

Title page [PLACEHOLDER]

How Do People Balance Death against Lesser Burdens?

Veronika Luptakova, Psychological and Behavioural Sciences, LSE

Alex Voorhoeve, Philosophy, Logic and Scientific Method, LSE (corresponding author)

a.e.voorhoeve@lse.ac.uk

Table of contents [PLACEHOLDER]

Acknowledgments.....	III
Abstract.....	IV
Introduction	1
1. Priority-Setting Principles	4
2. Previous Research on the Public’s Judgements	10
2a. Refusals to Prioritize by Severity.....	11
2b. Refusals to Fully Aggregate Claims	16
3. Our Experiment.....	20
3a. Objective	20
3b. Method	22
3c. Sample.....	27
3d. Analytical approach	28
4. Findings	29
5. Discussion and Conclusion	33
Acknowledgements.....	Error! Bookmark not defined.
Bibliography	39
Appendix	48

Acknowledgments

This paper was presented at the LSE, the Society for Experimental Philosophy Conference, the U.S. National Institutes of Health, and the University of York. We are grateful to our audiences and to Matthew Adler, Matteo Galizzi, Dario Krpan, Matthew Lindauer, and Michael Otsuka for comments. Alex Voorhoeve's work on this article was supported through the Bergen Centre for Ethics and Priority Setting's project 'Decision Support for Universal Health Coverage', funded by NORAD grant RAF-18/0009.

Abstract

Suppose one can fully alleviate either the very large individual health burdens of a first group or instead the significantly lesser burdens of a second group that is at least as numerous. In such cases, the most commonly applied principles for priority setting in health have two characteristics. First, when both groups are equally large, they prioritize alleviating the plight of the more severely burdened. Second, when both groups differ in size, these common principles are unlimited in their aggregation: one very large burden can be outweighed not merely by a large number of middling burdens, but also by a sufficiently large number of very minor burdens. While priority for the significantly more severely burdened when group sizes are equal is uncontroversial, many thinkers have challenged unlimited aggregation. There is, however, little evidence on the public's views. Here, we examine the responses to priority-setting dilemmas of a representative sample of the UK population ($n = 389$). We find that an overwhelming majority of respondents' views do not align with commonly used priority setting principles. A substantial share (forty-four per cent) of respondents do not always prioritize by severity when group sizes are equal. Moreover, among those who do prioritize by severity, most reject unlimited aggregation. A test for the influence of the assumed status quo did not yield a robust effect. Overall, our findings suggest that public support for commonly employed priority-setting principles cannot be assumed, and that unorthodox principles that place limits on the ability of very small burdens to outweigh very large ones may be more popular.

Introduction

In health policy, governments often face difficult decisions about who to prioritize for care when resources do not suffice to help everyone in need. In this chapter, we focus on circumstances in which a decision-maker must choose whether to save a first group of people from a greater individual harm or a second, no less numerous, group from a smaller individual harm. The membership of the two groups does not overlap, and in each case, saving group members from harm is assumed to fully eliminate the harm. Moreover, in our cases, the harms are such that if unaided, the members of the first group will be worse off than the members of the second group. An example is a case in which a decision-maker must choose whether to cure 100 young adults from a fatal illness or instead cure an equally-sized or larger group of young adults of an illness which presents a moderate individual health burden. We confine ourselves to cases of certainty, thereby ruling out decisions to give people mere chances at aid, for example by using a lottery to determine which group to assist. Rather, the decision-makers must simply decide which group to prioritize or instead indicate that they should have equal priority.

The principles of distributive justice most commonly employed to make such decisions in practice have two notable characteristics. First, when the two groups are equally large, these principles prioritize by severity. For example, they require saving a number of young people from early death rather than saving a same-sized group of young people from a lesser harm, such as paraplegia. Second, when both the number of people that can be saved from harm and the degree of harm they can be saved from varies, these commonly employed principles are fully aggregative: one death at a young age can be outweighed not merely by a large number of cases of substantial burdens such as paraplegia, but also by a sufficiently

large number of very minor burdens, such as a case of toenail fungus (Fleurbaey, Tungodden, & Vallentyne, 2009).

Prioritization by severity when the number of people one can save from harm is fixed does not occasion much controversy among leading thinkers; nor, to our knowledge, is it often challenged in practice. In contrast, unlimited aggregation has been challenged in both theory and practice. Many philosophers have proposed principles of distributive justice that embody a limited form of aggregation. According to such principles, preventing a large number of cases of substantial harm less bad than death, such as paraplegia, can take priority over preventing one early death. But no number of minor harms, such as cases of toenail fungus, can take priority over preventing one such death (see, e.g. Kamm, 1993; Scanlon, 1998; Otsuka, 2006; Lefkowitz, 2008; Temkin, 2012; Kelleher, 2014; Voorhoeve, 2014; Voorhoeve, 2017; Tadros, 2019; Walen, 2020; Brown, 2020; Rueger, 2020; Steuwer, 2021; Mann, 2021a, Mann, 2021b).¹ Moreover, at least one country, the Netherlands, has adopted such limitedly aggregative priority-setting principles (RVZ 2006; ZiN, 2017, ZiN, 2018, ZiN, 2020; Voorhoeve, 2020).²

One reason often cited in favour of such limited aggregation is that in blocking the ability of comparatively minor harms to collectively outweigh a much greater individual harm, it aligns with common-sense moral case judgements (see, e.g. Otsuka, 2004, p. 424–426; RVZ 2006, 82; Temkin 2012; ZiN 2017). In contrast, in permitting such outweighing of very grave

¹ Limited aggregative views have also faced substantial criticism. See, e.g. Tomlin (2017) and Horton (2018).

² The Netherlands' Institute for Health Care, *Zorginstituut Nederland*, or ZiN, advises the government on what should be covered in the country's mandatory package of health insurance. This package is heavily subsidized for the less well off. The limited aggregative criterion was first proposed by a previous instantiation of this institute, the Council for Public Health and Care, or *Raad voor Volksgezondheid en Zorg* (RVZ).

individual harms by a multitude of very minor harms, unlimited aggregative views are purported to be at odds with common-sense judgement.

This conformity (or lack thereof) of distributive principles with people's case judgements is important for several reasons. First, on the method of reflective equilibrium in theorizing about justice, one seeks principles of distributive justice that offer a rationale for considered, confidently held case judgements and that cohere with other attractive normative principles and still deeper moral ideals (Rawls, 1999; Daniels, 2013). Any inconsistency between such case judgements and apparently attractive principles requires resolution, by revising the case judgements, or by reconsidering one's principles of distributive justice. On this method, the supposed non-conformity of fully aggregative principles of distributive justice with widely held, considered case judgements may weaken our confidence in these principles. Moreover, the supposed conformity of limitedly aggregative principles with common case judgements may bolster such principles.

Second, priority-setting institutions make decisions in the public's name, using its resources. The legitimacy of these institutions depends in part on the public's view that they are employing reasonable decision-making principles (Van de Wijngaard, 2021). Moreover, when such institutions make decisions that offend the public's sense of justice, these institutions may find their decisions are unworkable.³

There is, however, a lack of evidence for the claims made about the public's attitudes towards unlimited aggregative principles. Neither the contributions to the philosophical

³ The well-known case of Oregon's 1990 Medicaid priority-setting exercise is an example. Officials there proposed to set priorities by simple cost-effectiveness. It emerged, however, that on this proposal, tooth capping should be prioritized over treatment for terminal appendicitis, because the former was estimated to be more cost-effective. This ordering of interventions led to outrage and the policy was subsequently withdrawn (Morell 1990; Ubel et al. 1996).

literature nor the Dutch priority-setting authority cite surveys of the public's views on this topic. Moreover, a recent review finds only a small number of relevant surveys, many of which have significant shortcomings, including small, non-representative samples and problematic framing (Voorhoeve, 2018). In this chapter, we aim to fill this lacuna in our knowledge. We examine the responses of a representative sample of the UK population ($n = 389$) to priority-setting dilemmas that have been structured to alleviate the framing issues in previous work. These dilemmas allow us to establish the extent to which respondents' views align with both prioritization by severity when group sizes are equal and with unlimited aggregation when they are not. We found to our surprise that forty-four per cent respondents do not always adhere to prioritization by severity. We also found that philosophers' and Dutch priority-setters' conjecture that unlimited aggregation is unpopular is borne out. For among those who *do* prioritize by severity, only one-fifth give responses that align with unlimited aggregation, while more than half offer limited aggregative responses. Our results therefore challenge the idea that commonly used priority setting principles can count on public support. They also suggest that principles embodying a limited form of aggregation may well garner more support.

We proceed as follows. In Section 1, we offer a brief overview of key priority setting principles. In Section 2, we summarize previous research and some of its limitations. In Section 3, we outline our experiment. In Section 4, we report our findings. In Section 5, we discuss their significance and limitations.

1. Priority-Setting Principles

The leading theories of distributive justice that respect the Pareto principle (the rule that if one can improve a person's well-being without reducing anyone else's, then one should do

so) can be arranged along a continuum by the degree of extra concern they show for the less well off. At one end of the continuum is utilitarianism, which requires maximizing the sum-total of well-being. Utilitarianism has no special concern for the less well off, as it regards a fixed increment of well-being as just as valuable when it accrues to a worse off person as when it accrues to a better off person. At the other end of this continuum is leximin, which requires maximizing the well-being of the least well off, and once this has been accomplished, of the second-least well off, and so on. One prominent example of a view in between these two end points is pluralist, moderate egalitarianism, which aims to both reduce inequality and improve total well-being (Tungodden, 2003). On this form of pluralist egalitarianism, a gain in well-being to someone who is worse off than others is more (but not infinitely more) valuable than an equally-sized gain to a better off person, as it both reduces inequality and improves well-being. Another example of a view on the spectrum is prioritarianism, which does not care about inequality itself, but which simply regards a gain in well-being as more important, the lower the level of well-being from which it arises, irrespective of whether there is anyone who is better off (or worse off) than the person being aided (Parfit, 1995; Adler, 2012).

If, as is common in health economics, we take quality-adjusted life years (or QALYs) as the relevant unit of health-related well-being, then utilitarianism is embodied in standard cost-effectiveness analysis which prioritizes interventions by their cost per QALY gained. An example is the approach championed by the World Health Organization (WHO)'s CHOICE project, which stands for CHOosing Interventions that are Cost-Effective (Betram and Edejer 2021). Views in between utilitarianism and leximin are embodied by common forms of equity-weighted cost-effectiveness analysis that give non-infinite additional weight to QALY gains to the worse off (Bognar & Hirose, 2014). An example is the priority-setting principles proposed by the Norwegian Committee on Priority Setting in the Health Sector, on which a QALY gained

by someone facing a large health burden is taken to be one and a half times more valuable than a QALY gained by someone facing a middling health burden, and three times more valuable than a QALY gained by someone facing a small health burden (Ottersen et al., 2016).

Insofar as practical policy-making in health is guided by explicit principles of distributive justice, utilitarianism in the form of standard cost-effectiveness or moderate pluralist egalitarianism/prioritarianism in the form of equity-weighted cost-effectiveness with non-infinite additional weights for improvements for the worse off dominate. In this paper, we shall focus on two characteristics of such commonly applied principles. First, in the choices of the kind outlined in the Introduction, when the groups facing the greater harm and the lesser harm are equal in size they prioritize by severity. For saving the group facing the greater individual burdens in such a case both maximizes total well-being and improves the lot of the worse off. Second, these common principles mandate saving a larger group from a lesser individual harm if the larger group is sufficiently numerous, even when the lesser harm is much less severe than the greater harm (Fleurbaey et al., 2009). The only difference is that for utilitarianism, fewer people are required for the balance to tip in favour of the people facing the lesser harm than for moderate egalitarianism or prioritarianism. To illustrate, consider a very mild chronic ailment that reduces health-related well-being in an otherwise healthy population by one-thousandth of a QALY for every year of its duration. Suppose that it starts at the age of twenty and lasts until the person dies suddenly (from other causes) at eighty. On the form of utilitarianism embodied in standard cost-effectiveness, it would take one thousand such cases to outweigh one death of a young person at twenty who, if saved, would otherwise live in good health until eighty. On the Norwegian proposal, the death of the young person would receive three times this weight, so it would take three thousand such cases.

As mentioned in the Introduction, from the 1990s onwards, a growing number of philosophers have argued that unlimited aggregation is problematic, because no number of such small harms to people who would in any case lead good lives should together outweigh one young person's death. Leximin shares this judgement, because in such cases, it prioritizes the young person facing death no matter how many others one could save from the minor ailment. But leximin's degree of priority for the worst off is extreme. For it demands that we eschew any amount of improvement in the second-least well off position in a distribution of well-being for the sake of the tiniest improvement in the very worst position. It also prioritizes aiding a lone person in the worst position over saving any number of people from lesser, but still very substantial health burdens, such as paraplegia. These problematic implications mean that it has not, to our knowledge, been employed by policy-makers as a guiding principle for priority-setting. They have also led critics of unlimited aggregation to propose other principles, on which some number of substantial impairments, but no number of very minor ones, can together outweigh a young person's untimely death.

The proposed limited aggregative principles can be divided into two broad types. One type posits an absolute threshold level of well-being and holds that an increment in well-being below this threshold, no matter how minute and no matter how few people would obtain it, always outweighs an increment above this threshold, no matter how sizeable the increment and no matter how numerous its recipients. When all gains accrue to people below the threshold, then the number of people benefited counts, as does the size of the gains and the level of well-being from which they take place (Crisp, 2003; Brown, 2005; Liao & Lim, 2022). A proposed rationale for this type of view is that harms that present substantial obstacles to leading an unqualifiedly good life are normatively different in kind from those that do not do

so, and that alleviating (in whole or in part) the former kind of harm should take lexical priority over alleviating the latter kind of harm.

A second type of view, in contrast, eschews an absolute threshold. Instead, it focuses on comparing the strength of the claim of a person who has most at stake in the decision with the strength of the claim of each person who has less at stake. It then holds that when the gap between the strength of the strongest claim and a weaker claim is too large, no number of the weaker claims can together outweigh the strongest claim. On this type of view, whether weaker claims aggregate therefore depends on their strength *relative* to the strongest competing claim. One example of this approach is Aggregate Relevant Claims, or ARC (Voorhoeve, 2014), which we therefore focus on for illustrative purposes. In our cases, ARC holds the following:

1. Each person whose fate is at stake has a claim.
2. People's claims compete just in case they cannot be jointly satisfied.
3. A person's claim is stronger:
 - 3.1 the greater the gain in their health-related well-being generated by aiding them; and
 - 3.2 the less health-related well-being they would enjoy if they were not aided.
4. A claim is relevant if and only if it is strong enough when compared to the strongest competing claim.
5. The decision-maker should satisfy the greatest sum of strength-weighted, relevant claims (Voorhoeve, 2014, p. 66).

ARC is inspired by the idea expressed in Nagel (1979, p. 118 and p. 126) that both an aggregative and a non-aggregative approach capture part of the demands of distributive

justice. On the aggregative approach, the equal moral importance of each person's well-being requires that we assign the same and positive marginal moral importance to every person's claim of a given strength. It meets this demand by holding that fulfilling two claims of a given strength is twice as important as meeting one such claim, and so on for all claims. The upshot is a requirement to meet the greatest sum of strength-weighted claims.

The non-aggregative approach embodies a different conception of equal concern which takes extremely seriously the distinctness of each individual's life. To fully respect the distinction between persons, it holds that one must compare each person's claim, taken separately, with each competing claim, considered separately (Nagel 1979, p. 116f). In this confrontation, a stronger claim always wins out. On this view, one therefore always has the strongest reason to meet the individually strongest claim.

Nagel (1979) suggests that a full theory of justice must incorporate elements of both approaches, while arbitrating between them in a reasonable way. ARC does so by maximizing the total fulfilment of strength-weighted claims under the restriction that the fulfilled claims are close enough in strength to be relevant to each other. It thereby follows the aggregative approach so long as one does not thereby depart too far from the demand of the non-aggregative approach.

Naturally, to offer guidance in particular cases as well as a fuller rationale for their views, proponents of ARC must explain how it determines which weaker claims are still relevant to competing, stronger claims, and which are not. A number of such proposals exist (Voorhoeve, 2014; Voorhoeve, 2017; Brown 2020). Proponents of threshold views most of course also try to meet the related challenge about where to locate the boundary between harms that substantially impair the ability to lead a good life and those that do not (Crisp,

2003; Liao & Lim, 2022). While there is also a lively debate about which of these types of limited aggregative views to favour, we shall set this debate aside in the remainder of this paper. For our concern is to uncover the extent to which subjects favour this family of unorthodox, limited aggregative views over unlimited aggregation. Our experiment therefore is designed not to distinguish between the forms of limited aggregation, just as it does not distinguish between the utilitarian, prioritarian or pluralist egalitarian views that make up the family of fully aggregative views.

Limited aggregative views require further specification to be applied to health care priority setting. An example is offered by the method used by Netherlands' Institute for Healthcare (ZiN, 2018, p. 4) for judging whether treatments are cost-effective. For individuals who suffer health burdens equal to or larger than 0.1 QALY per year, it employs a form of equity-weighted cost-effectiveness. Interventions that alleviate a burden of illness ranging from 0.1 to 0.4 QALY per year are judged cost-effective up to €20,000 per QALY gained; those that alleviate a burden of between 0.4 and 0.7 QALY per year are judged cost-effective up to €50,000 per QALY generated; and those that alleviate a burden of more than 0.7 QALY per year are judged cost-effective up to €80,000 per QALY gained. In line with a form of limited aggregation, it regards an annual individual burden of less than 0.1 QALY as not generating a call on public resources which could instead be used for life-saving interventions.⁴

2. Previous Research on the Public's Judgements

As mentioned in the Introduction, both the method of reflective equilibrium and considerations of legitimacy give us reason to consider people's moral preferences regarding

⁴ Voorhoeve (2020) defends a form of limited aggregation but argues that the threshold for relevance should be set in a different manner, and in many cases substantially lower, than it is in the Dutch context.

resource allocation decisions, and to incorporate them into public decision-making if they are reasonable. To do so, we must be able to reliably elicit these preferences and establish whether they are based on rational judgements and morally acceptable views (Dolan et al., 2003, p. 549).

In this section, we review the existing empirical literature on people's preferences about distributive justice in health care. Specifically, we focus on evidence, or lack thereof, for the assertion that people tend to prioritize by severity (subsection 2a) and that they fully aggregate existing claims on public resources (subsection 2b).

Decisions about the allocation of health care resources are about making trade-offs. However, when people are asked to make trade-offs, which are often difficult, many of them provide *non-trade* responses. These refusals to prioritize may be expressions of deeply held moral beliefs, but they may also represent "protest" responses when people are resistant to the idea of having to make difficult trade-offs (Damschroder et al., 2007). In order to properly interpret and account for people's preferences, it is important to understand what drives such non-trade responses. We now review two types of such responses that are pertinent to our investigation: so-called "equivalence refusals" and "off-scale refusals".

2a. Refusals to Prioritize by Severity

A prominent way in which individuals refuse to prioritize between two programmes that can fully cure the same number of people who have health conditions of different severity is to declare that these programmes should have equal priority (Damschroder et al. 2007, p. 266). This form of denial of priority by severity in same-sized group cases is known as an "equivalence refusal" and contradicts all commonly employed priority-setting principles as well as the unorthodox, limited aggregative ones considered here. (Note that not every

equivalence *response* to a priority-setting question is an equivalence *refusal*. The latter are, by definition, only those that cannot be explained through any of the theories of distributive justice outlined.)

Nord (1995) and Damschroder et al. (2007, p. 266) point out that it is unclear how to properly interpret such responses – whether to count them as considered moral judgements that express a strong preference for ensuring equal access to treatment for everyone in need (even when needs are substantially different), or whether to regard them as reflecting people’s unwillingness to make difficult choices.

In this subsection, we review existing empirical evidence on the prevalence of such refusals. We will use this as a proxy to estimate the share of the public who do not prioritize by severity in fixed-number cases.

We focus our review on empirical studies that involved elicitation using the Person Trade-Off (PTO) method. This method elicits moral preferences over inter-personal trade-offs by choosing between treatment programmes for two distinct groups of people, which generally differ in their size and/or in the severity of illness that they can be cured of. The PTO approach requires decision-makers to choose between (i) providing a large health benefit for a few people and, (ii) a smaller health benefit for many others (Schwarzinger et al., 2004, p. 172). Typically, respondents’ preferences are elicited either directly or in two steps. In the “direct” elicitation, respondents are presented with a description of two treatment programmes that are helping two different groups of patients and are asked that if one programme helps ten (or other baseline number of) patients suffering from condition X, how many patients suffering from condition Y would the second programme have to help in order for them to be indifferent between the two programmes. In the “step-wise” elicitation,

respondents are asked in step one to prioritize for treatment one of two same-sized groups that can be cured of ailments of differing severity. In step two, the characteristics of the group that was prioritized for treatment at step one is kept constant, but the number of people in the group that was deprioritized is varied to find the point of indifference, or “equal choiceworthiness” between the two programmes. By way of illustration: in the first step, a respondent is asked to prioritize either programme A, which cures 100 patients of a fatal illness or programme B, which cures 100 patients of paraplegia, or to indicate that these programmes should have equal priority. If they choose programme A, in the second step, this respondent is asked how many people would have to be cured of paraplegia to make them consider the two treatment programmes equally choiceworthy. If, instead, the respondent indicates that the two programmes are equally choiceworthy when they both save 100 people, even though they view early death as significantly more severe than paraplegia, then their response is counted as an equivalence refusal.

Only a limited number of PTO studies report the share of equivalence refusals among the public. In many studies, the condition that the equivalence refusals are at odds with all leading theories of distributive justice is not met or the share of the equivalence refusals is not reported.

For instance, some PTO studies do not directly compare treatment programmes that help the same number of people with conditions of differing severity. Instead, the numbers of people helped are adjusted based on the severity that respondents assigned to the respective health conditions. For instance, respondents who, on a scale on which the well-being of time alive in perfect health is rated at one and death at zero had previously indicated that a debilitating knee condition has a well-being value of 0.5, while a case of acute

appendicitis (which is terminal if untreated) has a well-being of zero, were then asked whether they would rather cure ten people acute appendicitis or twenty people of the knee condition or whether they were indifferent between the two (Ubel et al., 1996). Providing an equality response (indicating indifference) in such a case would not count as an equivalence *refusal* as one of the leading theories of justice – utilitarianism – would indicate equality.

Reviewed studies with a suitable design that reported the share of equivalence refusals provide a rather broad range, between zero per cent and forty-three per cent (Nord et al., 1993; Nord, 1995; Damschroder et al., 2005; Damschroder et al., 2007; Schwarzinger et al., 2003, Schwarzinger et al., 2004). Although all reviewed studies asked for trade-offs between health conditions that appeared to be clearly different in severity and asked respondents to assume the treatment programmes could cure (or prevent the onset of ill health for) the same number of people, they differed on several dimensions, such as, sample size, representativeness, framing of the question, including a life-threatening condition, comparing health conditions of the same or of different kinds (e.g. mild vs. moderate vs. severe shortness of breath or moderate leg pain vs. severe shortness of breath), and giving participants an opportunity to discuss, reflect on, and change their responses.

The estimates that are probably most representative of the views of the public come from two studies in Damschroder et al. (2007). Both studies were conducted using relatively large samples ($n = 388$ and 878) recruited from an internet panel and representative of the US population. They tested the impact of different contexts and framings on the share of equivalence refusals. They found that people were more likely to give equivalence refusals when making a decision from an *evaluator* perspective (i.e. being in a more detached role) than from a *decision-maker* perspective (i.e. being in a role directly responsible for the

decision), with the share of equivalence refusals being thirty-two per cent and twenty-one per cent, respectively. Furthermore, their results showed that reminding people of the consequences of their choice (i.e. that people in the group which was not selected will be untreated, *a rationing perspective*) resulted in an even lower share of equivalence refusals (twelve per cent), while asking people to make the choice from a *benefits perspective* (asking them which group of people received a greater benefit rather than which group should be prioritized) led to a higher share of equivalence refusals (forty-three per cent).

The results also show that people who felt outraged about the idea of rationing and those with lower numeracy were more likely to give equivalence refusals. Interestingly, those who reported finding the questions difficult to answer were *least* likely to give equivalence refusals (Damschroder et al., 2007, p. 270-271, 274), which might reflect their level of engagement.

One limitation of the reviewed studies is that the spectrum of harms was somewhat limited, as both very minor health complications and fatal conditions were omitted.⁵ This can have an impact on the share of equivalence refusals, as respondents tend to be sensitive to a relative difference in severity between the compared conditions. They were most likely to give equivalence refusals when the difference in severity was small (e.g. moderate vs. mild shortness of breath) and the least likely when the difference was large (e.g. quadriplegia vs. foot numbness) (Damschroder et al., 2007; see also Schwarzingler et al., 2004, Nord et al., 1993). It must be noted, however, that in some cases, comparing fatal diseases with other severe but not life-threatening conditions might be problematic as some respondents

⁵ Participants were asked to compare three conditions involving paralysis, ranging from complete foot numbness to leg paralysis (paraplegia) to quadriplegia (the “paralysis triad”) and three conditions involving shortness of breath, from mild to moderate to severe (the “shortness of breath triad”).

consider conditions such as paraplegia or quadriplegia to be worse than death (Damschroder et al., 2005; Schwarzinger et al., 2003). In such cases, one cannot say whether respondents really gave equivalence refusals or whether, instead, they judged the compared conditions to be approximately equally severe.

Lastly, available empirical research provides relatively little qualitative analysis. Therefore, we lack a solid understanding of what drives equivalence refusals, and whether they are protest responses or considered moral judgements that apply principles different from those outlined in Section 1.

In sum, only a handful of studies identify genuine equivalence refusals, which we take to be a major source of a refusal to prioritize by severity. This leaves a lacuna in our knowledge that we aim to fill. As we detail in Section 3, in our study design, we aimed at overcoming several of the identified weaknesses to arrive at a more precise estimate of the prevalence of equivalence refusals among the general population. We used a large, representative sample of the UK population; we included a wide spectrum of harms to let people compare conditions with a very large as well as rather small difference in severity; and we tested for individual perceptions of the severity of compared health conditions (rather than assuming these or using average assessments).

2b. Refusals to Fully Aggregate Claims

What are known as “off-scale refusals” arise when people indicate that the programme curing the more severe ailment should always be prioritized, no matter how large the number of people treated under the alternative programme which cures the less severe ailment. Such answers are inconsistent with fully aggregative views, but they are compatible with both leximin and limited aggregative approaches. The latter two differ, however, in the

circumstances under which they generate off-scale refusals. Leximin requires prioritizing the more severe condition in all circumstances, even when the somewhat less severe condition is still comparatively substantial. By contrast, limited aggregative approaches permit a large number of cases of the less severe condition to outweigh a small number of cases of the more severe condition when both conditions impose substantial harms. To find support for limited aggregation, one would therefore need to see the share of off-scale refusals increase when one moves from a choice between groups of people who each face substantial harm to a choice where the members of one group face very substantial harm and the members of the other group face only minor harm.

In this subsection, we review the three studies that come closest to providing evidence of the extent of such a pattern in off-scale refusals in healthcare priority setting. Ubel *et al.* (1996) surveyed a small sample of university students ($n = 42$), who were asked to evaluate a choice between curing ten individuals of an otherwise terminal case of appendicitis and curing a less burdensome health problem in a number of others. The lesser health burden in question was either very small (a ganglion cyst on a tendon which caused a bulge and occasional minor pain, but which did not limit activities) or substantial (a benign meningioma—a growth in the tissues around the brain—leading to regular, debilitating headaches though without an impact on life expectancy). Participants were asked which number of people cured of the illness causing the smaller burden would produce the same total benefit brought about by saving ten lives (Ubel *et al.*, 1996, p. 111). In line with the pattern predicted by limited aggregation, off-scale responses of “no number” or “infinity” were more frequent in the lives versus cyst case, with 40.5 per cent of responses being off-scale refusals, than in the lives versus meningioma case, with 4.8 per cent of responses being off-scale refusals.

As noted in Voorhoeve (2018, pp. 128-132), this study has two shortcomings. First, it involves a small convenience sample. Second, it uses a *benefit frame*, asking which number of lesser burdens one would need to alleviate to generate the same benefit as saving ten lives. But many defences of limited aggregation are deontological, or non-consequentialist (see, e.g. Kamm, 1993; Scanlon, 1998): they regard it as wrong to save a great many people from very minor harms rather than ten from death even if saving people from the very minor harms would produce a greater aggregate benefit. The survey fails to capture such deontologically-motivated responses, which would be better reflected in a *choice frame*, which requires individuals to specify for which number of people (if any) facing the lesser harm we should prioritize saving them rather than saving the ten from death.

In the study reported in Pinto-Prades and Lopez-Nicolas (1998), participants (Spanish students, $n = 83$) were told they could either establish a neonatal care unit which would save the lives of ten new-borns or implement a policy that would treat a “very large number (e.g. 100,000)” of others for an ailment that causes a smaller health loss. The latter loss varied; the smallest burden was living with moderate pain and discomfort which did not prevent the activities of daily life. They were then asked which programme they would *prefer* to see implemented. The paper reports only the majority answers for various trade-offs: a majority chose to save ten new-borns rather than treat the moderate chronic pain of a “very large number” of people and a majority was willing to alleviate the somewhat greater burdens (still short of death) of a multitude rather than save the lives of ten neonates (Pinto-Prades & Lopez-Nicolas, 1998, p. 290). It follows that at least some respondents must have expressed preferences that align with limited aggregation, holding that saving a number of lives takes priority over saving a very large number from small harms, while also holding that that saving

a very large number of people from serious harms should take priority over saving a small number of lives.

This *preference framing* may be regarded as superior to the benefit framing in Ubel et al. (1996) since it involves a judgement about which programme one would want to see prioritized. But it is still not ideal, since of course it is conceivable that someone would prefer (on other grounds, e.g. self-interest) to see a policy implemented that they regarded as morally impermissible.

There are several further weaknesses in the study design. For one, it again involved small convenience sample. Moreover, to test for off-scale refusals, respondents should have been asked whether they preferred to save the ten lives for *any* (natural) number of people they could cure of the lesser ailment. The magnitude of the benefits of the possible interventions were underdescribed. It was left open what the quality of life of the neonates would be if they were saved and the extent to which the lesser conditions would be alleviated by treatment. The fact that it involves neonatal death also muddies the waters, since on some moral views, neonates lack the mental capacities of persons, and the death of a non-person that lacks a conception of itself as a being that persists through time and that has, as yet, no desires and plans for the future is less bad than the death of a person (Millum, 2015).

Study 1 in Damschroder et al. (2007) improves on these studies in two respects. First, as mentioned in subsection 2a, it involved a larger sample ($n = 827$) chosen to represent the US population in various key respects. Second, it employed a choice frame and specified the magnitude of the benefits at stake for each person. Participants were asked to choose between saving a small number from a greater individual burden against saving a larger number from lesser burdens. The most substantial burden considered was quadriplegia and

the least significant was paralysis in one foot. As in Ubel et al. (1996), the overall pattern of responses suggested that a substantial share of participants chose in line with limited aggregative views. In the choice involving the greatest gap in severity of condition – curing ten quadriplegics or instead up to one million people of foot paralysis – forty per cent of respondents said that they would always cure the quadriplegics. By contrast, in the choice involving the smallest gap in severity – curing ten people of quadriplegia or up to one million of paraplegia – seven per cent said that they would always prioritize the former. Of course, these responses do not, strictly speaking, register off-scale refusals, since they involve an upper bound. But there is evidence that these were primarily off-scale refusals. For participants who chose to aid the quadriplegics rather than up to a million people with a lesser health burden were also asked whether there was a number of people “in the world” for which they would prioritize a programme alleviating the lesser burden over one saving the ten quadriplegics. Most (fifty-nine per cent) answered negatively (Damschroder et al. 2007, p. 270). The principal shortcoming of this study, however, is that the spectrum of harms is insufficiently wide. For permanent paralysis in one foot is still quite a substantial harm. The design therefore will fail to catch some respondents who hold that a multitude of cases of foot paralysis can outweigh a case of quadriplegia, but that no number of very minor ailments could outweigh loss of a young person’s life.

3. Our Experiment

3a. Objective

The reviewed empirical studies suggest that there are two distinct ways that individuals may choose contrary to all commonly employed principles of distributive justice in health. (1) Some refuse to prioritize by severity between groups that are equally large, even when the

difference in severity is substantial; and (2) Others, in contrast, refuse to let a multitude of small benefits to the better off, no matter how numerous this multitude, outweigh a fixed number of large benefits to the worse off, i.e. they fail or refuse to fully aggregate. We also outlined that (2) may be rationalized by the unorthodox, but still widely discussed, partially aggregative views outlined in Section 1.

The objective of our empirical research was to provide a better estimate of the prevalence of both types of departures among the general population through improvements in the research design, which would address some of the shortcomings identified in Section 2 in some or all of the previous studies, namely: non-representative samples; failing to check whether individuals ordered the severity of the conditions in question the way the survey designer intended; not clearly specifying the benefits at stake in the trade-off; using a benefit or preference frame rather than a choice frame; and unduly limiting the spectrum of harms under consideration.

We also aimed to test whether people's decisions are influenced by the assumed status quo, i.e. by what people believe is currently covered through the healthcare system.⁶ If people's views are disproportionately influenced by the status quo, it poses a problem for policy-makers, especially if there is a need for a reform, as people might be inclined to preserve possibly inequitable allocations of healthcare resources (Dolan & Robinson, 2001).⁷

⁶ For instance, if people say that *no* number of patients cured of a mild disease should compensate for not saving one person's life (an off-scale refusal), would they make the same judgement even if they learned that such trade-offs are currently being made (i.e. the status quo is that treatments for prevalent minor health conditions are funded but treatments for some rare and much more serious conditions are not covered)?

⁷ Dolan and Robinson (2001) make this point in relation to loss aversion, which is one of the potential drivers of status quo bias.

There is persuasive evidence from experimental (lab and field) research that human decisions are disproportionately influenced by the status quo.⁸ This has been tested and identified in a variety of decisions, including choosing investment portfolios, choosing between job offers, selecting a car colour, choosing auto insurance or a cancer testing programme for oneself (Samuelson & Zeckhauser, 1988; Kahneman, Knetsch, & Thaler, 1991; Johnson, Hershey, Meszaros, & Kunreuther, 1993; Salkeld, Ryan, & Short, 2000; Fernandez & Rodrik, 1991). However, the vast majority of the empirical research to date focuses on *non-moral* or purely prudential domains, mostly pertaining to self-regarding consumer or business choices. It is unclear, however, whether the status quo also influences decisions involving *moral* judgements involving interpersonal trade-offs (Caviola et al., 2014). Although the presumption seems to be that matters that influence decisions in the non-moral domain will also do so in the moral domain, the context-dependence of decision rules means this cannot be assumed. While moral philosophers have discussed and theoretically demonstrated the potential impact of the assumed status quo on moral judgements (e.g. Bostrom & Ord, 2006; Wasserman, 2015; Sparrow, 2015), the proposition requires further experimental confirmation.

3b. Method

We used an online questionnaire to capture participants' trade-offs between health care programmes treating conditions of different severity. The least severe condition – Nail Disease – caused minor, but still appreciable, discomfort. The intermediate condition –

⁸ Samuelson and Zeckhauser (1988) coined the term “status quo bias” to describe people’s tendency to *disproportionately* stick with the status quo. Decision-makers exhibit the bias to the extent to which they choose the status quo option “more frequently than would be predicted by a canonical model” (p. 8). This model assumes that people make choices solely based on their preferences. Therefore, irrelevant aspects of options under consideration, such as their order, framing, or being labelled as ‘status quo’, should not have an impact on the actual decision (Samuelson & Zeckhauser, 1988).

Paraplegia – was serious but not life-threatening. The most severe condition – Fatal Autoimmune Disorder – was deadly within a short period if left untreated. We asked participants to make two trade-offs, one between the most severe and the least severe health conditions (Lives vs Nails Scenario, large severity gap) and the other between the most severe condition and the serious but not life-threatening condition (Lives vs Paraplegia Scenario, moderate-to-small severity gap).

We believed it would be easier for people to imagine concrete health conditions. Therefore, we did not use a generic health status instrument, such as EurQoL, but instead we provided participants with a short description of three quite specific health conditions, including their typical symptoms and prevalence.⁹

⁹ **Nail disease:** A fungal nails infection that causes yellow discolouration and thickening of toenails. People suffering from the disease complain about unpleasant looks and mild itching but usually it is not painful and does not cause further health complications. It is quite a common disease, affecting around 15,000 out of 100,000 adults.

Partial body paralysis (Paraplegia): An impairment in motor or sensory function of lower part of the body. People with this condition cannot move their legs due to a spinal cord injury. Many patients suffering from this type of paralysis feel chronic pain that can be alleviated only partly. Although there are several health complications resulting from this condition (such as incontinence, depression, impotence, circulation disorders, etc.), it is typically not life-threatening. While it is a less common condition, it still affects significant number of people, about 40 out of 100,000 adults.

Fatal autoimmune disorder: A rare life-threatening autoimmune disorder that causes people to lose control over some of their muscles. It is characterized by painful involuntary muscle contractions in different parts of the body. Patients are gradually unable to perform normal activities such as walking, eating, or talking and in later stages have problems breathing. The disease progresses rather quickly, with average survival of about 30 months after the onset of symptoms. Currently its prevalence is very low, affecting only 2 out of 100,000 adults.

Note, that it has been shown that people care not only about the absolute number of lives saved but they care about how many lives were saved in relation to how many were at risk (Slovic & Västfjäll, 2013). Therefore, information of prevalence of a disease could possibly lead respondents to prefer programmes that cure people suffering from less prevalent conditions as the proportion of cured over affected was higher. However, we estimate the impact of this was minimal as we included the information on prevalence just to make the initial descriptions realistic and this information was not displayed at the point of making a decision. Furthermore, although in step 1, when the number of cured people was fixed to 100, they would cure a different proportion of all people affected by each condition, in step 2, respondents had an opportunity to indicate their own number of people that would have to be cured (in order for them to be indifferent between the programmes) and that could reflect for instance the prevalence of conditions, if that was their concern.

To make sure that respondents perceived the severity of the selected health conditions as we had intended, we asked them to perform an assessment of the severity for each condition at the beginning of the experiment. Specifically, they were asked to indicate “how difficult it is for a person to live with a respective health condition” on a scale from 0 to 100 (where 0 was “Not difficult at all” and 100 was “Extremely difficult”). Answers of those who did not indicate that living with the fatal disorder is more difficult than living with the partial paralysis and that living the partial paralysis is more difficult than living with the nail disease (i.e. their severity was not ordered as intended) were excluded from the analysis, as the study design would not be suitable to capture their moral judgements.¹⁰ Moreover, we believed that engaging in this assessment was likely to induce more careful consideration of the individual burden posed by these illnesses.

We used the PTO method to elicit people’s preferences.¹¹ We selected this method mainly for three reasons. First, it is a matching type of elicitation which should encourage deliberation more than simple choosing (Skedgel, 2013). Second, it prompts respondents to adopt a societal rather than a purely individual perspective and thus allows them to incorporate concerns for distributive justice (Dolan & Green, 1998; Singh et al., 2012). Third, it explicitly requires people to make a trade-off between two options, rather than assessing each one of them in isolation. Making such trade-offs is typically at the heart of decisions on

¹⁰ We had pre-tested this spectrum of health conditions before launching the survey to the online panel. Still, as we point out in the Findings section, there were a significant number of participants that perceived the spectrum of health conditions differently than intended.

¹¹ We could have asked them directly whether there is a number large enough that they would rather save that number of people from nail disease/ paraplegia than save someone from a fatal disorder, or what they considered permissible/ impermissible (Voorhoeve, 2018). However, we assumed that for an ordinary person it would be difficult to understand such philosophically framed questions.

distributive justice, including in health care, and therefore using PTO method is helpful (Payne et al., 1992; Schwarzinger et al., 2004; Singh et al., 2012).

Participants were instructed to assume for each decision that: (a) the treatment would be a full cure; (b) there was no alternative way to secure the treatment, and so patients suffering from the condition that is not chosen would go untreated; (c) the total cost of the programmes treating the different health conditions would be the same; and, (d) the budget had to be allocated to one condition or the other and could not be split.

Each participant was asked to answer two pairwise PTO questions. In Lives vs Nails, they compared a programme that would cure people of the nail disease and programme that would cure people of the fatal disorder. In Lives vs Paraplegia, they compared a programme that would cure people of paraplegia and programme that would cure people of the fatal disorder.

In each scenario, in the first stage, participants were asked which of the two programmes they would choose to fund if each programme would cure 100 people, or if instead they held that both programmes should have the same priority (i.e. an equivalence refusal). If they chose the programme that provided a cure for the fatal disorder, they were asked how many people would have to be cured under the other programme (i.e. treating the nails disease or paraplegia) in order for them to be indifferent between the two programmes (the matching task). They could either choose a specific number from the range provided (from 101 people to the total number of people in the healthcare system), or they could indicate that there is *some* number, but they did not know what exactly the number is, or they could select the option that there was *no* number large enough and that the programme curing the fatal disorder should always be prioritized (i.e. an off-scale refusal).

When participants selected a non-numerical response (an equivalence or off-scale refusal), they were asked to give a rationale for their response, if they could, before being directed to the next section of the survey.

The questions in both decisions were asked from the perspective of an ‘advisor.’ Participants were asked to imagine that a health care system director is deciding how to allocate budget for the upcoming period and were then asked which programme the director *should* choose. This perspective aligns well with the role of the public in healthcare policy decisions – people can typically express their judgements about how the decision-makers should decide, rather than having to decide themselves.¹² The decision was presented as one that is still to be made rather than one that has already been made, as we assumed the former would be more engaging.

As one of the robustness checks, we tested for the impact of the status quo on people’s allocation decisions. We used a between-subjects study design to avoid carryover effects that could affect within-subjects design. Participants were randomly allocated into three groups, each presented with different information on current coverage from the healthcare budget, i.e. the status quo. Participants in the Control group were informed that treatment for none of the health conditions under consideration is currently covered.¹³ Participants in the Less Severe Currently Covered group were informed that milder conditions

¹² It has been hypothesized that using a more detached perspective, such as the one of an advisor, could make it easier for respondents to discriminate between the groups of people as it would not force them to make ‘tragic choices’, which could result in a utility loss (Dolan et al., 2003, p. 550). However, empirical research conducted by Damschroder et al. (2007) have shown that while a detached perspective of an ‘evaluator’ leads to lower number of off-scale refusals (compared to a decision-maker’s perspective) it actually increases equivalence refusals.

¹³ Alternatively, the information about the status quo could have been omitted or both treatments could have been covered. However, the former alternative might prompt people to make their own assumptions about what the status quo is, and the latter could trigger loss aversion, either of which could confound the results.

out of those under consideration are currently covered as these affect a sufficiently large number of people but the most serious (fatal) health condition is not currently covered as it only affects very few patients. We conjectured that this version would incline participants towards an aggregative perspective (“aggregative nudge”), on which it is permissible to treat less serious diseases for a sufficiently large number of people instead of treating a serious health condition for very few. Participants in the Fatal Currently Covered group were informed that treatment for the most serious health condition under consideration is currently covered as it is fatal if untreated. We conjectured that this specification of the status quo would incline participants towards giving priority to those who have the most at stake (“anti-aggregative nudge”). As a result, we hypothesised that the number of off-scale refusals in the Less Severe Currently Covered group would be lower and that in the Fatal Currently Covered it would be higher (compared to the Control group), which would have an impact on the prevalence of choices in line with limited aggregative views.

At the end of the section in which the information on the status quo was provided, we included a test of whether people really understood what was assumed to be currently covered in the health care system (we will refer to this as the “status quo comprehension check”).

3c. Sample

With an aim to secure a sample size that is more representative of the general UK population than a convenience sample of university students, we used a paid panel of participants through Prolific Academic. Participants were required to be at least eighteen years old and have UK citizenship. Respondents were paid a fixed fee for participating in the study, provided they completed the survey, which took them around six minutes on average.

We received 446 responses from participants who gave their consent to participate in the research study. These were equally split between males and females¹⁴ and were randomly allocated to the Control, Less Severe Currently Covered and Fatal Currently Covered groups. We excluded two responses that were duplicates and two responses where sex was not specified correctly.¹⁵ We further excluded fifty-three respondents who did not judge the order of severity of selected health conditions as intended,¹⁶ reducing the sample size from 446 to 389 (Control group: $n=121$, Less Severe Currently Covered group: $n=137$, Fatal Currently Covered group: $n=131$). 321 of the remaining 389 subjects passed the status quo comprehension check.¹⁷

Overall, the sample was sufficiently representative of the UK population and well balanced across the treatment conditions. On average, the (remaining) subjects considered the difference in severity between the Nail disease and the Fatal disorder to be around eighty-two and between Paraplegia and the Fatal disorder around twenty-three, out of 100.

3d. Analytical approach

In our research set up, participants who gave no equivalence refusals and no off-scale refusals would typically align with an aggregative approach, such as utilitarianism, prioritarianism, or pluralist egalitarianism. Judgements of participants with no equivalence refusals and two off-

¹⁴ To ensure that the sample was gender balanced, we recruited male and female participants separately.

¹⁵ The participant used his/her unique Prolific Academic number to fill out the survey twice, once as a male, once as a female, which is an indication of cheating.

¹⁶ About half of the excluded responses considered paralysis and fatal disorder to be equally difficult to live with.

¹⁷ We have not excluded these responses from our baseline analysis as their failure to pass the comprehension check could also be due to selecting the wrong answer by accident (and there was no option to go back). They could also think it was a “trick” question (as it was a rather simple one). However, when analysing the experimental data, we accounted for the potential impact of a failed status quo comprehension check on the treatment effect.

scale refusals would typically correspond to leximin. Participants who gave no equivalence refusals and gave an off-scale refusal in Lives vs Nails but not in Lives vs Paraplegia Scenario would be considered to hold limited aggregative views, such as the threshold view or ARC described at the end of Section 1.

When analysing equivalence refusals, we can adopt two classifications. Under a broader classification, we consider an equivalence refusal any response from a participant who both correctly ordered the health conditions (i.e. judged the Nail disease to be less severe than Paraplegia and Paraplegia to be less severe than Fatal Autoimmune Disorder) and indicated in a given scenario that both groups should have the same priority when curing the same number of people. Under a narrower classification, we do not count as equivalence refusals the responses of those who indicated that the difference in severity between the two health conditions, which they judged to be of equal priority, is relatively small¹⁸.

To analyse the impact of the status quo on people's preferences, we calculate the average of the sum of off-scale refusals in each treatment group separately and compare it with the average for the control group.

4. Findings

Figure 1.1 summarizes the observed pattern of responses. It is noteworthy that relatively few respondents align with unlimited aggregation (ten per cent of all included respondents on the broad classification of equivalence refusals; twelve per cent on the narrow classification). At

¹⁸ We considered the conditions to be perceived as close in severity when the difference in a respondent's severity ratings for the respective conditions was ten points or less on a 100-point scale. The ten-point threshold was chosen *post hoc* based on our intuitive assessment, to ensure our estimates of equivalence refusals are reasonably conservative; it could have been somewhat smaller or larger.

the same time, a small share (twelve per cent) conforms to leximin. The most common responses align with limited aggregative views (twenty-five per cent on the broad classification of equivalence refusals, thirty-three per cent on the narrow classification). It is also striking that, even on the narrow classification of equivalence refusals, taken together, the responses that align with either aggregative views, leximin, or limited aggregative views make up only slightly more than half of all responses (fifty-six per cent). This means that just under half of respondents (forty-four per cent) gave a combination of answers that cannot be rationalized by any of the philosophical principles surveyed in Section 1. The main cause of this departure from these philosophical principles was respondents' failure to prioritize by severity.¹⁹ The most common instance of such failure were equivalence refusals, when respondents indicated an equal priority for programmes curing the same number of people with conditions of different severity.

As described above, we used two classifications to identify equivalence refusals. Based on the broader classification, we identified about six per cent of responses to the Lives vs Nails Scenario and about thirty-nine per cent of responses to the Lives vs Paraplegia Scenario as equivalence refusals. When we excluded responses of those who considered paraplegia and the fatal autoimmune disorder to be close in severity, the share of equivalence refusals for the second scenario dropped to twenty-nine per cent.²⁰ We consider this share to be a rather conservative estimate of genuine equivalence refusals, which we take to represent the most frequent case of a failure to prioritize by severity.

¹⁹ A very small share of respondents (one per cent) departed from these principles by prioritizing by severity but then holding that some number of cases of toenail fungus, but no number of cases of paraplegia, can outweigh saving 100 people from the fatal disease.

²⁰ No responses were excluded from the Lives vs Nails scenario as the perceived difference between the nail disease and the fatal autoimmune disorder was always greater than 10 points.

In line with the previous research reviewed in Section 2, we found that (i) equivalence refusals were much more likely when the severity difference was rather small (twenty-nine to thirty-nine per cent) than when it was large (six per cent) and (ii) off-scale refusals were much more likely when the severity difference was large (seventy-one per cent) than when it was small (thirteen per cent).

Figure 1.1. Matrix of responses to prioritization decisions [\[here\]](#)

Total n=389 (100 per cent)

		Lives vs Paraplegia				
		Paraplegia for equal numbers	Equal priority for equal numbers	Lives for equal numbers		
				Paraplegia for a number > 100	Always Lives	
Lives vs Nails	Nails for equal numbers	1 %	0	0	0	
	Equal priority for equal numbers	1 %	4 %	1 %	0	
	Lives for equal numbers	Nails for a number > 100	5 %	6 %	2 %	10 %
		Always Lives	6 %	20 %	8 %	25 %

Notes: Fields in light yellow represent responses with no explanation in terms of the philosophical principles surveyed in Section 1. Fields highlighted with a dashed line represent equivalence refusals. The blue field represents responses that correspond to the leximin principle. Orange fields represent responses that correspond to fully aggregative views: utilitarianism, prioritarianism or pluralist egalitarianism, using broad classification of equivalence refusals (only dark orange) or the narrow one (both light and dark orange). Green fields represent responses that correspond to limited aggregative views such as the threshold view or ARC on the broad (dark green only) or narrow classification of equivalence refusals (both light and dark green). Percentages displayed are rounded, therefore they do not add up to 100 per cent.

We now turn to our embedded experiment, which tested for the influence of the status quo by providing participants with different information on what is currently covered

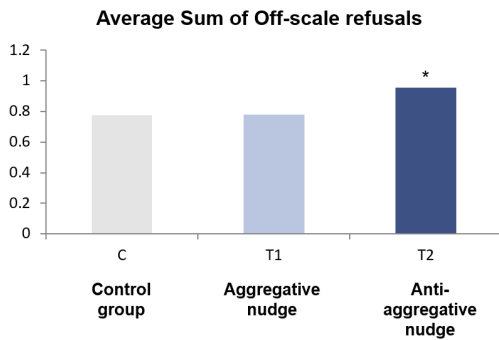
under the health care system. The observed pattern of responses aligns to some extent with our hypothesis that the assumed status quo influences the rate of off-scale refusals. When considering a base model without any covariates and including all respondents who properly ordered the health conditions, telling respondents that the fatal condition was currently covered (but the other conditions were not)—the anti-aggregative nudge—had a statistically significant treatment effect in the predicted direction, i.e. it made off-scale refusals more likely compared to the control ($p=0.028$). However, telling them that the less severe conditions were current covered—the aggregative nudge—had no significant treatment effect. However, when we further analysed the data for the anti-aggregative nudge²¹ and controlled for all and selected sets of covariates, the treatment effect was no longer statistically significant. The model with the highest explanatory power was obtained when the following explanatory variables were included: age, financial situation, self-reported health assessment, and religiosity²² ($n=252$, $p=0.056$, $\text{adj } R^2=0.0288$). Also, when we excluded from the analysis those participants who failed the status quo comprehension check, the treatment effect of the anti-aggregative nudge was not statistically significant ($n=200$, $p=0.06$; see [Figure 1.2](#) and [Table 1.1](#) in [Appendix A](#)). In sum, we find no robust evidence for a status quo effect.

Figure 1.2. Experimental findings - the impact of status quo is as predicted but not robust [\[here\]](#)

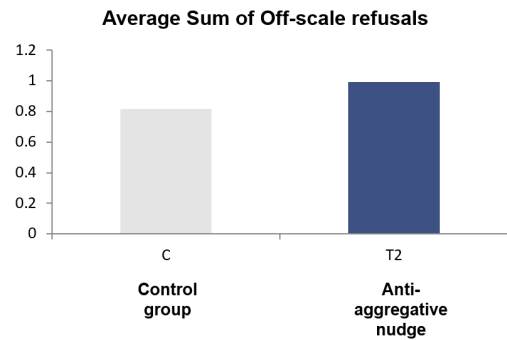
²¹ Comparing the results for Control group and Fatal Currently Covered group

²² Note that our sample was not fully balanced on age and ethnicity. The inclusion of all covariates ($p=0.086$) would have resulted in a sizeable decrease of $\text{adj } R^2$ (-0.002).

Both treatments
Including all participants, $n = 389$



Anti-aggregative nudge only
Including only participants who passed status quo comprehension check, $n = 200$



* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5. Discussion and Conclusion

Our aim in this chapter was to provide empirical evidence on the public’s degree of support for the most commonly used principles of distributive justice in health care. Our study design eliminated some of the shortcomings identified in existing research, such as small, unrepresentative samples, poor framing, and limited spectra of harms. We analysed the responses of a representative sample of the UK population to priority-setting dilemmas about the provision of healthcare resources. The dilemmas were designed to check participants’ conformity with two features of commonly employed principles: (i) prioritization by severity when groups of patients are of the same size and differ substantially in the severity of the harms they face; and (ii) unlimited aggregation when the groups differ in size and in the severity of the harms they face.

In our survey, an overwhelming share of the public (eighty-eight per cent) did not make choices in full conformity with these commonly used principles. These “nonconformists” fall into three camps: forty-four per cent of all subjects chose in ways that clearly depart from prioritization by severity; thirty-three per cent of all subjects prioritized by severity but accepted only a limited form of aggregation, on which only substantial harms

could together outweigh saving a life; and a further twelve per cent accepted prioritization by severity but rejected unlimited aggregation by giving lexical priority to saving lives.

We found the widespread rejection of prioritization by severity to be particularly surprising, as the principle is, to our knowledge, not often disputed in health care priority setting policy discussions and is also relatively uncontroversial in philosophical debates on distributive justice. There are, however, uncommon philosophical views which might motivate at least some such departures from prioritization by severity. For example, Taurek (1977) holds that when a bystander can save either a first group from a harm or a second, equally large group of a smaller harm, this bystander may, to guide their decision-making, adopt the permissible moral perspective of any of the people whose well-being is at stake. Taurek takes this permissible moral perspective to include a significant (though limited) degree of special concern for one's own well-being over the well-being of a stranger—for example, he holds that a person may permissibly prioritize saving themselves from a substantial harm over saving a stranger from an even larger harm such as imminent death. It follows that it is permissible, on a Taurekian view, for the bystander to save either group so long as the harm faced by the group facing the smaller harm is such that an individual facing that harm could permissibly save themselves from this harm rather than a stranger from the larger harm. If one assumes that it is permissible for a person to save themselves from paraplegia rather than a stranger from the fatal autoimmune disorder, but it is not permissible for a person to save themselves from the nail disease rather than a stranger from this fatal disorder, then the upshot of a Taurekian view would be that while it is required to always prioritize the fatal disorder over the nail disease, it is permitted to give equal priority to saving 100 people from paraplegia over saving 100 from the fatal disorder. Such a Taurekian view may therefore account for the around twenty per cent of subjects who give an answer of this

kind while judging that the difference in severity between these conditions is substantial. Indeed, on this Taurekian view, it is also permissible to prioritize the paraplegics over those with a fatal disorder in this case (as a further six per cent of respondents did); just as it is permissible to prioritize those with the fatal disorder (as many other respondents who conform to prioritization by severity did). Given the range of responses that such a Taurekian view permits, our experiment is not well-suited to fully distinguish subjects motivated by this view from those motivated by other views; nonetheless, the fact that it can account for roughly a quarter of the responses that none of the other surveyed views can accommodate makes the question of the degree to which people support it worthy of future investigation.²³

It is also important that thirty-three per cent of respondents favoured limited aggregation. This offers some support for the conjectures of philosophers that many people's morality incorporates a form of limited aggregation. And, to return to the policy in the Netherlands which does not permit small harms to outweigh life-saving treatments mentioned in the Introduction, if one assumes the UK and Dutch publics hold similar views on distributive justice in health, our findings can also bolster the Dutch priority setting authority's previously unsubstantiated claim that their policy has significant public support.

²³ Refusals to prioritize by severity may, of course have other motivations. For instance, a respondent who prioritized saving 100 people from paraplegia over saving 100 people from the fatal autoimmune disorder (while considering living with the later condition to be more difficult than living with the former) argued that "the people with the fatal disease will not suffer indefinitely...the partial paralysis is not fatal, so the suffering would be prolonged without a cure." The respondent chose saving lives over curing nails in Lives vs Nails scenario (with an off-scale refusal) but chose curing paraplegia over saving lives in Lives vs Paraplegia scenario (for equal numbers). The respondent passed the comprehension check and provided the following severity assessments: 9 – 62 – 81 for nail disease, paraplegia, and fatal disorder, respectively, which means there was a sufficient difference in perceived severity between paraplegia and fatal disorder and it is unlikely that paraplegia was judged to be worse than death.

We emphasize, however, that we find substantial pluralism among the views expressed. Indeed, our findings suggest that this pluralism is more extensive than is commonly recognized. Discussions of moral pluralism in health care priority setting sometimes assume that we are operating within the confines of the continuum of views that runs from utilitarianism to leximin; the only dimension of moral pluralism that is then considered, and that public opinion is then surveyed on, is the degree of priority to the worse off (see, e.g. Adler, 2012, pp. 392-399; Dolan & Tsuchiya 2011; Norheim, 2013; Robson et al. 2017). This narrow framing ignores the very large proportion of the public (seventy-six per cent)²⁴ who, in our survey, espouse views that do not align with these core assumptions.

Given that the legitimacy of priority-setting arrangements depends in part on public acceptance of these arrangements, our findings raise the question whether the common practice of using unlimited aggregative distributive principles to set health care priorities has sufficient legitimacy. To us, our findings also suggest that limited aggregative views, such as those embodied in the Dutch priority setting principles, are worthy of consideration, since they have an apparently reasonable philosophical rationale and may have broader public support.

In sum, we submit that empirical investigations like ours can have value in two principal ways. First, they may help us to better understand the public's views and the concomitant question of which policies can be said to be implemented in their name with their support. Second, insofar as they reveal an apparent mismatch between orthodox views and the public's choices, they may encourage further scrutiny of unorthodox views by

²⁴ Those who chose in line with full aggregation is twelve per cent a further twelve per cent aligned with leximin. It follows that seventy-six per cent cannot be placed on the continuum that runs from utilitarianism to leximin.

philosophers and psychologists, to see whether these views have previously unrecognized merits, or whether, instead, these unorthodox views flow from biased or otherwise unreasonable judgements. In this way, empirical studies of the public's views of distributive justice are an impetus to engage in an interdisciplinary and inclusive search for reflective equilibrium.

While our results offer novel insights into the public's views and may serve to stimulate debate about it, it is important to highlight several limitations of our research. First, using a fatal disease as the most severe condition, while common in the literature using the person-trade-off methodology, may have forced respondents to deal with complicated questions regarding comparing the harm of early death with the harms of life with a substantial impairment such as paraplegia. For this reason, it might be advisable in future research to make the most severe condition an enduring, non-fatal one.

Second, although we included several prompts for deliberation, such as the initial severity assessment, using a matching rather than simple choice task, as well as asking respondents to justify their choices, we did not include an element of discussion or qualitative research techniques to clarify people's judgements. Previous research has shown that such discussions or joint deliberations might have an impact on people's preferences (Nord, 1995; Schwarzinger et al., 2003) and that qualitative exploration might uncover potential misinterpretations of people's true judgements (Thompson, 2022).

Third, although our findings show that a much larger share of the public prefers limited aggregation to full aggregation, the design of our study did not allow us to generate insight into people's reasoning about how to draw the line between harms that do and those that do not aggregate against more severe harms.

Fourth, we only used two decisions and three specific conditions. This does not allow us to confidently extrapolate to other conditions or to other contexts of distributive justice.

Despite these limitations, our results indicate that a very large share of the public hold views about the application of distributive justice in healthcare that depart from the principles that are normally used. This invites further empirical investigation of the prevalence and stability of these views, and of people's reasons for holding them. We have outlined several suggestions for further research, which would address some of the limitations of our study and enhance our understanding of people's considered judgements and preferences for the allocation of healthcare resources. This includes providing more space for deliberation and discussion, using multiple scenarios, and a study design that would distinguish between those who use an absolute threshold from those who make the permissibility of aggregation depend on the relative strength of competing claims. This more advanced empirical research should, we propose, go hand in hand with greater philosophical scrutiny of the merits and demerits of views of distributive justice that deny an obligation to help the more severely burdened (when numbers are equal) and that limit the extent to which premature death can be outweighed by lesser burdens.

Bibliography

Adler, M. (2012). *Well-Being and Fair Distribution: Beyond Cost-Benefit Analysis*. Oxford, England: Oxford University Press.

Betram, M., & Edejer, T. T. T. (2021). Introduction to the Special Issue on “The World Health Organization Choosing Interventions That Are Cost-Effective (WHO-CHOICE) Update”. *International Journal of Health Policy and Management*, *10*, 670-672. doi: 10.34172/IJHPM.2021.105

Bognar, G., & Hirose, I. (2014). *The Ethics of Health Care Rationing*. Routledge.

Bostrom, N., & Ord, T. (2006). The reversal test: eliminating status quo bias in applied ethics. *Ethics*, *116*(4), 656-679. doi: 10.1086/505233

Brown, C. (2005). Priority or Sufficiency... or Both? *Economics and Philosophy*, *21*(2), 199-220. doi: 10.1017/S0266267105000568

Brown, C. (2020). Is Close Enough Good Enough? *Economics and Philosophy*, *36*, 29–59. doi: 10.1017/S0266267119000099

Caviola, L., Mannino, A., Savulescu, J., & Faulmüller, N. (2014). Cognitive biases can affect moral intuitions about cognitive enhancement. *Frontiers in systems neuroscience*, *8*(195), 1-5. doi: 10.3389/fnsys.2014.00195

Crisp, R. (2003). Equality, Priority, and Compassion. *Ethics*, *113*(4), 745-763. doi: 10.1086/373954

- Damschroder, L. J., Roberts, T. R., Goldstein, C. C., Miklosovic, M. E., & Ubel, P. A. (2005). Trading people versus trading time: What is the difference?. *Population health metrics*, 3(1), 10. doi: 10.1186/1478-7954-3-10
- Damschroder, L. J., Roberts, T. R., Zikmund-Fisher, B. J., & Ubel, P. A. (2007). Why people refuse to make tradeoffs in person tradeoff elicitation: A matter of perspective? *Medical Decision Making*, 27(3), 266-280. doi: 10.1177/0272989x07300601
- Daniels, N. (2013). "Reflective Equilibrium." The Stanford Encyclopedia of Philosophy (Winter 2013 Edition), editeded.). Edited by Edward N. Zalta (ed.), URL = <
<http://plato.stanford.edu/archives/win2013/entries/reflective-equilibrium/>>./.
- Dolan, P., & Green, C. (1998). Using the person trade-off approach to examine differences between individual and social values. *Health Economics*, 7(4), 307-312. doi: 10.1002/(SICI)1099-1050(199806)7:4<307::AID-HEC345>3.0.CO;2-N
- Dolan, P., & Tsuchiya A. (2011). Determining the parameters in a social welfare function using stated preference data: an application to health. *Applied Economics*, 43, 2241–2250. doi: 10.1080/00036840903166244
- Dolan, P., Olsen, J. A., Menzel, P., & Richardson, J. (2003). An inquiry into the different perspectives that can be used when eliciting preferences in health. *Health Economics*, 12(7), 545-551. doi: 10.1002/hec.760
- Dolan, P., & Robinson, A. (2001). The measurement of preferences over the distribution of benefits: the importance of the reference point. *European Economic Review*, 45, 1697-1709. doi: 10.1016/s0014-2921(00)00052-0

- Fernandez, R., & Rodrik, D. (1991). Resistance to Reform: Status Quo Bias in the Presence of Individual- Specific Uncertainty. *The American Economic Review*, 81(5), 1146-1155. doi: 10.1257/0002828041464425
- Fleurbaey, M., Tungodden, B., & Vallentyne, P. (2009). On the Possibility of Nonaggregative Priority for the Worst Off. *Social Philosophy and Policy*, 26, 258-285. doi: 10.1017/s0265052509090116
- Horton, J. (2018). Always Aggregate. *Philosophy and Public Affairs* 46(2), 160-74. doi: 10.1111/papa.12116
- Johnson, E., Hershey, J., Meszaros, J., & Kunreuther, J. (1993). Framing, probability distortions, and insurance decisions. *Journal of Risk and Uncertainty*, 7(1), 35-51. doi: 10.1007/978-94-011-2192-7_3
- Kahneman, D., Knetsch, J., & Thaler, R. (1991). Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias. *Journal of Economic Perspectives*, 5(1), 193-206. doi: 10.1017/cbo9780511803475.009
- Kamm, F. M. (1993). *Morality, Mortality, Vol. I*. Oxford: Oxford University Press.
- Kelleher, P. (2014). Relevance and Non-Consequentialist Aggregation. *Utilitas*, 26(4), 385–408. doi: 10.1017/S0953820814000144
- Lefkowitz, D. (2008). On the Concept of a Morally Relevant Harm, *Utilitas*, 20(4), 409–423. doi: 10.1017/s0953820808003245
- Mann, K. (2021a). The Relevance View: Defended and Extended. *Utilitas*, 33(1), 101-110. doi: 10.1017/s095382082000028x

- Mann, K. (2021b). Relevance and Nonbinary Choices. *Ethics*, 132(2), 382-413. doi: 10.1086/716873
- Millum, J. (2015). Age and Death: A Defense of Gradualism. *Utilitas*, 27(3), 279-297. doi: 10.1017/s0953820815000047
- Liao, S.M. & Lim, J. E. (2022). Lives, Limbs, and Liver Spots: The Threshold Approach to Limited Aggregation. Manuscript shared by author.
- Morell, V. (1990). Oregon puts bold health plan on ice. *Science*, 249, 468-471. doi: 10.1126/science.2382125
- Nagel, T. (1979). *Mortal Questions*. Cambridge, England: Cambridge University Press
- ~~Nebel, J. M. (2015). Status quo bias, rationality, and conservatism about value. *Ethics* 125(2), 449-476. Doi: 10.1086/678482~~
- Nord, E. (1995). The person-trade-off approach to valuing health care programmes. *Medical decision making*, 15(3), 201-208. doi: 10.1177/0272989x9501500302
- ~~Nord, E., & Johansen, R. (2014). Concerns for severity in priority setting in health care: a review of trade-off data in preference studies and implications for societal willingness to pay for a QALY. *Health policy*, 116(2-3), 281-288. doi: 10.1016/j.healthpol.2014.02.009~~
- Nord, E., Richardson, J., & Macarounas-Kirchmann, K. (1993). Social evaluation of health care versus personal evaluation of health states: evidence on the validity of four health-state scaling instruments using Norwegian and Australian surveys. *International journal of technology assessment in health care*, 9(4), 463-478. doi: 10.1017/s0266462300005390

- Norheim, O.F. (2013). Atkinson's Index Applied to Health. Can Measures of Economic Inequality Help Us Understand Tradeoffs in Health Care Priority Setting? In N. Eyal, S. A. Hurst, O. F. Norheim, & D. Wikler (Eds), *Inequalities in Health: Concepts, Measures, and Ethics* (pp. 214-230). Oxford, England: Oxford University Press.
- Otsuka, M. (2004). Skepticism about Saving the Greater Number. *Philosophy and Public Affairs*, 32(4), 413–426. doi: 10.1111/j.1088-4963.2004.00020.x
- Otsuka, M. (2006). Saving Lives, Moral Theory, and the Claims of Individuals. *Philosophy and Public Affairs* 34(2), 109–135. doi: 10.1111/j.1088-4963.2006.00058.x
- Ottersen, T., Førde, R., Kakad, M., Kjellevoid, A., Melberg, H. O., Moen, A., ... Norheim, O.F. (2016). A new proposal for priority setting in Norway: Open and fair. *Health Policy*, 120(3), 246-251, doi: 10.1016/j.healthpol.2016.01.012
- Payne, J., Bettman, J., & Johnson, E. (1992). Behavioral Decision Research: A Constructive Processing Perspective. *Annual Review of Psychology*, 43(1), 87-131. doi: 10.1146/annurev.ps.43.020192.000511
- Parfit, D. (1995). *Equality or Priority?* The Lindley Lecture. The University of Kansas.
- Pinto-Prades, J. L., & Lopez-Nicolas, A. (1998). More evidence of the plateau effect: a social perspective. *Medical Decision Making*, 18(3), 287-294. doi: 10.1177/0272989x9801800306
- Robson, M., Asaria, M., Cookson, R., Tsuchiya, A. & Ali, S. (2017). Eliciting the level of health inequality aversion in England. *Health Economics*, 26(10), 1328–1334. doi: 10.1002/hec.3430

- Rueger, K. (2020). Aggregation with Constraints. *Utilitas*, 32(4), 454-471. doi: 10.1017/s095382082000014x
- Salkeld, G., Ryan, M., & Short, L. (2000). The veil of experience: Do consumers prefer what they know best? *Health Economics*, 9(3), 267-270. doi: 10.1002/(sici)1099-1050(200004)9:3<267::aid-hec511>3.0.co;2-h
- Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, 1(1), 7-59. doi: 10.1007/bf00055564
- Scanlon, T.M. (1998). *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
- Schwarzinger, M., Lanoë, J., Nord, E., & Durand-Zaleski, I. (2004). Lack of multiplicative transitivity in person trade-off responses. *Health Economics*, 13(2), 171-181. doi: 10.1002/hec.808
- Schwarzinger, M., Stouthard, M. EA., Burström, K., & Nord, E. (2003). Cross-national agreement on disability weights: The European Disability Weights Project. *Population Health Metrics*, 1(1), 9. doi: 10.1186/1478-7954-1-9
- Singh, J., Lord, J., Longworth, L., Orr, S., McGarry, T., Sheldon, R., & Buxton, M. (2012). Does responsibility affect the public's valuation of health care interventions? A relative valuation approach to health care safety. *Value in Health*, 15(5), 690-698. doi: 10.1016/j.jval.2012.02.005
- Skedgel, C. D. (2013). *Estimating societal preferences for the allocation of healthcare resources using stated preference methods*. (Doctoral dissertation). University of Sheffield.

- Slovic, P., & Västfjäll, D. (2013). The more who die, the less we care: Psychic numbing and genocide. In A. Oliver (Ed.), *Behavioural Public Policy* (pp. 94-114). Cambridge, England: Cambridge University Press. doi: 10.1017/cbo9781107337190.005
- Sparrow, R. (2015). Imposing genetic diversity. *The American Journal of Bioethics*, 15(6), 2–10. doi: 10.1080/15265161.2015.1028658
- Steuwer, B. (2021). Aggregation, Balancing and Respect for the Claims of Individuals. *Utilitas*, 33(1), 17-34. doi: 10.1017/s0953820820000217
- Tadros, V. (2019). Localized Restricted Aggregation. In D. Sobel, P. Vallentyne, & S. Wall (Eds.), *Oxford Studies in Political Philosophy, Vol. 5* (pp. 171-204). Oxford, England: Oxford University Press. doi: 10.1093/oso/9780198841425.003.0007
- Taurek, J. (1977). Should the Numbers Count? *Philosophy and Public Affairs*, 6, 293–316.
- Temkin, L. (2012). *Rethinking the Good*. Oxford, England: Oxford University Press.
- Thompson, K. (2022). Qualitative methods show that surveys misrepresent “ought implies can” judgments. *Philosophical Psychology*, 1-29. doi: 10.1080/09515089.2022.2036714
- Tomlin, P. (2017). On Limited Aggregation. *Philosophy and Public Affairs* 45(3), 232-260. doi: 10.1111/papa.12097
- Tungodden, B. (2003). The value of equality. *Economics and Philosophy*, 19(1), 1–44.
- Ubel, P. A., Loewenstein, G., Scanlon, D., & Kamlet, M. (1996). Individual utilities are inconsistent with rationing choices: A partial explanation of why Oregon’s cost

- effectiveness list failed. *Medical Decision Making*, 16(2), 108–116. doi: 10.1177/0272989x9601600202
- Van de Wijngaard, Q. (2021). *Priority-setting in the Netherlands: A Case for Public Participation and Against Proportional Shortfall*. (Master's thesis), Erasmus University Rotterdam.
- Voorhoeve, A. (2014). How Should We Aggregate Competing Claims? *Ethics*, 125(1), 64–87. doi: 10.1086/677022
- Voorhoeve, A. (2017). Why One Should Count Only Claims with which One Can Sympathize. *Public Health Ethics* 10(2), 148–156. Doi: 10.1093/phe/phw006
- Voorhoeve, A. (2018). Balancing Small against Large Burdens. *Behavioural Public Policy*, 2(1), 125–142. doi: 10.1017/bpp.2017.4
- Voorhoeve, A. (2020). Healthy Nails versus Long Lives: An Analysis of a Dutch Priority Setting Proposal. In N. Eyal, S. Hurst, C. Murray, S.A. Schroeder, and D. Wikler (Eds.), *Measuring the Global Burden of Disease: Philosophical Dimensions* (pp. 273-292). doi: 10.1093/med/9780190082543.003.0016
- Walen, A. (2020). Risks and Weak Aggregation: Why Different Models of Risk Suit Different Types of Cases. *Ethics*, 131, 62-86. doi: 10.1086/709985
- Wasserman, D. (2015). Disability, diversity, and preference for the status quo: bias or justifiable preference? *The American Journal of Bioethics*, 15(6), 11-12. doi: 10.1080/15265161.2015.1028676
- ZiN. [Zorginstituut Nederland] (2015a). Van goede zorg verzekerd: Hoe zorginstituut Nederland adviseert over de inhoud van het basispakket [Good care assured: How the

Netherlands Healthcare Institute advises on the content of the basic package]. (pp. 1-25, Rep.). Diemen: Zorginstituut Nederland.

ZiN. (2015b). Kosteneffectiviteit in de praktijk [Cost-effectiveness in practice]. (pp. 1-19, Rep.). Diemen: Zorginstituut Nederland.

ZiN. (2017). Pakketadvies in de praktijk: Wikken en wegen voor een rechtvaardig pakket [Package advice in practice: Deliberation and weighing for a fair package]. (pp. 1-25, Rep.). Diemen: Zorginstituut Nederland.

ZiN. (2018). Ziektelast in de praktijk: De theorie en praktijk van het berekenen van ziektelast bij pakketbeoordelingen [Burden of disease in practice: Theory and practice of calculating the burden of disease in package assessments]. (pp. 1-25, Rep.). Diemen: Zorginstituut Nederland.

ZiN. (2020). *Beoordeling Van nieuwe zorg* [Review of new care]. Retrieved from <https://www.zorginstituutnederland.nl/over-ons/werkwijzen-en-procedures/adviseren-over-en-verduidelijken-van-het-basispakket-aan-zorg/beoordeling-van-nieuwe-zorg>

Appendix

Table 1.1 Results of regression analyses

Regression models: The Sum of Off-Scale Refusals, all models for Treatment 2, including all participants (models 1 – 4) and including only those participants who passed the comprehension check (models 5 – 7)

(1) Base model (2) Basic demographics (3) Best model (in terms of the highest adjusted R²) (4) All covariates (5) Base model (6) Best model (7) All covariates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Treatment	0.177*	0.161	0.150	0.147	0.174	0.134	0.109
	(0.0800)	(0.0835)	(0.0825)	(0.0854)	(0.0921)	(0.0982)	(0.103)
Sex		0		0			0
		(.)		(.)			(.)
2		-0.0125		-0.0517			-0.0966

		(0.0828)		(0.0861)		(0.0996)
Age	0	0	0	0	0	0
	(.)	(.)	(.)	(.)	(.)	(.)
2	-0.0203	0.0318	-0.0363	-0.0460	-0.147	
	(0.0993)	(0.0986)	(0.106)	(0.116)	(0.125)	
3	0.230*	0.267*	0.233	0.186	0.141	
	(0.115)	(0.117)	(0.123)	(0.137)	(0.143)	
4	0.0465	0.0801	-0.0104	0.0147	-0.114	
	(0.178)	(0.175)	(0.188)	(0.198)	(0.212)	
5	0.138	0.195	0.104	0.100	-0.00639	
	(0.462)	(0.457)	(0.470)	(0.471)	(0.481)	
Ethnicity	0		0		0	

	(.)	(.)	(.)
2	0.182 (0.328)	0.0963 (0.341)	0.159 (0.473)
3	-0.411 (0.330)	-0.342 (0.339)	-0.479 (0.407)
4	0.0105 (0.223)	0.0793 (0.246)	0.140 (0.275)
5	0.162 (0.685)	0.249 (0.700)	0.175 (0.723)
Education	0 (.)	0 (.)	0 (.)
2	0.131	0.0557	0.122

		(0.102)		(0.108)		(0.130)
3		0.101		0.00612		0.0514
		(0.138)		(0.147)		(0.168)
4		0.129		0.118		0.217
		(0.245)		(0.247)		(0.274)
5		0.158		0.0320		-0.0572
		(0.183)		(0.194)		(0.227)
Financial		0	0	0	0	0
situation		(.)	(.)	(.)	(.)	(.)
2		-0.285	-0.298	-0.282	-0.349	-0.313
		(0.188)	(0.185)	(0.192)	(0.216)	(0.225)

3	-0.211 (0.170)	-0.224 (0.171)	-0.213 (0.176)	-0.259 (0.202)	-0.237 (0.206)
4	-0.313 (0.179)	-0.311 (0.179)	-0.314 (0.186)	-0.346 (0.211)	-0.320 (0.219)
5	0.00162 (0.250)	0.0882 (0.251)	0.0893 (0.259)	-0.139 (0.290)	-0.130 (0.298)
6	-0.329 (0.360)	-0.305 (0.358)	-0.266 (0.368)	-0.399 (0.381)	-0.378 (0.391)
7		0 (.)	0 (.)	0 (.)	0 (.)
Religion		0.0793 (0.264)	0.0996 (0.275)	0.0610 (0.312)	0.111 (0.321)

2		-0.123 (0.250)	-0.0819 (0.263)	-0.119 (0.296)	-0.128 (0.306)
3		-0.0882 (0.234)	-0.111 (0.253)	-0.0461 (0.272)	-0.100 (0.291)
4		-0.168 (0.208)	-0.193 (0.229)	-0.136 (0.240)	-0.177 (0.262)
5		-0.877 (0.497)	-0.935 (0.531)	-0.933 (0.522)	-1.144* (0.570)
Health		0 (.)	0 (.)	0 (.)	0 (.)
2		-0.284* (0.137)	-0.248 (0.145)	-0.225 (0.158)	-0.189 (0.166)

3		-0.195 (0.145)	-0.161 (0.152)	-0.119 (0.166)	-0.0710 (0.173)
4		-0.266 (0.181)	-0.222 (0.191)	-0.156 (0.212)	-0.0885 (0.225)
5		-0.375 (0.198)	-0.311 (0.209)	-0.217 (0.239)	-0.170 (0.252)
Political views			0 (.)		0 (.)
2			-0.110 (0.133)		-0.113 (0.164)
3			0.0127 (0.171)		-0.0792 (0.204)

4					-0.186 (0.154)			-0.330 (0.188)
5					-0.0738 (0.221)			-0.00108 (0.279)
_cons	0.777*** (0.0577)	0.879** (0.276)	1.323*** (0.275)	1.362*** (0.382)	0.817*** (0.0707)	1.377*** (0.311)	1.423** (0.439)	
<i>N</i>	252	252	252	252	200	200	200	
adj. R^2	0.015	0.006	0.029	-0.002	0.013	-0.007	-0.018	

Standard errors in parentheses
 * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$