of determining the effectiveness of specific treatments or interventions in clinical trials, QALYs are generally preferred. The health state questionnaires they rely on, including the EQ-5D, are capable of detecting more subtle changes in patient quality of life over shorter periods of time. This allows healthcare professionals to make nuanced comparisons between patients experiencing the same disease.[25]

While QALYs and DALYs represent important and encouraging steps to measure and model impact in the health sector beyond financial return, there are a number of ways these metrics can fall short of reliably capturing patient quality of life. In the following sections, I will first address the practical implications of prioritizing public preference measures above patient experience measures, before taking a step back to evaluate the three most compelling theoretical arguments ushered in defense of the former. Ultimately, I hope to demonstrate that all three arguments fail to provide strong moral reasons to prioritize public preferences in healthcare decision-making.

1.2 The valley between public preference and patient experience

In deciding which metric to prioritize in the allocation of medical resources – public preferences or patient experiences – it is crucial to note that the two are often misaligned.

For example, in general population surveys, physical health is often considered to be more determinant of quality of life than mental health. To most people, a life that is physically challenging usually seems harder than one that is mentally

---

[22] Neumann et al. (2009).
[23] Vos et al. (2017).
[24] Haagsma et al. (2015).
[25] Gold et al. (2002).

challenging.[26] One meta-analysis found that, on average, severe pain is judged to be 1.4*x* worse than severe anxiety or depression. Severe immobility is judged to be 1.6*x* worse.[27] However, when patients are asked to review the quality of their lives for themselves, the story can change dramatically.[28] An empirical analysis of 15,184 patient self-reports found that severe anxiety and depression had negative effects on patient subjective wellbeing that were more than 10*x* worse than severe pain. Once the effects of depression and anxiety were controlled for, severe immobility was found to have no effect at all.[29] This result was replicated in a similar analysis of American adults.[30]

Patients with colostomies, dialysis, and rheumatoid arthritis also generally report higher levels of wellbeing than expected.[31] On the other hand, patients with conditions that affect social functioning generally reported lower levels of wellbeing than expected. Limited social functioning – defined as not being able to see friends or relatives on a regular basis – can have negative effects on self-reported quality of life that are 1.5*x* more severe than limited physical functioning. However, in public preference surveys, limited physical functioning is deemed to be 1.3*x* worse.[32] Other gaps between public preferences and patient experience have been observed for visual impairments, heart disease, and asthma.[33]

Several explanations have been put forth to explain these divergences.[34] Here I mention two in particular. One frequently cited culprit is the focusing illusion - a cognitive bias that occurs when our attention is directed to a specific event or aspect of our lives in such a way that makes it seem more important than it actually is.[35] This tendency can severely distort health state valuations based on public preferences. When samples of the general population are presented with descriptions of health states and asked to judge which one is worse, their attention is immediately drawn to the disability as if they were the only thing that matters. Other domains of life that may matter more, such as personal relationships and social context, are not taken into account.

Second, in practice, while the most extreme cases of pain and discomfort can be severely debilitating, most physical symptoms only tend to affect patient wellbeing in the moment.[36] They are unlikely to be endless daily sources of frustration. On the other hand, almost by definition, mental health conditions constantly demand attention. Their enduring presence can make them exceedingly difficult to get used to.[37] For those with little experience of mental or physical disability, these dynamics can be difficult to imagine or appreciate, especially when prompted to do so in the context of short surveys. As a result, when members of the public are asked to judge the severity of health states, they may be more prone to overweight the effect of physical disabilities and underweight the effect of social or mental disabilities.

Given these observed discrepancies between public preferences and patient experiences, prioritizing one can often come at the cost of deprioritizing the other. When we then consider the arguments provided in defense of public preferences for QALY and DALY calculations, the stakes are raised considerably. We need to have strong reasons to jettison the self-reports of patients in favor of the stated preferences of the public, or vice versa. Treating them equally cannot solve the problem. Giving equal weight to public preference and patient experience also demands justification. If one party is uniquely ill-equipped to make these sorts of judgements, then simply aggregating them together is not going to get us any closer to the truth. If the goal is to understand the laws of particle physics, polling the public would seem like an odd strategy. Weighing public preferences and patient experiences equally is only justified if we have good reason to do so. As I intend to argue, we do not. Nevertheless, given their widespread implementation, it is worth giving the most common

---

[26] Xie et al. (2014). Shaw et al. (2005).

[27] Xie et al. (2014). See Table 2.

[28] Binder & Coad (2013); Graham et al (2011).

[29] Mukuria & Brazier (2013). See Table 3.

[30] Dolan & Metcalfe (2012).

[31] Brazier et al. (2005); Menzel et al. (2002).

[32] Mukuria & Brazier (2013). See Table 4.

[33] Halpern & Arnold (2008).

[34] For thorough discussions, see Dolan & Kahneman (2008); Brazier et al (2005); Menzel et al. (2002).

[35] Dolan & Kahneman (2008).

[36] Dolan (2011); Dolan & Kahneman (2008).

[37] Dolan (2011); Levack et al. (2004).

arguments put forth in defense of public preference measures a proper hearing. In what follows, I will therefore take a step back to consider the most powerful theoretical justifications ushered in their defense.

One of the most influential guidelines on the use of cost-effectiveness analysis in healthcare is provided in *Cost-Effectiveness in Health and Medicine*.[38] Written by a panel of experts from decision science, economics, ethics, psychology, and medicine, the volume defends the use of community preferences to value health for several interrelated reasons. The main arguments cluster around three main themes, what I consider to be principles of democracy, impartiality, and adaptation.[39] In the remaining sections of this part of the thesis, I will address each individually. I intend to demonstrate that all three arguments fail to provide sufficient reason to prioritize public preferences in cost-effectiveness analysis. However, the purpose of this exercise is not merely intended to be destructive. In evaluating these arguments on their own terms, I will also attempt to unearth and clarify important lessons and takeaways that can be learned from them to inform the design and development of new metrics of cost-effectiveness that may be able to perform better.

1.3 The democratic argument

The democratic argument can be summarized as follows:

> DEMOCRATIC ARGUMENT: Since members of the public are the funders and potential beneficiaries of healthcare systems, public preferences should inform healthcare policy.[40]

I take this argument to be democratic in that it seems to reflect two of the most powerful arguments put forth in favor democracy itself. The first is simply that the public should decide where and how their own money is spent. One of the most important features of any democratic system is the commitment to decision-making processes in which the interests of taxpayers are fairly represented. In the present context, this would imply that publicly funded healthcare systems should reflect public priorities. I take this to be the most overt and important implication of the democratic argument.

The second is more subtle. Democratic processes of decision-making may also be justified inasmuch as they improve the quality of outcomes.[41] Some matters of policy can be better informed by deferring to the "dispersed knowledge of citizens" than to any individual or centralized authority.[42] In the present context, since members of the public are potential beneficiaries of healthcare systems, they have a stake in ensuring that the system functions well. Relying on democratic procedures of decision-making may be the best way to ensure that medical resources are put to the best use possible.[43] I consider democratic procedures in this context to be the elicitation of preference valuations of health states from representative samples of the public. However, as I will discuss, this argument could also be used to justify preference valuations of health states by experts or patients themselves. Perhaps it is their dispersed knowledge that should be accumulated to improve the quality of outcomes.

Both of these claims are perfectly reasonable. Public preferences should play a role in decision-making and democratic processes can improve the quality of outcomes. However, relying on preference valuations of health states to inform cost-effectiveness analysis is neither implied by the first claim, nor capable of satisfying the second. In what follows, I will take them each individuals.

---

[38] Neumann et al. (2016).

[39] Neumann et al. (2016). Variations of these arguments have also been defended elsewhere. For discussions, see Brazier et al. (2005); Menzel et al. (2002); Gandjour (2010); Dolan (1999); Nord (1999); Hadorn (1991); Salomon et al. (2010); Gold et al. (1996).

[40] The specific wording of Neuman et al. (2016) is "It is preferred that analysts measuring preferences directly use a representative sample of the general public, who are suitably informed about the health states being valued...the reasoning is that they are the ultimate payers of the publicly funded health care system and potential patients. As costs are borne by potential beneficiaries of publicly funded healthcare systems prior to the potential for benefit, so health effects should also be considered in an ex-ante way" (p. 182). This has also been summarized elsewhere by Mukuria & Brazier (2013) as "...the cost and consequences of health care are borne by the general population and their preferences should therefore inform decision-making" (p. 98). Salomon et al. (2012) states: "Much of the scientific literature about health-state weights uses the responses of the general public, on the basis of the argument that, in a democratic society, the views of the general public are relevant in comparative assessments that inform public policy" (p. 2131). This claim has also been referred to elsewhere as the "insurance principle" (Gold et al. 1996).

[41] This has elsewhere been characterized as the epistemic argument for democracy (Holst & Molander, 2019).

[42] Schwartzberg, M. (2015, p. 196).

[43] While this latter claim is not explicitly addressed by Neuman et al. (2016), it does seem worthy of consideration. The philosopher Daniel Hausman characterizes this argument as the potential for public preferences to serve as "*data* concerning the values of health states" (2015, p. 98).

It would be hard to deny that public priorities should inform public policy. However, there are surely limitations to this general principle. Even the most committed democrats would not suggest that the public sign off on every potentially relevant policy decision. We don't expect public audiences to determine surgery protocols, establish treatment plans, or prescribe medications. Why then should we expect them to determine the severity of disabilities? There seems to be an important distinction to be made in this regard between setting the normative objectives of healthcare policy and implementing them. Public preferences can and perhaps should inform the former, but not necessarily the latter. This latter category is the domain in which preference valuations of health states belongs.

There are a number of ways public preferences could, and perhaps should, be worked into the policymaking process. How should the needs of the elderly be weighed against the needs of the young? To what extent should disadvantaged groups be given priority? What level of healthcare should be provided to every citizen? These may be important questions for the public to weigh in on. All of them set clear normative objectives. They cannot easily be answered be appealing to expert opinion. Yet these are clearly distinct from asking the public to decide who is healthier, someone who has cramping pain and a bloated feeling in the belly, or someone who has swollen legs with hard and thick skin, which causes difficulty in moving around.[44] The basic aims of democratic involvement can surely be accommodated without having to rely on the public to make these sorts of judgements. In fact, in empirical studies, community members generally do not indicate any inclination to do so. One analysis in the United Kingdom found that while most respondents want to be involved in healthcare decision-making at the system level, few indicate a desire to evaluate the health status of patients.[45]

If the implication of the democratic argument is that public priorities ought to be reflected in healthcare policy, I agree. Yet when researchers ask the public to judge health states, they are not asking them to set normative priorities. The public may well have a role to play in determining the overall level of funding, or deciding on patterns of distribution. But the practice of valuing health already presumes the normative objective of using cost-effectiveness analysis to allocate resources. Once these types of normative objectives are locked in, the demands of the democratic argument would appear to be met. There seems to be no reason to continue insisting on democratic procedures of decision-making to value health states. That is, unless of course doing so could improve the quality of outcomes.

This brings us to the second implication of the democratic argument. Even if public preference valuations of health states are not necessarily implied by democratic principles of representation, they may still be justified if harvesting public knowledge provides the best means of understanding and approximating patient quality of life. Yet this too seems doubtful. We generally do not defer to the opinions of those on matters for which they have no relevant experience. By definition, representative samples of the general public are likely to have had little, if any, personal contact with the diseases and disabilities they are being asked to evaluate. One or two sentence descriptions of health states seems highly unlikely to provide a suitable antidote for inexperience. If the objective is to promote positive health outcomes and minimize negative health outcomes, we ought to be tracking the reality of health. Failing to do so surely cannot improve the quality of outcomes. But why then should we assume that aggregating public preferences would meaningfully track the reality of health? As I discussed in the last section, we have good reason to believe that it doesn't.

While this line of thinking sheds considerable doubt on the usefulness of public preferences to value health, it is not yet an argument for rejecting preferences outright. It may still be worth considering the epistemic merits of democratic arguments for eliciting the preferences of experts and patients. In other words, instead of asking the public to decide which diseases and disabilities are better or worse than others, why not simply ask those with more relevant experience?

One group that immediately comes to mind is healthcare professionals and practitioners. Even if the public can't be counted on to reliably assess the effects of disease and disability, perhaps employing democratic procedures of decision-making that harness the collective knowledge of experts could do the trick. In the first study using DALYs to estimate

---

[44] These were lay descriptions employed by Salomon et al. (2012) to describe symptomatic intestinal nematode infections and symptomatic lymphatic filariasis (Appendix, p. 10).

[45] Litva et al. (2002).

disease burdens, the authors did, in fact, rely on a panel of experts to compare and contrast disability levels associated with different conditions.[46] However, this approach came under heavy fire in subsequent years.[47] Importantly, the relevant expertise in this context is not biology or chemistry. It is an understanding of the experience of living with disease. In this light, one could argue that because medical experts spend a great deal of time with patients, they are in a better position to judge their quality of life than the public. This seems reasonable. Nevertheless, two concerns persist.

For one, doctors are not ordinary people. They are highly educated, generally well paid, and perhaps most importantly, trained to view disease and disability through the lens of medicine. As noted by the economist Daniel Mont, medical experts are uniquely prone to consider "disability as a disorder to be cured rather than a functional status that needs to be accommodated."[48] This inherent bias could lead to an overestimation of the burden of certain diseases, or even an underestimation if the disability in question isn't particular visible to the medical community. In this context, conditions that affect mental and social health again seem particularly at risk of being overlooked or misunderstood. A number of investigations in recent years have also exposed deeply troubling breakdowns of understanding between physicians and patients. As summarized by the philosophers Ian James Kidd and Havi Carel: "Patients continue to voice epistemic concerns, most obviously through the vast body of pathographic literature – including online patient fora, blogs, and narratives – which consistently attest to persistent experiences of feeling ignored, marginalized, or epistemically excluded by health professionals. The UK Patients' Association, for instance, lists complaints about communication between patients and health professionals as a frequent complaint received by the association."[49] All of this suggests that while medical experts and healthcare practitioners may be in a better position to judge patient wellbeing than the public, there is still no guarantee that even their judgements would be free of bias or reliably capable of tracking the reality of patient experience.[50]

Second, most healthcare professionals also only have intimate knowledge of the diseases and disabilities within their particular areas of expertise. Few have extensive experience interacting with all manner of patients.[51] Stated preference techniques are again deliberately designed to pit one disease against another in order to elicit comparative judgements. It seems highly unlikely that healthcare professionals would be equipped to make meaningful comparisons for anywhere close to the number of diseases and disabilities required to inform resource allocation decisions. And even if we were able to overlook all of these concerns, expert opinion could still only be considered a proxy for patient experience. Even the most careful and considerate physicians would have to base their assessments of patient wellbeing largely on patient testimony. It is, in fact, patients, not medical doctors, who are truly the experts on the experience of living with their own conditions. So why not simply go straight to the source? Why not attempt to elicit stated preferences directly from patients themselves? Perhaps employing democratic processes that harness the dispersed knowledge of patients can improve the quality of outcomes in healthcare decision-making, even if the public and expert cannot. While patient preferences would likely be the most promising preference candidate, they too are likely to be unreliable.

First, numerous psychological studies have demonstrated that how we expect to feel in the future, and how we end up feeling, are rarely aligned.[52] We struggle to predict the valence and duration of our future emotional states and even fail to accurately predict which emotions we end up experiencing.[53] For the purpose of valuing health states, this is particularly troubling as many stated preference techniques explicitly require trade-offs to be made in terms of future circumstances. Both the Time-Trade Off (TTO) and Standard Gamble (SG) ask respondents to imagine how many years of healthy life they would be willing to give up in order to avoid living their remaining life years with a given disease or disability. For members of the public, or medical experts who have had little or no experience with the disease or disability in question,

---

[46] Murray et al. (1996).

[47] For notable critiques, see Mont (2007) and King & Bertino (2008).

[48] Mont (2007, p. 1660).

[49] Kidd & Carel (2016, p. 173).

[50] This hypothesis has also been borne out in empirical investigations. See Levak et al. (2004).

[51] The most likely group in this respect would probably be general practitioners (GPs). However, even among this group, it seems highly unlikely that GPs would have spent sufficient amounts of time interacting with anywhere close to the variety of patients required to make meaningful and informed comparative valuations of diseases and disabilities for cost-effectiveness analysis.

[52] For a relevant summary of empirical findings, see Wilson & Gilbert (2003).

[53] Wilson & Gilbert (2003).

this task seems next to impossible. Yet even for patients, empirical research has demonstrated that how patients expect to feel in the face of an event tends to be strongly influenced by their present health status and can often change as the disease progresses over time.[54] There is no guarantee that the patient preferences remain constant. Stated preferences relying on unrealistic future predictions therefore seem likely to produce biased results, even if they are elicited from patients themselves.

Second, patients are also unlikely to be experts on the experience of living with conditions other than their own. Once again, stated preference techniques explicitly require trade-offs and comparisons to be made between different states of health. Relying on patients with one condition to assess their quality of life, in relation to patients with another condition, may be just as likely to produce unreliable outcomes as deferring to the opinion of the general public or medical experts. Even asking patients to compare their present wellbeing to their imagined quality of life if they were perfectly healthy seems questionable. Individuals living with a disease or disability may be no more capable of reliably estimating their quality of life without their own illness as they would be able to imagine their quality of life with someone else's.

All of these considerations provide strong reasons to not only deprioritize the stated preferences of the public, but also to deprioritize stated preferences altogether. Here we considered two possible interpretations of the democratic argument. Neither justified the use of preferences to value health. First, the inclination to reflect public priorities in healthcare policy does not imply that the public should value health states. Second, aggregating public preferences also seems unlikely to track the reality of health. So much for democratic arguments in favor of public preference metrics. But what of expert or patient preference metrics? These too proved questionable. Concerns of cognitive bias and inexperience with diverse health conditions render both expert preferences and patient preferences uniquely unfit for the objectives of cost-effectiveness analysis. We seem to be left with little reason to rely on preferences at all as a means to allocate medical resources, at least on grounds of democracy.

It is worth noting that this analysis did not reject either dimension of the democratic argument outright. There may well be a role for the public to play in setting priorities in healthcare and harnessing dispersed knowledge may still be called for if doing so can improve the quality of outcomes. In the second half of the thesis, I will return to both dimensions in designing a new metric of patient experience. Nevertheless, before moving on, it is still worth addressing the remaining two arguments often put forth in favor of community preferences – the impartiality argument and the adaptation argument.

1.4 The impartiality argument

In their recommendations, the Second Panel on Cost-Effectiveness in Health and Medicine goes on to defend community preferences on Rawlsian grounds.[55] Quoting Rawls*,* the authors note that "parties situated behind a veil of ignorance ... do not know how the various alternatives will affect their particular case, and they are obliged to evaluate principles solely on the basis of general considerations."[56] Along these lines, it is often assumed that community members may be less prone to biased reasoning when evaluating health states than patients, as their impartial position situates them behind a veil of ignorance. On the other hand, "those who have experienced illness may be inclined to overstate their illness or disability, either because their experience of illness changes their valuation of health status or because of a desire to influence health policy in a manner that will result in personal benefit, for instance by having more healthcare resources allocated to their problem or condition."[57] Taking these two considerations together, the impartiality argument can be summarized as follows:

> IMPARTIALITY ARGUMENT: Because community members do not know what diseases or disabilities they may have in the future, the evaluation of health states they provide is more likely to be impartial than that provided by patients.

---

[54] Winter at al. (2009); Halpern & Arnold (2008).

[55] This argument has been discussed elsewhere in Gandjour (2010) and Daniels (1985).

[56] Rawls (1971, p. 136-137) quoted in Neumann et al (2016, p. 182).

[57] Neumann et al. (2016, p. 182-183).

It is worth noting that, if correct, this argument would seem to undermine the reliability of both patient preferences and patient experiences. If impartiality is the concern, patient valuations of health states may be ruled out altogether. However, the impartiality argument fails on three fronts. First, there is no reason to believe that patients would, or even could, knowingly mischaracterize the severity of their illness with the intention of channeling resources to their cause. Second, there is no reason to believe that members of the public are actually capable of giving truly impartial assessments of health states. Third, the implicit assumption that impartiality would even be a virtue in this context is misplaced. When it comes to judging the severity of disease, impartiality may be precisely what we should try to avoid.

First, as the authors themselves go on to acknowledge, there is absolutely no evidence to suggest that patients intentionally lie about their quality of life.[58] In fact, both preference and experience based survey instruments are deliberately designed to control for these sorts of influences by not explicitly informing participants what their responses will be used for. Respondents are simply asked to judge disability levels or assess their quality of life with little in the way of additional clarification.[59] Assuming that patients would be able to infer the distributive aims of the exercise and provide self-interested answers with the express purpose of influencing the societal allocation of medical resources seems like a stretch. If this is the motivation for using public preferences to value health states, it hardly seems justified.

Second, the notion that public preferences valuations are in fact impartial also seems questionable. Stated public preferences are likely to come furnished with all sorts of preexisting beliefs, desires, experiences, knowledge, biases, and judgements. There is no way to guarantee, much less assume that members of the public would or even could jettison these aspects of their characters to impartially judge the health status of others. The exercise of weighing health conditions is itself not really set up to reflect anything like the Rawlsian original position. Respondents are only asked to judge diseases and disabilities, nothing more or less. This creates somewhat of a dilemma. If community members are not asked to be mindful of their personal commitments, public evaluations of health states may be more likely to be influenced by cognitive and cultural bias. On the other hand, if researchers attempt to simulate survey environments that control for them, they may inadvertently create new problems somewhere else.

For example, perhaps at the beginning of each preference survey, community members could be asked to set aside any preexisting beliefs or biases that could potentially influence their responses. It is hard to imagine why or how this would succeed. Even the most well-intentioned survey designers and survey respondents are unlikely to overcome the powerful undercurrents of cognitive limitations and cultural influence. Debiasing techniques in which subjects are trained to overcome irrational thinking have also mostly proven ineffective, and often backfire.[60] Providing additional information, or challenging respondents to reconsider their views, or disclosing personal motivations, can provoke faulty reasoning somewhere else. Attempting to control for all potential sources of biases would quickly become a game of whack-a-mole. Debiasing attempts also require knowing which biases to target. Inevitably, cultural blind spots are bound to persist. When designing or responding to stated preference surveys, we may simply not know what we are missing. At least inasmuch as impartiality demands fair and unbiased reasoning that is unfettered by personal commitments and motivations, public preferences may be unable to rise to the occasion.

This brings us to the third flaw of the impartiality argument. In Rawls' thought experiment, because theoretical persons behind a veil of ignorance are unable to assess their position in society, their self-interest guides them towards allocations of resources that would be deemed fair to all members of the society. In this context, impartiality would appear to be a virtue. However, when it comes to judging the disability levels of health states, impartiality not only doesn't seem like a

---

[58] Neumann et al. (2016, p. 182-183).

[59] The precise wording used in the survey conducted by Salomon et al. (2012) to estimate disability weights was: "Now, we want to learn how people compare different health problems. A person's health may limit how well parts of his body or his mind works. As a result, some people are not able to do all of the things in life that others may do, and some people are more severely limited than others. I am going to ask you a series of questions about different health problems. In each question I will describe two different people to you. You should imagine that these two people have the same number of years left to live, and that they will experience the health problems that I describe for the rest of their lives. I will ask you to tell me which person you think is healthier overall, in terms of having fewer physical or mental limitations on what they can do in life. Some of the questions may be easy to answer, while others may be harder. There are no right or wrong answers to these questions. Instead, we are interested in finding out your personal views." (Appendix, p. 2).

[60] Case studies relevant to healthcare are discussed in Kahneman (2011), Levy (2014), and Caplan (2014).

virtue, it may even be a vice. If community members are deemed to be impartial because they have no experience or relevant knowledge of the diseases and disabilities they are being asked to consider, then the criterion of impartiality seems at best irrelevant, at worst harmful.

Imagine you've just opened a new restaurant and you ask me which dish to serve on opening night: a seafood spaghetti with lobster and parsley or a sweet potato curry with kale and coriander. I'm from out of town, although one day I may wish to visit. I have no knowledge of the local cuisine or any of the surrounding restaurants in the area. I am completely impartial as to which dish is served. Right away it seems unclear why you would then be asking me for my opinion in the first place. The fact that I am impartial does not seem to render me any more capable of making the right decision. Nevertheless, you don't feel comfortable making the decision on your own and I seem like a reasonable person, so you figure it can't hurt. I decide on the lobster. I've enjoyed lobster before and associate it with luxury. Fair enough. You order the relevant ingredients and prepare the dish for opening night. Unfortunately, in the town you've just opened your restaurant in, lobster is generally considered to be low class. Twice as many customers would have preferred the curry. Opening night is a failure.

What went wrong here? We seem to have confused impartiality for epistemic reliability. Impartiality did not render me a reliable source. Instead, my own inexperience rendered me uniquely incapable of being a reliable source. Instead of seeking my advice, you probably should have sought the opinions of local residents who were in a better position to know what I did not. In the arena of healthcare, for the purpose of judging which diseases and disabilities are better or worse than others, the fact that community members are impartial would also seem to provide little reason to give their preferences any more weight. Inasmuch as impartiality is rooted in inexperience, it seems like precisely the wrong sort of virtue to prioritize. This is, again, not to say that the public should play no role in determining how medical resources are funded or distributed. In setting explicit normative objectives, impartiality may be worth pursuing. Yet that is not what public valuations of health states set out accomplish. When it comes to valuing health, the fact that the public are presumed to be impartial does not render community members better adjudicators of disease and disability.

Unlike the democratic argument, in this case the impartiality argument does seem mostly unjustifiable. First, motivated reasoning on the part of patients would appear to be an illegitimate concern. Second, public preferences are still unlikely to be completely impartial inasmuch as they are influenced and determined by cognitive and cultural bias. Third, even if both of these concerns can be overlooked, the fact that community members do not know which diseases or disabilities they may develop in the future seems to undermine their ability to meaningfully assess the value of health in the first place.[61] However, before we can soundly and conclusively set them aside, public preferences ought to be evaluated in light of the third and perhaps most influential argument ushered in their defense: the adaptation argument.

1.5 The adaptation argument

At this point we seem to have identified no compelling reason why principles of democracy imply public valuations of health states, much less sufficient reason to assume that public preferences would, or even should, be impartial. Nevertheless, one may still be tempted to argue that public preferences are more suitable for decision-making purposes, given the well-documented phenomenon of adaptation to disease and disability. In this case, the argument for community preferences would not necessarily be that they are reliable, but rather that patient valuations are unreliable.[62] This argument can be summarized as follows:

---

[61] Throughout this section, I have also taken onboard the assumption that community members do not know if and to what extent they will need to rely on the healthcare system in the future. Yet this is of course debatable. It seems reasonable to imagine that members of the public would have at least some inclination of the diseases and disabilities they could likely be afflicted with, and therefore just as likely to be driven by self-motivated reasoning as patients themselves. Alternatively, it also seems possible they members of the public have friends or family members under the care of medical professionals and may therefore wish to provide motivated answers in the service of benefiting them. However, once again, as there is little evidence to suggest that survey respondents – patients or community members – actually respond to surveys in this way, I have set these concerns aside.

[62] In other words, while the democratic and impartiality arguments present positive claims in favor of community preferences, the adaptation argument puts forth a negative claim against patient experiences.

> ADAPTATION ARGUMENT: The fact that patients adapt to their conditions renders them incapable of reliably evaluating their own quality of life.

Adaptation broadly refers to the degree to which individuals get used to changing life circumstances over time. In the context of health, a wide body of empirical evidence has demonstrated that people with disabilities and chronic conditions do tend to adjust to their conditions. This is reflected both in measures of patient preferences – "adaptive preferences" – and in measures of patient experience – "hedonic adaptation."[63] As evidence of the former, many members of the deaf community do not consider deafness to be a disability, but rather a cultural identity. When asked, they indicate no preference to be cured or treated.[64] As evidence of the latter, patients with paraplegia also often report feeling just as satisfied with their lives as able-bodied counterparts.[65] In perhaps the most striking example, lottery winners were found to be only marginally happier than paralyzed accident victims two years after the fact.[66]

There has been much discussion about whether the issue of adaptation renders measures of subjective experience unsuitable for policymaking.[67] To many commentators, there seems to be something inherently wrong, or at least misleading, about these instruments if they are incapable of meaningfully distinguishing between a happy disabled person and a happy able-bodied person. However, while examples of positive adaptation to health states is widely discussed, it is worth noting that some patients, particularly those with mental and social disabilities, do not and in fact cannot adapt to their conditions.[68] As Daniel Kahneman and colleagues point out, the very notion of adaptation to depression seems almost nonsensical: "Because depression is itself a subjective state, it is difficult to interpret what adapting to depression could mean – that a person doesn't feel bad when he feels bad?"[69] It is also worth noting that not all patients with physical disabilities get used to them. Several studies have shown that adaptation to severe multiple sclerosis and degenerative disorders generally does not occur.[70] Put simply, while some patients do seem to adapt to their conditions, others do not. The concern that adaptation may lead to biased estimations of disease burdens is therefore not universal. The question is not whether we should trust the preferences or experiences of all patients, but rather whether we should trust those of patients who adapt to their conditions.

Underlying this question are two distinct but related concerns. The first is that people born with disabilities have no experience of a life without them, and are therefore unable to reliably imagine just how well off they could otherwise be. The second is that patients who adapt to their conditions are rendered psychologically incapable of forming rational preferences or assessing their own quality of life. Both are unfounded.

First, the concern that patients can't possibly imagine a life without their disability could surely cut in the other direction. If able-bodied members of the public have no experience living with physical disabilities, why shouldn't their preferences be mistrusted for the same reason? If the underlying worry is that most people simply do not have the experience or capacity to judge the quality of life of others in dramatically different life circumstances, this would seem to provide strong reasons to abandon the enterprise health valuations based on stated preferences altogether. This is of course largely in line with the position I have been advocating.

The second concern may be more worthy of consideration. The worry here is not that all individuals are incapable of rationally imagining their wellbeing in relation to other possible states of being, but rather that those with disabilities or chronic diseases are uniquely incapable given the adaptive nature of their conditions. Yet this too seems unwarranted.

---

[63] Discussions of both phenomena in the context of healthcare are provided by Mitchell (2018).
[64] Lane (2002).
[65] Damschroder et al. (2005).
[66] Brickman et al. (1978).
[67] Sen (1990); Nussbaum (2001).
[68] Dolan & Kahneman (2008); Menzel et al (2002).
[69] Kahneman et al. (1999, p. 311-312).
[70] Ibid.

At this juncture, it is worth bringing Miranda Fricker's theory of epistemic injustice into the frame. Fricker defines epistemic injustice as "a wrong done to someone specifically in their capacity as a knower."[71] She distinguishes two specific types: testimonial and hermeneutical. Testimonial injustice can occur when individual or group testimony is not given credibility due to their social position in society. For example, the police may be less inclined to believe the grievances of minority groups. Hermeneutical injustice can occur when some individuals or groups do not have the interpretive tools or conceptual capacities to understand and appreciate the experience of others. The classic example in this respect is suffering from sexual harassment in a culture that lacks a concept for it. As characterized by the bioethicist Jackie Leach Scully, both testimonial and hermeneutical injustice can be considered part of the broader phenomenon of "epistemic exclusion" in which "certain kinds of knowledge are refused admission into the general shared stock."[72]

In the health sector, patients are at high risk of experiencing both forms of injustice.[73] As noted by philosophers Ian James Kidd and Havi Carel, there are "...distinctive features of the social group of ill persons [that] render them vulnerable to forms of negative prejudices and stereotypes."[74] Disabled people in particular are frequently judged to be globally incompetent, even when they only experience a single and domain specific form of disability.[75] In most societies around the world, illness, disease, and disability are often explicitly or implicitly considered to reflect psychological frailty, vulnerability, lack of agency, or incapacitation. Disability (even by its very name) is generally considered to be *a priori* suboptimal.

However, while there are of course forms of disability for which epistemic unreliability is an inherent feature – e.g. brain injury, dementia, psychosis – there seems to be no medical or scientific basis to assume that most patients would be incapable of rationally assessing their own quality of life. Critics might contend that the phenomenon of adaptation itself ought to be taken as evidence of patients' psychological unreliability. Yet this would seem to beg the question. We cannot use the assumption that disability is suboptimal to undermine patients' assessments of their own wellbeing, since the fact that disability is suboptimal is precisely what is in question in the first place.[76] Once again, we generally do not assume that able-bodied individuals are unreliable judges of their own experience simply because they have never been disabled. Why then should we assume the reverse to be true? When patient self-reports are considered to be mistaken or misguided simply because their conditions are stigmatized or misunderstood, concerns of testimonial injustice seem paramount to consider. For the purposes of cost-effectiveness analysis, the fact that we are confused or surprised by patient assessments of their own wellbeing does not provide strong moral reasons to disregard their testimony as invalid.

All the while, asking community members to judge the severity of health states for which they have no direct experience could also provide fertile ground for hermeneutical injustice. Again, if members of the public have had direct experience with the mental, social, and physical disabilities they are asked to consider, how then can they be expected to meaningfully judge or interpret the experience of those who do? In the language of epistemic justice, community members are not equipped with the hermeneutical tools to fully appreciate and understand the experiences of the patients' lives they are being asked to evaluate. In turn, prioritizing community preferences over patient self-reports can serve to reinforce an institutional approach to disability focused on medical intervention and alleviation as opposed to societal understanding and accommodation.[77] As the philosopher Laura Cupples powerfully summarizes:

> "Our social scripts tell us disabled people are both dependent and incompetent, and this makes it difficult for us to imagine them living independently, being valuable members of the workforce, or effectively parenting their own children. If we are to pursue epistemic and social justice for disabled people and people with chronic illnesses—

---

[71] Fricker (2008, p. 9).

[72] Scully (2018).

[73] Carel & Kidd (2014).

[74] Kidd & Carel (2017, p. 176).

[75] Scully (2018); Toombs (1995).

[76] Barnes (2009).

[77] It is this final step that can convert hermeneutical disadvantage into hermeneutical injustice. Patients may be hermeneutically disadvantaged without experiencing any specific social distortion of power, which Fricker defines as the key criterion of epistemic injustice (2008, p. 151-152). In the present context, relying on public valuations of health states surely puts patients at a hermeneutical disadvantage, but can also create a social distortion by reinforcing a medical (mis)understanding of disease and disability. However, the distinction between hermeneutical injustice and hermeneutical disadvantage is often debated. For relevant discussions, see Kidd et al. (2017, p. 176-178).

> including the right to live and work in the community, and the right to build their own families—we must work to expand our social imaginary and to write new social and cultural scripts surrounding the experience of disability. We must engage with disabled people as knowers and give the hermeneutic resources they develop to explain their own experiences proper uptake."[78]

In other words, it may not be patients who misunderstand their own quality of life, but we who misunderstand what quality of life really is. In this respect, granting credibility to the stated preferences of community members while disregarding the experiences of patients seems ethically unacceptable. Given the potential for both testimonial and hermeneutical injustice, we ought not demonstrate a bias towards accepting the stated preferences of community members, but rather towards accepting the self-reports of patients.

1.6 Conclusion

All of these arguments strongly suggest that estimations of disease burdens using QALYs or DALYs are likely to be descriptively and normatively suboptimal, as they are inherently reliant on community preferences and a relatively narrow conception of health and wellbeing. The ultimate implication of using them to make resource allocation decisions is that diseases and symptoms that the public cannot easily understand may not become adequately funded or sufficiently understood, precisely because no one knows how to properly evaluate the potential impact of investing in efforts to address them.

The analysis presented thus far would also seem to leave us without options from the perspective of preference satisfaction. If neither public, nor expert, not even patient preferences can serve as justifiable metrics for healthcare rationing, we must either abandon the project of cost-effectiveness altogether or look for more suitable metrics elsewhere. In the remainder of this thesis, I will take the latter perspective and develop an approach to cost-effectiveness rooted in metrics of patient experience.

In developing this account, it is worth reflecting on the key lessons and takeaways to be learned from the analysis presented thus far. First, we identified no strong reason to reject the democratic argument altogether. Any justifiable approach to cost-effectiveness analysis should likely make room for public priorities to be reflected in the decision-making process. Even if the public is not equipped to value health states, the wholesale discounting of public priorities does not seem warranted. Second, using democratic procedures of decision-making to harness collective knowledge may be justified if doing so can improve the quality of outcomes. The fact that preferences are generally unreliable does not imply that other forms of dispersed knowledge have nothing to contribute. Third, while impartiality does not necessarily seem like a virtue in valuing health, we may still wish to limit opportunities for motivated reasoning. Any approach to cost-effectiveness analysis that can be easily gamed or taken advantage of may not be likely to succeed. Fourth, although adaptation appears to be less of a concern than it is made out to be, we should still be on the lookout for other types of cognitive and cultural bias. Faulty reasoning and cultural blind spots could still undermine the reliability of health state valuations if they are left entirely unchecked. Fifth, and perhaps most importantly, we should not defer to those on matters for which they have no relevant experience. This last point served to undercut many, if not most, of the arguments put forth in defense of preference valuations of health states. It may also prove particularly difficult to solve. For the enterprise of cost-effectiveness analysis to function properly, we need to be able to make comparisons between an enormous array of diseases and disabilities. Yet neither the public, nor medical experts, nor even patients themselves seem qualified to make these comparisons for us. We need to devise another system of comparative evaluation that can avoid these pitfalls. In the next section, I will lay out a patient-centered framework for cost-effectiveness analysis in healthcare that strives to meet these goals.

---

[78] Cupples (2020, p. 11).

Part II: Valuing experience

In this latter half of the thesis, I will develop an approach to healthcare rationing rooted in patient experience. The broad strokes of the argument presented thus far can be summarized as follows:

(1) Cost-effectiveness metrics in healthcare should reflect the value of health
(2) The value of health can be measured in terms its impact on wellbeing
(3) Patient experience is a better proxy for wellbeing than public preferences
(4) Cost-effectiveness metrics in healthcare should reflect patient experience

This discussion has mostly focused on the third premise. In the first part, I argued that public preferences are ill-poised to serve as a metric for allocating resources inasmuch as they fail to meaningfully reflect or represent any justifiable notion of quality of life. However, in doing so, I have mostly shown that public preference metrics are flawed. I have not yet shown that patient experience metrics can do any better. This task will occupy the remainder of the thesis.

To this end, it is worth reiterating the key takeaways that emerged from the first half. If our objective is to develop an approach to cost-effectiveness analysis based on patient experience, there are at least five criteria it should strive to meet. First, it should carve out some space for public involvement. Second, it should harness the dispersed knowledge of citizens if doing so can meaningfully improve the quality outcomes. Third, it should limit opportunities for motivated reasoning. Fourth, it should minimize cognitive and cultural bias.[79] Fifth, it should be informed by reliable sources. However, it does not seem sufficient to suggest that an approach to cost-effectiveness analysis based on patient experience would be theoretically justifiable. It must also be practically feasible. Bridging the gap between theory and practice is often messy. Even if patient experience metrics are normatively defensible, the specifics of their implementation could render them practically untenable. To this point, we may wish to add a sixth criterion: feasibility. Whatever our ethical justifications for metrics of patient experience may be, they should not get lost in translation once we attempt to implement them in the real-world.

In the sections that follow, I will sketch out a patient-centered approach to cost-effectiveness analysis that strives to meet these demands. The discussion will be primarily concerned with the benefits and drawbacks of using patient experience metrics to inform healthcare policy. From this perspective, classic objections to subjective interpretations of wellbeing involving brains in vats and experience machines have reduced force.[80] The objective will not be to establish subjective experience as the only or even most reliable indicator of wellbeing in any and all circumstances, but rather to evaluate its potential as a practically useful and ethically justifiable means by which to allocate medical resources, one that has comparative advantages over prevailing approaches.

I will build this account from the bottom-up, beginning with feasibility. Since any approach to cost-effectiveness that is not feasible would be dead in the water, feasibility seems like a good place to start. In Section 2.1, I will therefore provide an overview of the measures of subjective wellbeing currently on offer in economics and psychology and evaluate their potential to serve as measures of patient experience in the health sector. I will argue that life evaluations are uniquely well-suited to the aims of cost-effectiveness analysis in that they are easy to implement and meaningfully reflective of patient experience. In Section 2.2, I will use these insights to introduce a new measure of patient experience, one that is capable of harnessing the dispersed knowledge of citizens and making comparative evaluations between diseases and disabilities without relying on faulty reasoning or unreliable testimony. Finally, in Section 2.3, I will articulate two underspecified distributive dimensions of this approach that may be informed by public priorities. The thesis will conclude with several recommendations and avenues for further research.

---

[79] Remember again here, adaptation is not necessarily the concern. Rather, I have in mind the sorts of focusing illusions and forecasting errors that can sway public preference valuations of health states as discussed in Section 1.2.

[80] Arguments to this effect are presented in Nozick (1974) and Parfit (1984).

2.1 Subjective wellbeing

In this section, I will introduce the most commonly used measures of subjective wellbeing in economics and psychology and assess their feasibility for the purposes of cost-effectiveness in healthcare. This discussion of feasibility will mostly be centered around two main themes. I will assess the degree to which subjective wellbeing measures can be practically implemented for healthcare policymaking and meaningfully reflective of patient experience. Ultimately, I will argue that life evaluation measures have comparative advantages on both accounts. However, they are also not necessarily completely immune to bias. I will address possible sources of cognitive and cultural bias as well as possible remedies.

Subjective wellbeing – or simply "happiness" – broadly refers to an understanding of quality of life that places a high value on first-person experience. Empirical studies on subjective wellbeing typically take the form: Wellbeing = r(h) where wellbeing is represented as some reporting function (r) of underlying true happiness (h).[81] In practice, measures of subjective wellbeing generally fall into one of two categories: affect or evaluation. Affect refers to the day-to-day experience positive or negative emotions, while life evaluation refers to a global assessment of life as a whole.[82]

The model presented in this half of the thesis is motivated by the need to develop metrics of patient experience for resource allocation decisions in the health sector. For this purpose, life evaluations seem uniquely well-suited for three main reasons. They are easy to work with, reflect patient values, and offer a holistic understanding of quality of life. In what follows, I will address each point individually by evaluating the most common measures of affect and life evaluation as employed in empirical research and comparing their relative benefits and limitations.

Measures of affect are generally concerned with monitoring the quality of emotional states, or simply mood. For example, the Positive and Negative Affect Schedule (PANAS) asks respondents to report the extent to which they felt each of the following ways in the previous week: interested, distressed, excited, upset, strong, guilty, scared, hostile, enthusiastic, proud, irritable, alert, ashamed, inspired, nervous, determined, attentive, jittery, active, and afraid.[83] Responses to each question are coded on a five-point scale from "not at all" to "extremely" and then aggregated to provide an overall indication of positive or negative mood. However, because emotional experience is expected to be highly variable, isolated measurements are unlikely to be reliable indicators of underlying subjective wellbeing. The psychologist Daniel Kahneman and economist Paul Dolan therefore advocate the use of experience sampling methods (ESM) in which respondents' moods are repeatedly monitored through the day, or daily reconstruction methods (DRM) in which respondents are asked to review their emotional experiences each night before they go to sleep.[84]

However, for the purposes of cost-effectiveness analysis, this is bound to present practical difficulties. Regularly asking patients to complete long questionnaires on a daily basis would seem like an enormously costly undertaking. While clinical trials may be perfectly suitable settings for these types of assessments, using them to make comparisons of disease burdens between countries, over time, and across patient groups would likely be ruled out, at least in the short term. The scope of an approach to cost-effectiveness analysis rooted in affect-based measures of experience would therefore likely be quite limited.

Life evaluations, on the other hand, are generally easier to work with. For one, they tend to be a lot shorter. One of the most widely used measures in this respect, the Cantril Ladder, presents respondents with the following prompt: "Think of life as a ladder. Suppose the top of the ladder represents the best possible life for you and the bottom of the ladder the worst possible life. Where on the ladder do you feel you stand at the present time?"[85] Responses are coded on an 11-point scale. Another single-item measure, life satisfaction, asks respondents to report how satisfied they are with their lives on a

---

[81] Dolan et al. (2008).

[82] Following Haybron (2020), in this discussion I group measures of life satisfaction and life meaning (eudemonia) under the heading of life evaluations as both are reflective evaluative judgements of life as a whole. However, they are often teased apart in empirical research (Busseri & Sadava, 2011).

[83] Crawford & Henry (2008).

[84] Dolan and Kahneman (2008).

[85] Cantril (1965).

scale from 0 to 10, with 0 being completely dissatisfied and 10 being completely satisfied.[86] Over the last three decades, their simplicity has allowed life evaluations to be used in over 160 countries by the Gallup World Poll, United States Social Survey, UK Office for National Statistics, European Social Survey, Eurobarometer, World Values Survey, and Pew Global Attitudes Survey, among others. Unlike affect, life evaluations also tend to remain mostly stable over time, limiting the need for ongoing and repeated administration.[87] For the purposes of healthcare decision-making, using them to inform cost-effectiveness analysis is therefore poised to be a much more feasible endeavor.

Even more importantly, life evaluations seem capable of offering more meaningful understandings of patient experience than affect. I take "meaningful" here to be an important dimension of feasibility. If we advocate for patient experience metrics in theory but only end up measuring how often patients' legs fall asleep, we would hardly be justified in using them as a means by which to allocate medical resources. It matters not only that our measures of experience are reliable, but also that we are measuring the right things. In this respect, life evaluations seem preferable to affect for two reasons. They are more reflective of patient values and offer a more holistic understanding of quality of life.[88]

While affect measures can tell us about the experienced emotional character of patients' lives, they cannot provide insight into which emotions the person experiencing them actually considers to be valuable. It would be foolish to suggest that patients should be having constantly pleasant emotional experiences. As the philosopher Daniel Hausman points out, "If all that matters is mood, the treatment of choice for nonfatal illnesses would be valium laced with opium."[89] We need to have some meaningful way of distinguishing between emotional states that are more or less beneficial than others. However, if measures of affect were given the central role in healthcare policymaking, it is unclear how this could be accomplished. How much positive affect would new treatments need to create to be deemed cost-effective?

Life evaluations may be able offer a solution. Using life evaluations as measures of patient experience can allow us to determine the extent to which different emotional states actually benefit the person experiencing them. Someone addicted to opioids may exhibit high degrees of positive affect, but profoundly low levels of life satisfaction.[90] The utility or disutility of emotional experiences is likely to be cashed out in terms of how patients evaluate their lives.[91] Life evaluations can therefore also offer healthcare decision-makers a way to understand patient wellbeing without having to adopt one particular definition of the good life for everyone or rely on unreliable sources to define it for them. As long as we are committed to the assumption that patients would judge their lives to be going well by their own standards if they really were in fact going well, life evaluations can offer important insights into the causes and contents of their wellbeing. As noted by the philosopher Daniel Haybron, "The great attraction of life satisfaction is that it doesn't presume that only pleasure ultimately matters for you, and it doesn't reduce life to nothing more than a summation of moments, with no regard to the bigger picture. If you regard your life as going well for you, by whatever standard you deem fitting, then life satisfaction will seemingly reflect that."[92] This also promises to avoid fundamental concerns of epistemic injustice associated with public preference valuations. Affording patients the opportunity to decide for themselves what makes life worth living, and in turn using these judgements as the basis upon which to allocate medical resources, can ensure that energy and investment is directed towards the most important determinants of patient quality of life as determined by those most qualified to decide – patients themselves.

At this point, life evaluations seem preferable to affect inasmuch as they are easier to use, reflective of patient values, and offer more holistic representations of patient wellbeing. Yet this last point may still need some additional clarification. What reason do we have to believe that life evaluations actually measure what they intend to? Are they really capturing

---

[86] Annex A in OECD (2013).

[87] Diener et al. (2013).

[88] Raibley (2010.).

[89] Hausman (2015, p. 118)

[90] In empirical research, drug addiction generally has profoundly negative effects on subjective life evaluations. See Luty & Arokiadass, (2008); De Maeyer et al. (2010); Binder & Coad (2013).

[91] In practice, life evaluations and affect do generally tend to correlate well with each other, though not always. For a thorough discussion, see Margolis et al. (2020).

[92] Haybron (2016).

overall judgements of individual wellbeing or something else entirely? At least five strands of evidence are worth considering in this respect. First, life evaluations tend to correlate with third party reports. Studies in which family members, friends, and professional interviewers are asked to predict the life evaluations of others generally align with individual self-reports.[93] Second, life evaluations also reflect known group differences. People with low levels of income, poor health, few social connections, and limited economic opportunities generally evaluate their lives to be worse than those in more fortunate life circumstances.[94] Third, on a national level, life evaluations reflect objective conditions. Countries with low levels of corruption and conflict, strong civil liberties and political rights, and high degrees of social capital consistently score higher than those without them.[95] Fourth, life evaluation scores associate with other important psychological markers including optimism, loneliness, and trust.[96] Fifth, life evaluations predict future behaviors including school retention, voting activity, and suicide rates.[97] All of this suggests that life evaluations are not only reliable, but that they are valid and holistic indicators of overall experience.

Life evaluations perform so well because they are capable of reflecting what matters most to people without having to ask them directly. This is one of the most important advantages of using experience measures to conduct health state valuations as opposed to preference measures.[98] When asked to imagine how we might feel in any number of circumstances, an array of cognitive and cultural biases may ultimately lead us to misunderstand, misremember, or misrepresent crucial realities of lived experience. When asked to review how our own lives are going by our own standards, we tend to perform much better.[99]

However, this does not mean that life evaluations are completely immune to bias. On an individual level, some studies have indicated that they can be influenced by unrelated situational factors including the time of day, the day of the week, and weather conditions at the time of judgement.[100] On an international level, cross-country comparisons of wellbeing may also be subject to cultural differences. For example, members of more collectivist societies seem less likely to report low levels of subjective wellbeing than members of individualistic societies since the former are more likely to consider unhappiness socially undesirable.[101] In the present context, there are two related lessons to take away from these observations. First, health state valuations for cost-effectiveness analysis should be based on as many patient self-reports as possible. Across large enough samples, minor situational influences are largely expected to cancel each other out. Second, when evaluating the influence of health on wellbeing, it is also important to control for cross-country cultural differences that could potentially bias responses. These issues will be discussed in greater detail in the next section.

In this section, I considered feasibility in terms of two dimensions: practicality and meaningfulness. I argued that life evaluations met both standards for three reasons. They are easy to use, reflective of patient values, and holistically representative of patient experience. This is not to say that life evaluations are capable of representing everything there is to know or care about wellbeing. Instead, they are intended here merely as useful indicators of patient experience that have unique comparative advantages over other approaches. However, for the purposes of allocating medical resources, simply comparing average life evaluations between patient groups does not seem sufficient. Here we are interested not only in overall differences in wellbeing, but also the extent to which those differences can be attributed to differences in health. In the next section, I will therefore expand on this analysis to introduce a new methodology that is both informed by patient experience and designed to suit to the needs of cost-effectiveness analysis in healthcare.

---

[93] Diener et al. (2013).
[94] Dolan et al. (2008); Oswald (2010).
[95] Helliwell et al. (2020).
[96] Dolan et al. (2008); Oswald (2010); Lykken & Tellegen (1996).
[97] Diener et al. (2013).
[98] Bronsteen et al. (2013).
[99] Wilson & Gilbert (2003).
[100] Diener et al. (2013); Schwarz et al., 1987. However, there is some debate as to the replicability of these findings (Hudson et al. 2020).
[101] Lau et al. (2005).

2.2 Wellbeing Adjusted Life Years

In this section, I will conceptualize patient experience in terms of Wellbeing Adjusted Life Years (WALYs).[102] This discussion will proceed in two steps. First, I will lay out the methodology to calculate WALYs for the purposes of cost-effectiveness analysis in healthcare. Second, I will evaluate its performance against the six normative criteria introduced earlier on. I will conclude by commenting on two distributive implications of this approach in the next section, and discuss how they may be informed and developed in future research.

Like QALYs and DALYs, WALYs attempt to represent two properties of human life in one unit: quality and time. However, unlike QALYs and DALYs, WALYs consider quality of life in terms of patient experiences as opposed to public preferences. The number of WALYs experienced by an individual in one year may range from 0 to 1, with one WALY representing one year in full wellbeing. This initial framework gives rise to the following two formulas:

(1) $\text{WALYs}_{\text{experienced}} = \text{actual wellbeing} / \text{potential wellbeing}$
(2) $\text{WALYs}_{\text{lost}} = 1 - \text{actual wellbeing} / \text{potential wellbeing}$

As I will demonstrate, these formulas can be applied at both individual and societal levels. WALYs can be used to understand the amount of wellbeing experienced or lost by one patient or within a given patient population.

Actual wellbeing in this context can be understood in terms of average life evaluations of the patient group. However, potential wellbeing is slightly trickier. For reasons discussed in the last chapter, asking patients how well-off they think they might without their conditions would produce biased results.[103] We also cannot simply assume that patients with disabilities or chronic diseases would be perfectly happy without them. There is more to life than health. Any reasonable approach to cost-effectiveness analysis would have to account for the extent to which high or low levels of wellbeing can actually be attributable to the disease or disability in question.

In fact, precisely estimating the effect of health on patient wellbeing would require observing the same person in two states of the world, healthy and unhealthy. This is obviously impossible. However, we may still be able to approximate this difference by other means. For this purpose, researchers generally have two tools at their disposal. They can run experiments or analyse data. In the first instance, by randomly allocating individuals into two groups and exposing one to an intervention but not the other, researchers can estimate the causal effect of the intervention. In the present context, this would obviously be wildly unethical. We cannot deliberately infect people with disease or disable them. Instead, we have to analyse data. Fortunately, large-scale anonymized datasets containing detailed information on subjective wellbeing for thousands of respondents are routinely collected by governments, private institutions, and public universities around the world.[104] By comparing the life evaluations of patients to the life evaluations of healthy counterparts who are similar to them in all other relevant respects – gender, age, education, employment, marital status, income, country, etc. – we can isolate differences in wellbeing between both groups that can be attributed to differences in health.[105] This approach is designed to mimic experimental designs without the need for deliberate intervention.[106] In the present context, since we are considering actual wellbeing in terms of the average patient life evaluations, we can then consider potential wellbeing in terms of average life evaluations of healthy counterparts.

To bring this model more into focus, consider an illustrative example of diabetes and Parkinson's. In 2017, European adults with diabetes reported average life satisfaction scores of 7.1 out of 10, while Parkinson's patients reported average

---

[102] An initial concept for Wellbeing Adjusted Life Years was presented in Happiness Research Institute and Leaps by Bayer (2020). This approach also builds off earlier work by Ruud Veenhoven (2005) as well as Richard Layard and colleagues (2020). However, while all of these tools aim for a more global representation of wellbeing for the purposes public policymaking, the scope of the present discussion is limited to its usefulness as a metric of patient experience for the purposes of cost-effectiveness analysis in healthcare.

[103] See Section 1.3.

[104] For a thorough review of relevant datasets on subjective wellbeing and health, see Powdthavee (2015).

[105] In practice, this is often accomplished with matching designs in which researchers detect and control for all differences in subjective wellbeing levels between patients and healthy counterparts except for those that may be attributable to disease and disability (Binder & Coad, 2013; Oakes & Johnson, 2006).

[106] Binder & Coad, 2013; Oakes & Johnson, 2006).

life satisfaction scores of 6.6.[107] Diabetes patients were also on average 0.3 points less satisfied with their lives than healthy counterparts, while this difference was 0.6 points for Parkinson's patients.[108] These inputs can then be used to estimate the amount of wellbeing lost due to disease for both groups of patients. WALYs lost due to Parkinson's is given by $1 - \frac{6.6}{7.2} = 0.09$, while WALYs lost due to diabetes is given by $1 - \frac{7.1}{7.4} = 0.04$. This would be equivalent to saying that Parkinson's patients lose 9% of the potential wellbeing they could have otherwise experienced, while diabetes patients lost 4%.

Before diving into the normative implications of this approach, it is also worth considering how these formulas could be used to assess wellbeing burdens of disease on a population level. According to data provided by the Institute for Health Metrics and Evaluation, in 2017 there were 1.4 million European adults living Parkinson's disease and 22.9 million with diabetes.[109] By multiplying average individual WALY losses by the total number of European adults with each condition, we then find that 126,000 WALYs were lost due to Parkinson's in 2017, while 916,000 WALYs were lost due to diabetes. In turn, these sorts of estimates could be used to evaluate the cost-effectiveness of new treatments, interventions, and medications in terms of the cost per WALY gained.

With this framework in mind, we can now evaluate its performance against the six normative criteria introduced earlier on. I will take them here in reverse order. First, any approach to cost-effectiveness analysis must be feasible. In the last section, I considered feasibility in terms of practicality and meaningfulness. Life evaluations met both standards in that they are easy to use, reflective of patient values, and holistically representative of quality of life. Since life evaluations are used as the currency of value to calculate WALYs, these benefits carry over to WALYs as well.

Second, the model must be informed by reliable sources. Many of the challenges associated with public preference metrics was their deference to community members on matters for which they had no relevant experience. WALYs are designed to avoid this pitfall. In the present context, it is patients, not members of the public or medical experts, who are relied upon to evaluate their own experiences. By allowing patients to decide for themselves what makes life worth living, WALYs are poised avoid concerns of testimonial and hermeneutical injustice that beleaguered public preference metrics.

Third, we should limit cognitive and cultural bias. In this case, by comparing patient life evaluations to the life evaluations of healthy counterparts, we are now also able to infer the underlying effects of disease and disability without having to rely on questionable stated preference comparisons. This promises to avoid many of the biases that can influence QALYs and DALYs. However, life evaluations are still not completely immune to error. To avoid cognitive bias, the methodology outlined in this section recommended the use of large-scale datasets to estimate average levels of wellbeing across as many patients and healthy counterparts as possible. This approach is less vulnerable to random situational influences as they are expected to cancel each other out across large enough samples. To avoid cultural bias, this section also recommended that patients be compared to healthy counterparts who are similar to them in all relevant respects. This would surely include country. For example, it would be inappropriate to estimate the effect of diabetes by comparing patients in Austria to healthy people in China. Patients should only be compared with healthy counterparts who live in the same country as them as to control for any relevant cross-cultural differences.[110]

Fourth, we should limit opportunities for motivated reasoning. While this criterion seems less urgent than the others, it may still be worth accounting for. In this context, patients and healthy counterparts should again only be asked to report on their quality of life without being informed of the distributive aims of the exercise. This is generally standard practice

---

[107] Data drawn from the Survey of Health, Aging, and Retirement in Europe (SHARE). For more details regarding SHARE data and methodology, visit *www.share-project.org*. Data only representative for adults over the age of 50.

[108] Author's estimations using nearest-neighbor propensity score matching with covariates for gender, age, residential area, income, wealth, education, employment, marital status, number of children, country, and year. The model compared 2,343 Parkinson's patients and 36,882 diabetes patients to 2,935 and 36,943 healthy counterparts, respectively. Differences were found to be significant at a confidence level of 99.9%. Life satisfaction measured on a scale from 0 to 10.

[109] For more information, visit www.healthdata.org.

[110] This is also standard practice in empirical happiness research. For a review of relevant best practices and recommendations, see OECD (2013).

in subjective wellbeing research.[111] However, this is not to say that patients or the public should have no say in how resources are distributed. I will address this issue in detail in the next section.

Fifth, cost-effectiveness analysis should employ democratic procedures of decision-making if doing so can improve the quality of outcomes. This was one of the two most important implications of the democratic argument for public preferences. While I argued that accumulating preference valuations of health states was not in fact capable of improving the quality of outcomes, I found nothing particularly objectionable with democratic epistemology. In this case, WALYs employ democratic processes of reasoning in two relevant respects. First, democratic procedures of decision-making are employed to harness the collective knowledge of patients. However, in this case, the relevant knowledge we are accumulating is not preferences, but experiences. Since patients are the experts on their own experiences, gathering as much data on these experiences as possible can only serve to improve the quality of outcomes. Second, to estimate the effects of disease and disability, we are now also harnessing the collective knowledge of healthy counterparts. This allows us to infer potential patient wellbeing without having to rely on any individual or centralized authority.[112] Even though we abandoned measures of preference for measures of experience, we need not abandon the enterprise of democratic epistemology. The approach to cost-effectiveness analysis I have outlined in this section is in fact deliberately designed to harness the collective knowledge of citizens. Informing the model with the experienced knowledge of patients and healthy counterparts can only serve to improve our understanding of health of disease, which can in turn help to improve the quality of outcomes.

Sixth, cost-effectiveness analysis should reflect public priorities. This was the first democratic defence we considered in favour of public preference metrics. There I argued that democratic principles of representation need not imply public preference valuations of health states. Nevertheless, rejecting wholesale the idea that public priorities should be reflected in the policymaking process may be too quick. An approach to medical resource allocation that leaves no room for public input would seem unwarranted. At least as long as the public foots the bill, they ought to have some say in determining how health policy is carried out. Yet if valuations of health are based on metrics of patient experience, it remains unclear if and to what extent the public can contribute. The fact that healthy counterparts are surveyed about their experiences seems unsatisfactory as a benchmark of democratic representation. The public deserves more to do. In the next section, I will therefore develop a tentative framework for evaluating the appropriateness of public involvement in healthcare decision-making and apply it to two potential case studies of distribution.

2.3 Preferences for distributions

Thus far, this thesis has been primarily concerned with metrics of resource distribution in the health sector. I have mostly focused on which goods to care about from the perspective of cost-effectiveness analysis and how best to measure them. However, patterns of resource distribution can have equally important normative consequences. While any metric can theoretically be paired with any pattern, these relationships can be more or less fitting depending on how the metric is designed and applied in practice. Although a complete analysis of distributional patterns in healthcare is well beyond the scope of this thesis, in this section I will briefly comment on two distributive implications of using WALYs as a measure of cost-effectiveness in healthcare – equity and aggregation – and discuss how public preferences could be leveraged to inform them.[113]

However, it may first be worth clarifying why public preferences ought to be considered relevant in this context to begin with. In the original discussion of the democratic argument, I made a distinction between setting the normative objectives of healthcare policy and implementing them.[114] I made the tentative claim that public preferences should inform the former, but not necessarily the latter. The practice of valuing health states itself presupposes the normative objectives of cost-effectiveness analysis. While the normative objective of promoting positive health outcomes and minimizing negative

---

[111] OECD (2013).

[112] That is, except statisticians.

[113] For lengthier discussions on both matters, see Sassi et al. (2001) and Daniels (1985).

[114] Section 1.3.

health outcomes could very well be up for public debate, deciding who is or is not healthy should not be. Along similar lines, I consider distributive aims in this context to be normative objectives. While letting the public weigh in on whether or not disadvantaged groups should be given priority may be justified, relying on public preferences to decide who is and is not disadvantaged may not be. The argument here is not that the public should always or necessarily set normative objectives in healthcare. Depending on the context, there may be good reason to be skeptical of public involvement. Nevertheless, I intend to take seriously the impetus for democratic representation in publicly funded healthcare systems and offer the two following case studies as potential opportunities for public influence.

An old an unsettled debate in the literature on cost-effectiveness analysis concerns the trade-off between efficiency – understood as health maximization – and equity – understood as health equality.[115] It is often argued that the two are misaligned. Maximizing health benefits can reinforce health inequalities. While this trade-off is primarily one of distributive priorities, cost-effectiveness metrics can also have implicit equity assumptions baked in. WALYs are no exception. Because WALYs are calculated using the ratio of actual to potential wellbeing, gains at the lower end of the spectrum are inevitably weighted more than gains at the top. For example, a one point increase in wellbeing from 4 to 5 would represent a gain of $1 - \frac{4}{5} = 0.2$ WALYs, while a one point increase in wellbeing from 9 to 10 would only represent a gain of $1 - \frac{9}{10} = 0.1$ WALYs. By design, WALYs will therefore always attach more weight to helping those who are worse off. This a sharp departure from QALYs and DALYs. Both measures weigh gains delivered at the top the spectrum equally to gains delivered at the bottom. Interventions that raise patient quality of life from 0.3 QALYs to 0.4 QALYs and those that raise patient quality of life from 0.8 QALYs to 0.9 QALYs are deemed equally effective. A broken leg is therefore considered to have the same impact on quality of life whether you are poor or rich, young or old, married or single, with or without children. As the saying goes, "a QALY is a QALY is a QALY.[116]

These design differences can have important distributive consequences when implemented in practice. However, while equity concerns have traditionally been left unaccounted for in cost-effectiveness analysis using QALYs and DALYs, they generally find high levels of public support.[117] When asked, most people consider patients in more dire health to be more deserving of medical resources those who are relatively better off.[118] Several studies have also found that delivering health benefits to the least privileged members of society tends to be viewed as more important than maximizing overall health.[119] Although WALYs seem better equipped to handle these types of concerns than QALYs or DALYs, there may still be some room for additional clarification. For example, empirical research could seek to investigate the extent to which the equity weights built into WALYs actually align with public priorities. If two societal groups experience the same disease, but one group reports life evaluations that are on average twice as high as the other, is it really appropriate to consider treatment for the former twice as valuable as treatment for the latter? Does the calculus change if this background difference in wellbeing is attributable to poverty or addiction? Philosophers may have diverging opinions. WALYs make their own assumptions. But in taxpayer funded healthcare systems, taking stock of public opinion may also be warranted. All of these questions set explicit normative priorities. Public preferences may have an important role to play in answering them.

Another important distributive implication of using WALYs as a measure of cost-effectiveness analysis is the degree to which wellbeing burdens ought to be aggregated across individuals. In the earlier example, while Parkinson's was deemed to be significantly more burdensome on an individual level, it was dwarfed by diabetes on a population level given the stark divergence in prevalence between both diseases. Aggregation is in fact likely to be an inevitable outgrowth of any approach to cost-effectiveness analysis. While some degree of aggregation is generally considered to be intuitive – e.g. all else being equal, curing disease in one hundred patients seems better than curing disease in ten patients – profound disagreement exists as to its limits.[120] Curing enough headaches may seem like it should never outweigh the benefits of

---

[115] Sassi et al. (2001); Wagstaff (1991).

[116] Weinstein, (1988).

[117] Sassi et al. (2001).

[118] Ubel et al. (1999); Gyrd-Hansen (2004).

[119] Dolan, Cookson & Ferguson (1999); Abásolo & Tsuchiya (2004); Lindholm & Rosen (1998).

[120] For discussions, see Voorhoeve (2014); Pummer & MacAskil (forthcoming).

saving a life. While WALYs on their own do not necessarily offer normative guidance on this issue, they do provide two important perspectives from which to consider it. While QALYs are almost exclusively employed to measure quality of life on an individual level, and DALYs are almost exclusively employed to make population level comparisons, WALYs are designed to be a one-stop-shop for both. By allowing for simple and straight forward comparisons of individual and population level wellbeing burdens, WALYs can make transparent the consequences and dynamics of aggregation that are often hidden from view in other measures. In practice, this can allow for more informed public debate regarding the extent to which medical resources ought to be distributed to prioritize individual or population level burdens of disease.

To briefly review, in the first half of this thesis, I argued that democratic representation was not irrelevant to the aims of cost-effectiveness analysis in healthcare, but rather that public preference valuations of health states appeared to be an inappropriate way to go about achieving it. Instead, I offered a tentative test for evaluating questions that the public should or should weigh in on. When it comes to setting explicit normative objectives in healthcare systems, community members as payers and potential beneficiaries have a strong claim to at least some degree of influence. However, once the normative objectives of healthcare analysis are locked in, the role of public preferences is diminished. In this section, I offered two potential avenues for public influence in specifying an approach to cost-effectiveness rooted in patient experience. Both concerned distributive patterns of allocation. The public may have a role to play in influencing the extent to which WALYs are responsive to equity concerns and the degree to which gains or losses in wellbeing are aggregated across populations. As of now, these issues remain largely unspecified in this thesis. I highlight them here as potentially fruitful avenues of future research. While we may have reasons to be sceptical of public involvement or justified in attempting to set normative priorities by other means, public preferences may not have seen their last day in court.

Conclusion

This thesis has attempted to shed light on some of the most important normative issues underlying the widely implemented practice of cost-effectiveness analysis in healthcare. While new attempts to measure and model cost-effectiveness in terms of quality of life represent promising developments, oft relied upon metrics of distribution have foundational limitations. This thesis primarily considered the trade-offs between prioritizing public preferences over patient experiences. While the former hold considerably sway in determining the global distribution of medical resources, there are a number of reasons to be profoundly skeptical of the understanding of quality of life they provide. Integrating literature from philosophy, economics, and psychology, I discussed three of the most powerful arguments ushered in defense of public preference valuations of health states. These clustered around three themes: democracy, impartiality, and adaptation. Ultimately, I argued that all three fail to provide strong moral reasons to jettison patient self-reports in favor of stated public preferences. I also noted that these failings cannot easily be addressed by appealing to the preferences of experts, or even of patients themselves.

However, this exercise was not merely instead to be destructive. I also highlighted key normative criteria that should inform the design and development of a new metric of cost-effectiveness rooted in patient experience. For any metric to be successful, it should leave room for public involvement, harness collective knowledge, limit bias and motivated reasoning, rely on reliable testimony, and be both practically and meaningfully feasible. In the second half of the thesis, I introduced a new preliminary framework that may be able to meet these demands. This approach relied on life evaluations of patients and healthy counterparts to estimate average and potential wellbeing in terms of Wellbeing Adjusted Life Years (WALYs). While preliminary, this approach is poised to outperform public preference metrics in many of the most important normative and descriptive goals they set out for themselves. I also highlighted two as yet underspecified distributive dimensions of this model that could be informed by public priorities and future research. Although more normative work needs to be done, the deficiencies of prevailing approaches to medical resource allocation in the health sector urgently call for creative solutions to improve the metrics upon which they rely. In an age of ever expanding complexity, policymakers need systematic ways of evaluating their decisions. This discussion has sought to highlight key potential areas of normative improvement upon which new approaches to cost-effectiveness analysis in healthcare can be developed in the years to come.

Bibliography

Abásolo, I., & A. Tsuchiya. (2004). Exploring social welfare functions and violation of monotonicity: An example from inequalities in health. *Journal of Health Economics, 23*(2), 313–29.

Alex G, M. J., & Wyrwich, K. W. (2003). Health utility measures and the standard gamble. *Academic emergency medicine*, *10*(4), 360-363.

Attema, A. E., Edelaar-Peeters, Y., Versteegh, M. M., & Stolk, E. A. (2013). Time trade-off: one methodology, different methods. *The European Journal of Health Economics*, *14*(1), 53-64.

Barnes, E. (2009). Disability and adaptive preference. *Philosophical perspectives*, *23*, 1-22.

Bickenbach, J. (2016) Disability and health care rationing. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Retrieved from: https://plato.stanford.edu/entries/disability-care-rationing/

Binder, M., & Coad, A. (2013). "I'm afraid I have bad news for you..." Estimating the impact of different health impairments on subjective well-being. *Social Science & Medicine*, *87*, 155-167.

Brazier, J. E., Rowen, D., Lloyd, A., & Karimi, M. (2019). Future directions in valuing benefits for estimating QALYs: is time up for the EQ-5D?. *Value in Health*, *22*(1), 62-68.

Brazier, J., Akehurst, R., Brennan, A., Dolan, P., Claxton, K., McCabe, C., ... & Tsuchyia, A. (2005). Should patients have a greater role in valuing health states?. *Applied health economics and health policy*, *4*(4), 201-208.

Brickman, P., Coates, D., & Janoff-Bulman, R. (1978). Lottery winners and accident victims: Is happiness relative?. *Journal of personality and social psychology*, *36*(8), 917.

Bronsteen, J., Buddafusco, C. & Masur, J.S. (2013). Well-Being Analysis vs. Cost-Benefit Analysis. *Duke Law Journal, 62*, 1603–1689

Busseri, M. A., & Sadava, S. W. (2011). A review of the tripartite structure of subjective well-being: Implications for conceptualization, operationalization, analysis, and synthesis. *Personality and social psychology review*, *15*(3), 290-314.

Cantril, H. (1965). *The pattern of human concerns*. New Brunswick, NJ: Rutgers University Press.

Caplan, A. L. (2014). Why autonomy needs help. *Journal of medical ethics*, *40*(5), 301-302.

Carel, H., & Kidd, I. J. (2014). Epistemic injustice in healthcare: a philosophial analysis. *Medicine, Health Care and Philosophy*, *17*(4), 529-540.

Claxton, K. et al. (2013). *Methods for the Estimation of the NICE Cost Effectiveness Threshold. Centre for Health Economics*. York: University of York.

Cowell, F. A., Fleurbaey, M., & Tungodden, B. (2010). *The tyranny puzzle in welfare economics: an empirical investigation*. STICERD, LSE.

Craig, B. M., Busschbach, J. J., & Salomon, J. A. (2009). Modeling ranking, time trade-off and visual analogue scale values for EQ-5D health states: A review and comparison of methods. *Medical care*, *47*(6), 634.

Crawford, J. R., & Henry, J. D. (2004). The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large non‐clinical sample. *British journal of clinical psychology*, *43*(3), 245-265.

Damschroder, L. J., Zikmund-Fisher, B. J., & Ubel, P. A. (2005). The impact of considering adaptation in health state valuation. *Social science & medicine*, *61*(2), 267-277.

Daniels N. (1985). *Just Healthcare*. New York: Cambridge University Press.

De Maeyer, J., Vanderplasschen, W., & Broekaert, E. (2010). Quality of life among opiate-dependent individuals: A review of the literature. *International journal of drug policy*, *21*(5), 364-380.

Diener, E., Inglehart, R., & Tay, L. (2013). Theory and validity of life satisfaction scales. *Social Indicators Research*, *112*(3), 497-527.

Dolan P, & Tsuchiya A. (2006). The elicitation of distributional judgements in the context of economic evaluation. In: Jones A (ed). *Companion to Health Economics*. Elgar.

Dolan, P. (1999). Whose preferences count?. *Medical Decision Making*, *19*(4), 482-486.

Dolan, P. (2011). Using happiness to value health. *Monographs*.

Dolan, P., & Kahneman, D. (2008). Interpretations of utility and their implications for the valuation of health. *The Economic Journal, 118*(525), 215-234.

Dolan, P., & Metcalfe, R. (2012). Valuing health: a brief report on subjective well-being versus preferences. *Medical Decision Making*, *32*(4), 578-582.

Dolan, P., Peasgood, T., & White, M. (2008). Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. *Journal of economic psychology*, *29*(1), 94-122.

Dolan, P., R. Cookson and B. Ferguson (1999), 'Effect of discussion and deliberation on the public's view of priority setting in health care: focus group study', *British Medical Journal, 318*(7188), 916–19.

Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.

Gandjour, A. (2010). Theoretical foundation of patient v. population preferences in calculating QALYs. *Medical Decision Making*, *30*(4), E57-E63.

Gold, M. R., J. E. Siegel, L. B. Russell, and M. C. Weinstein, eds. (1996). *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press.

Gold, M. R., Stevenson, D., & Fryback, D. G. (2002). HALYS and QALYS and DALYS, Oh My: similarities and differences in summary measures of population Health. *Annual review of public health*, *23*(1), 115-134.

Grad, F. P. (2002). The preamble of the constitution of the World Health Organization. *Bulletin of the World Health Organization*, *80*, 981-981.

Graham, C., Higuera, L., & Lora, E. (2011). Which health conditions cause the most unhappiness?. *Health economics*, *20*(12), 1431-1447.

Gyrd-Hansen, D. (2004). Investigating the social value of health changes. *Journal of Health Economics, 23*(6), 1101–16.

Haagsma, J. A., De Noordhout, C. M., Polinder, S., Vos, T., Havelaar, A. H., Cassini, A., ... & Salomon, J. A. (2015). Assessing disability weights based on the responses of 30,660 people from four European countries. *Population health metrics*, *13*(1), 10.

Hadorn, D. C. (1991). The role of public values in setting health care priorities. *Social science & medicine*, *32*(7), 773-781.

Halpern, J., & Arnold, R. M. (2008). Affective forecasting: an unrecognized challenge in making serious health decisions. *Journal of general internal medicine*, *23*(10), 1708-1712.

Happiness Research Institute & Leaps by Bayer (2020) *Wellbeing Adjusted Life Years,* Berlin: Leaps by Bayer.

Hausman, D. M. (2015). *Valuing health: Well-being, freedom, and suffering*. Oxford University Press.

Haybron, D. (2016). Mental state approaches to wellbeing. *The Oxford Handbook of Well-Being and Public Policy,* 347-378.

Haybron, D. (2020). Happiness. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Retrieved from: https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=happiness

Helliwell, J., Layard, R., Sachs, J., & De Neve, J. (2020). *World Happiness Report*. New York: Sustainable Development Solutions Network.

Hendriks, M. (2015). The happiness of international migrants: A review of research findings. *Migration Studies*, *3*(3), 343-369.

Holst, C., & Molander, A. (2019). Epistemic democracy and the role of experts. *Contemporary Political Theory*, *18*(4), 541-561.

Hudson. Hudson, N. W., Anusic, I., Lucas, R. E., & Donnellan, M. B. (2020). Comparing the Reliability and Validity of Global Self-Report Measures of Subjective Well-Being With Experiential Day Reconstruction Measures. *Assessment*, *27*(1), 102-116.

Kahneman, D. (2011). *Thinking, Fast and Slow*. Macmillan.

Kahneman, D., Diener, E., & Schwarz, N. (Eds.). (1999). *Well-being: Foundations of hedonic psychology*. Russell Sage Foundation.

Kidd, I. J., & Carel, H. (2017). "Epistemic Injustice and Illness." *Journal of Applied Philosophy 34*(2): 172-190.

Kidd, I. J., Medina, J., & Pohlhaus Jr, G. (Eds.). (2017). *The Routledge handbook of epistemic injustice*. Taylor & Francis.

King, C. H., & Bertino, A. M. (2008). Asymmetries of poverty: why global burden of disease valuations underestimate the burden of neglected tropical diseases. *PLoS Negl Trop Dis*, *2*(3), 209.

Kovács, J. (1998). The concept of health and disease. *Medicine, Health Care and Philosophy*, *1*(1), 31-39.

Lane, H. (2002). Do deaf people have a disability?. *Sign language studies*, *2*(4), 356-379.

Lau, A. L., Cummins, R. A., & Mcpherson, W. (2005). An investigation into the cross-cultural equivalence of the Personal Wellbeing Index. *Social Indicators Research*, *72*(3), 403-430.

Layard, R., Clark, A. E., De Neve, J. E., Krekel, C., Fancourt, D., Hey, N., & O'Donnell, G. (2020). When to Release the Lockdown? A Wellbeing Framework for Analysing Costs and Benefits (No. 13186). IZA Discussion Papers.

Levack, P., Graham, J., & Kidd, J. (2004). Listen to the patient: quality of life of patients with recently diagnosed malignant cord compression in relation to their disability. *Palliative Medicine*, *18*(7), 594-601.

Levy, N. (2014). Forced to be free? Increasing patient autonomy by constraining it. *Journal of medical ethics*, *40*(5), 293-300.

Lindholm, L. and M. Rosen (1998), 'On the measurement of the nation's equity adjusted health', *Health Economics, 7*(7), 621–8.

Litva, A., Coast, J., Donovan, J., Eyles, J., Shepherd, M., Tacchi, J., ... & Morgan, K. (2002). 'The public is too subjective': public involvement at different levels of health-care decision making. *Social Science & Medicine*, *54*(12), 1825-1837.

Luty, J., & Arokiadass, S. M. R. (2008). Satisfaction with life and opioid dependence. *Substance abuse treatment, prevention, and policy*, *3*(1), 1-4.

Lykken, D., & Tellegen, A. (1996). Happiness is a stochastic phenomenon. *Psychological Science, 7*, 186–189.

Margolis, S., Schwitzgebel, E., Ozer, D. J., & Lyubomirsky, S. (2020). Empirical relationships among five types of well-being. *Measuring Well-being: Interdisciplinary Perspectives From the Social Sciences and the Humanities. Oxford University Press forthcoming*.

Marseille, E., & Kahn, J. G. (2019). Utilitarianism and the ethical foundations of cost-effectiveness analysis in resource allocation for global health. *Philosophy, Ethics, and Humanities in Medicine*, *14*(1), 5.

Menzel, P., Dolan, P., Richardson, J., & Olsen, J. A. (2002). The role of adaptation to disability and disease in health state valuation: a preliminary normative analysis. *Social science & medicine*, *55*(12), 2149-2158.

Mitchell, P. (2018). Adaptive Preferences, Adapted Preferences. *Mind*, *127*(508), 1003-1025.

Mont, D. (2007). Measuring health and disability. *The Lancet*, *369*(9573), 1658-1663.

Mukuria, C., & Brazier, J. (2013). Valuing the EQ-5D and the SF-6D health states using subjective well-being: a secondary analysis of patient data. *Social Science & Medicine*, *77*, 97-105.

Murray, C. J., Lopez, A. D., & World Health Organization. (1996). *The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020: summary*. World Health Organization.

Neumann, P. J., Fang, C. H., & Cohen, J. T. (2009). 30 years of pharmaceutical cost-utility analyses. *Pharmacoeconomics*, *27*(10), 861-872.

Neumann, P. J., Sanders, G. D., Russell, L. B., Siegel, J. E., & Ganiats, T. G. (Eds.). (2016). *Cost-effectiveness in health and medicine*. Oxford University Press.

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

Nozick, R. (1974). *Anarchy, state, and utopia*. New York, NY: Basic Books.

Nussbaum, M. C. (2001). Symposium on Amartya Sen's philosophy: Adaptive preferences and women's options. *Economics & Philosophy*, *17*(1), 67-88.

Oakes, J. M., & Johnson, P. J. (2006). Propensity score matching for social epidemiology. *Methods in social epidemiology*, *1*, 370-393.

OECD (2013), *OECD Guidelines on Measuring Subjective Well-being*, Paris: OECD Publishing. Retrieved from: https://read.oecd-ilibrary.org/economics/oecd-guidelines-on-measuring-subjective-well-being_9789264191655-en

Oswald, A. J., & Wu, S. (2010). Objective confirmation of subjective measures of human well-being: Evidence of the U.S.A. *Science, 327*, 576–579.

Parfit, D. (1984). *Reasons and persons*. Oxford University Press: Oxford.

Powdthavee, N. (2015). Would You Like to Know What Makes People Happy? An Overview of the Datasets on Subjective Well-Being. *Australian Economic Review*, *48*(3), 314-320.

Pummer, T. & MacAskil, W. (forthcoming). Effective Altruism. In H. LaFollette (ed.) *International Encyclopedia of Ethics*.

Rabin, R., & Charro, F. D. (2001). EQ-SD: a measure of health status from the EuroQol Group. *Annals of Medicine*, *33*(5), 337-343.

Raibley, J. (2010). Well-being and the priority of values. *Social theory and practice*, *36*(4), 593-620.

Rawls, J. (1971). *A theory of justice*. Harvard university press.

Rowen, D., Zouraq, I. A., Chevrou-Severac, H., & van Hout, B. (2017). International regulations and recommendations for utility data for health technology assessment. *Pharmacoeconomics*, *35*(1), 11-19.

Salomon, J. A., Vos, T., Hogan, D. R., Gagnon, M., Naghavi, M., Mokdad, A., ... & Farje, M. R. (2012). Common values in assessing health outcomes from disease and injury: disability weights measurement study for the Global Burden of Disease Study 2010. *The Lancet*, *380*(9859), 2129-2143.

Sassi, F., Archard, L., & Le Grand, J. (2001). Equity and the economic evaluation of healthcare. *Health technology assessment (Winchester, England)*, *5*(3), 1.

Schwartzberg, M. (2015). Epistemic democracy and its challenges. *Annual review of political science*, *18*, 187-203.

Schwarz, N., Strack, F., Kommer, D., & Wagner, D. (1987). Soccer, rooms, and the quality of your life: Mood effects on judgments of satisfaction with life in general and with specific domains. *European Journal of Social Psychology, 17*, 69–79.

Scully JL. (2018). From "She would say that, wouldn't she?" to "Does she take sugar?" Epistemic injustice and disability. *International Journal of Feminist Approaches to Bioethics, 11*(1).

Sen, A. (1990). Development as capability expansion. *The Community Development Reader*, 41-58.

Shaw, J. W., Johnson, J. A., & Coons, S. J. (2005). US valuation of the EQ-5D health states: development and testing of the D1 valuation model. *Medical care*, 203-220.

Toombs, S. K. (1995). The Lived Experience of Disability. *Human Studies 18*(1): 9-23.

Ubel, P. A. (1999). How stable are people's preferences for giving priority to severely ill patients? *Social Science & Medicine, 49*(7), 895–903

Ubel, P. A. (2001). *Pricing life: why it's time for health care rationing*. MIT Press.

Veenhoven, R. (2005). Inequality of happiness in nations. *Journal of Happiness Studies*, *6*(4), 351-355.

Voorhoeve, A. (2014). How should we aggregate competing claims?. *Ethics*, *125*(1), 64-87.

Vos, T., Abajobir, A. A., Abate, K. H., Abbafati, C., Abbas, K. M., Abd-Allah, F., ... & Aboyans, V. (2017). Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet*, *390*(10100), 1211-1259.

Wagstaff, A. (1991). QALYs and the equity-efficiency trade-off. *Journal of health economics*, *10*(1), 21-41.

Weinstein, M. C. (1988). A QALY is a QALY–or is it?. *Journal of health economics*, *7*(3), 289-290.

Whitehead, S. J., & Ali, S. (2010). Health outcomes in economic evaluation: the QALY and utilities. *British Medical Bulletin*, *96*(1), 5-21.

Wilson, T. D., & Gilbert, D. T. (2003). Affective Forecasting. *Social Psychology*, *35*, 345.

Winter, L., Moss, M. S., & Hoffman, C. (2009). Affective forecasting and advance care planning: anticipating quality of life in future health statuses. *Journal of health psychology*, *14*(3), 447-456.

Xie, F., Gaebel, K., Perampaladas, K., Doble, B., & Pullenayegum, E. (2014). Comparing EQ-5D valuation studies: a systematic review and methodological reporting checklist. *Medical Decision Making*, *34* (1), 8-20.

Xu, K. et al. (2018). Public Spending on Health: A Closer Look at Global Trends. *World Health Organization*. Retrieved from: https://www.who.int/health_financing/documents/health-expenditure-report-2018/en/

Zhao, Y., Feng, H. M., Qu, J., Luo, X., Ma, W. J., & Tian, J. H. (2018). A systematic review of pharmacoeconomic guidelines. *Journal of medical economics*, *21* (1), 85-96.