OHE

Understanding societal preferences for priority by disease severity in England & Wales

CONTRACT RESEARCH REPORT
NOVEMBER 2024

Helen Hayes
Priscila Radu
David Mott
Chris Skedgel

ohe.org

NOVEMBER 2024

Understanding societal preferences for priority by disease severity in England & Wales

Helen Hayes
Office of Health Economics, London

Priscila Radu
Office of Health Economics, London

David Mott
Office of Health Economics, London

Chris Skedgel
Office of Health Economics, London

Please cite this report as:
Hayes H, Radu P, Mott D, Skedgel C, 2024. Understanding societal preferences for priority by disease severity in England & Wales. OHE Contract Research Report, London: Office of Health Economics. Available at: https://www.ohe.org/publications/understanding-societal-preferences-for-priority-by-disease-severity-in-england-wales/

Corresponding Author:
Chris Skedgel
cskedgel@ohe.org

About OHE Contract Research Reports

Many of the studies OHE performs are proprietary and the results are not released publicly. Studies of interest to a wide audience, however, may be made available, in whole or in part, with the client's permission. They may be published by OHE alone, jointly with the client, or externally in scholarly publications. Publication is at the client's discretion.

Studies published by OHE as OHE Contract Research Reports are subject to internal quality assurance and undergo external review, usually by a member of OHE's Editorial Panel. Any views expressed are those of the authors and do not necessarily reflect the views of OHE as an organisation.

Funding and Acknowledgements

This consulting report was commissioned and funded by *The Association of the British Pharmaceutical Industry (ABPI)*.

We gratefully acknowledge our independent project advisors, Dr Vivian Reckers-Droog of Erasmus University, and Professor Peter Moffatt of the University of East Anglia, for their advice and guidance around the design of the study and the interpretation of the results. We also acknowledge Paul Schneider of Valorem Health for his outstanding support with survey programming and administration, and Professor Richard Norman of Curtin University for early input on our approach and study design.

For further information please contact:

Dr Chris Skedgel
Director, OHE
Honorary Research Fellow, University of East Anglia
Visiting Scientist, Medical University of Vienna

Tel +44 (0)207 747 7015
Email cskedgel@ohe.org

Table of Contents

Executive Summary ..... **iv**
1 Background **7**
2 Methods **11**
2.1 Stated preference task format 11
2.2 Survey development 11
2.2.1 QALY shortfall presentation and questionnaire wording 11
2.2.2 The final questionnaire 12
2.3 Data 17
2.3.1 Outcome variables 17
2.3.2 Demographic data 17
2.3.3 Value statements 18
2.4 Statistical analysis 18
2.4.1 Sample summary 18
2.4.2 Cut-offs 18
2.4.3 Shortfall (severity) weights 18
3 Results **20**
3.1 Sample summary 20
3.1.1 Demographic data 20
3.1.2 Value statements 22
3.2 Severity Cut-offs 23
3.3 Shortfall Weights 24
3.4 Sensitivity checks 26
4 Discussion **27**
References **30**
Appendices **32**

Executive Summary

Background

- The National Institute for Health and Care Excellence (NICE)'s severity modifier assigns greater weight to health gains to patients with more severe diseases, making it more likely that treatments for those diseases will be recommended for coverage on the National Health Service (NHS).
- For a disease to qualify for the severity modifier, patients must be expected to lose substantial amounts of their future health. NICE, however, has provided little explanation of how the eligibility cutoffs and weights were determined, or how they may be aligned with societal preferences.
- We surveyed a representative sample of the population of England and Wales to understand how the public would prioritise health gains for more severe diseases.

Key findings

- We find that NICE's current severity modifier does not appear to be well aligned with the public's preference for prioritising health gains for more severe diseases in England and Wales.
- Members of the general population give priority at a substantially lower severity threshold, and the public assigns greater relative value to health gains at almost every level of severity, compared with NICE's current severity modifier.
- If NICE seeks to align the value and priority assigned to new medicines and technologies with societal preferences, these results suggest a need for NICE to reassess its criteria for the severity modifier.

How does NICE's severity modifier work?

NICE implemented a new severity modifier in February 2022 as part of a wider update to its methods. Its aim was to broaden the definition of severity beyond the existing end-of-life severity modifier to also consider improvements in quality of life.

The revised approach considers two different – but related – measures of disease severity: absolute shortfall (AS) and proportional shortfall (PS). Greater value is assigned to health gains for patients with greater absolute (AS) or proportional (PS) health shortfalls, based on whichever shortfall is greater.

Under NICE's current criteria, patients who lose a substantial amount of their future health can qualify for a value 'multiplier' of either 1.2 or 1.7. NICE applies this multiplier to health gains when calculating the cost-effectiveness of a treatment. This has the effect of *increasing* the quantity of quality-adjusted life years (QALYs) gained with treatment and

decreasing the effective cost per QALY gained, making a recommendation by NICE more likely relative to treatments for conditions that do not qualify for the modifier.

Patients expected to lose between 85% and 95% of their (discounted) expected lifetime quality-adjusted life years (QALYs) or between 12 and 18 (discounted) QALYs relative to an individual without the disease receive a multiplier of 1.2.

Patients who are expected to lose more than 95% of their (discounted) expected lifetime QALYs or more than 18 (discounted) QALYs relative to an individual without the disease receive a multiplier of 1.7.

	No multiplier	**1.2 value multiplier**	**1.7 value multiplier**
Absolute shortfall (AS)	<12 discounted QALYs	12-18 discounted QALYs	≥18 discounted QALYs
Proportional shortfall (PS)	<85% PS	85% - 95% PS	≥95% PS

Gaining a better understanding of UK societal preferences

We conducted a quantitative preference study to understand how well the criteria of NICE's severity modifier align with UK societal preferences. Our primary objectives were to estimate:

1. Cut-offs for 'severe' and 'very severe' categories by AS and PS.
2. The relative value of health gains across the range of possible shortfall scores, in comparison to a low-severity (20% PS) group.

We used a Person Trade-Off (PTO) approach to understand the value of health gains to patients with a greater versus lesser future health in an age-sex representative sample of the England & Wales general population (complete case analysis N=990). We also elicited their views on severity thresholds and key principles of resource allocation.

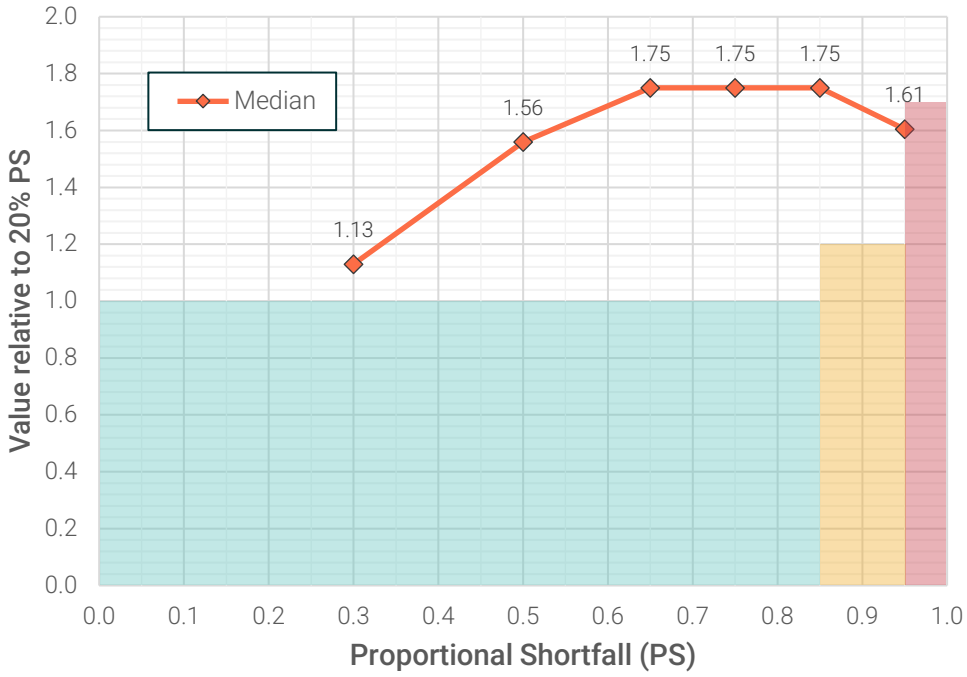
What our research tells us about the UK values health gains and losses

We found that members of the general population give priority at a substantially lower shortfall thresholds, and that the public assigns greater relative value to health gains to more severe conditions at almost every level of severity, compared with NICE's current severity modifier.

We identified the shortfall 'cutoffs' for what respondents considered "severe" or "very severe". On the proportional shortfall (PS) scale, the public indicated that 'severe' health states started around 50% PS (compared to NICE's 85%) and that 'very severe' health states begin around 65% (compared to NICE's 95%).

We also found a fairly rapid increase in the relative value of treating more severe patient groups, even at relatively moderate levels of shortfall, and that relative value plateaus at about 1.7 beyond a PS of 65%. This plateau in value around 65% PS is consistent with the

cutoff identified in a separate task and reinforces the notion of critical threshold in public preferences at this shortfall.



** Excludes mean weights. See Figure 11 in the main text for mean and median weights.*

Policy implications and next steps

Our results suggest that NICE's current severity modifier is not well aligned with the UK public's preference for prioritising health gains for more severe diseases. If NICE seeks to align the value and priority assigned to new medicines and technologies with societal preferences, these results suggest a need for NICE to reassess its criteria for the severity modifier.

1 Background

Some Health Technology Assessment (HTA) bodies are adopting value modifiers that explicitly assign additional value to health gains that accrue to patients in more severe health states (Zhang and Garau, 2020). Norway, for example, has (informally) implemented a modifier that increases the acceptable cost-effectiveness threshold (CET) according to categories of "absolute shortfall" (AS) (Norwegian Ministry of Health and Care Services, 2017). AS is the difference between the expected quantity of quality-adjusted life years (QALYs) remaining *with a disease and with the standard of care* (SoC) but *not* the new treatment, and the expected quantity of QALYs remaining *in the absence of the disease* (Skedgel et al., 2022):

$$\text{Absolute shortfall (AS)} = \text{Remaining QALY expectation in absence of disease} - \text{Remaining QALY expectation with disease and SoC}$$

The original proposal for a Norwegian severity modifier ("the Magnussen Committee"), suggested 6 severity classes, between an AS of less than 4 QALYs to more than 20 QALYs, and that the maximum CET should be three times the baseline threshold (Norwegian Ministry of Health and Care Services, 2017). The acceptable CET increases from 275,000 to 825,000 Norwegian Kroner (approximately £19,300 to £58,000) over this range.

The Netherlands has implemented a similar modifier based on "proportional shortfall" (PS), where PS represents disease-related QALY loss (i.e. AS) as a proportion of expected QALYs remaining in the absence of disease (Skedgel et al., 2022):

$$\text{Proportional shortfall (PS)} = \frac{\text{Disease-related QALY loss with SoC (AS)}}{\text{Remaining QALY expectation in absence of disease}}$$

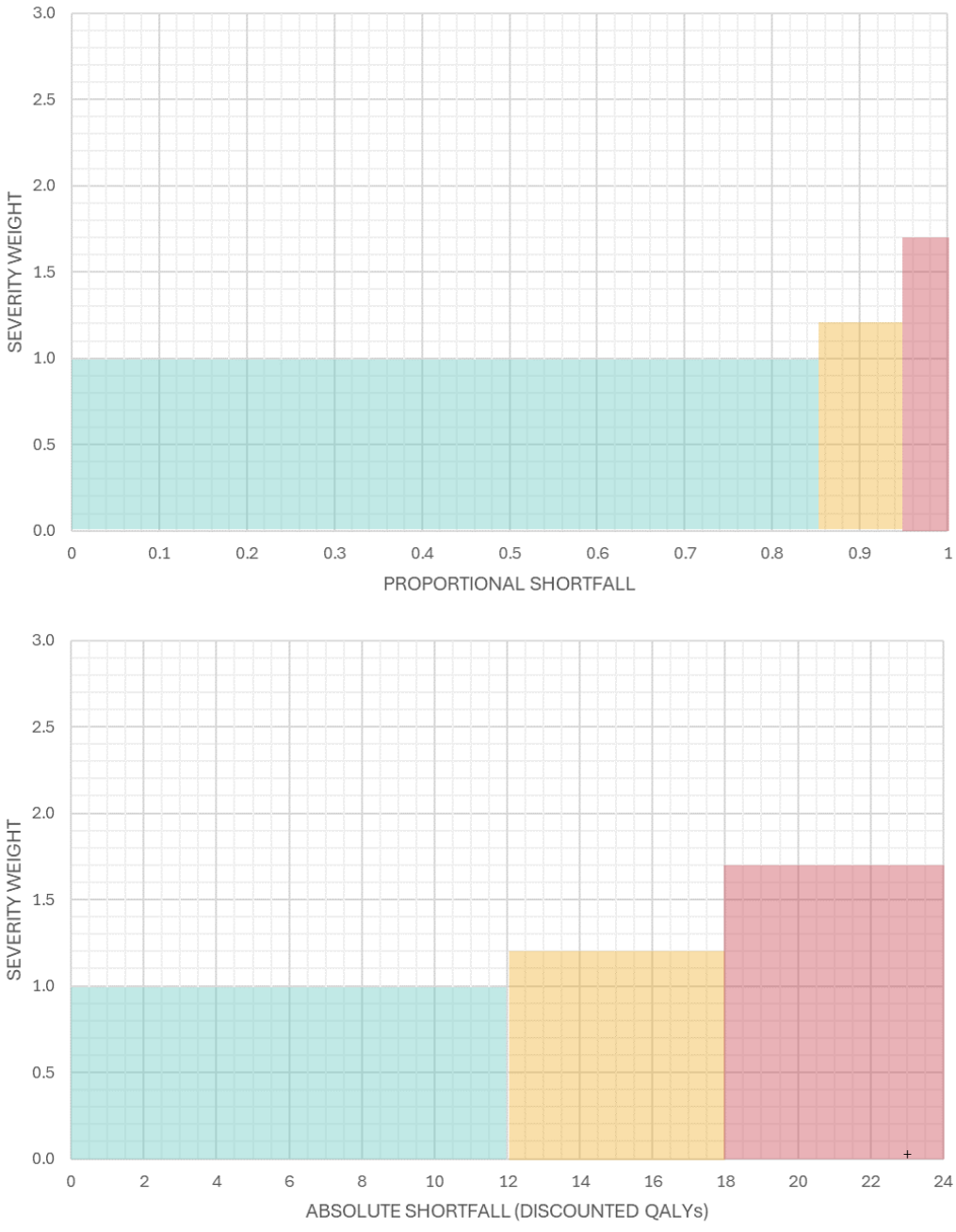
The Zorginstituut Nederland (ZIN) in the Netherlands divides PS into 4 categories: <10% PS, 11-40% PS, 41-70% PS and >71% PS. Technologies that target conditions with a PS less than 10% are, in principle, not considered for reimbursement, whilst the other categories are assessed against CETs of €20,000, €50,000 and €80,000 per QALY gained, respectively (approximately £17,000, £42,000 and £67,000) (Reckers-Droog, Van Exel and Brouwer, 2019).

As we have noted elsewhere (Skedgel et al., 2022), both approaches to capturing shortfall potentially advantage or disadvantage different age groups. As AS is constrained by remaining life expectancy, it may disadvantage older patients, who will tend to have less remaining life expectancy to lose than younger patients. Indeed, "for older patients, it becomes increasingly difficult, if not impossible, to fall into the highest severity class. For example, 65-year-old patients may, on average, have no more than 15 QALYs left to lose" (Reckers-Droog, Van Exel and Brouwer, 2019). Conversely, as PS tends to increase as remaining life expectancy decreases, it may disadvantage younger patients, who will tend to have more remaining life expectancy.

There is some anecdotal evidence that ZIN sees consideration of PS when assessing severity, which may be more likely to disadvantage (typically younger) patients with greater life expectancy, as balancing, to some degree, the use of QALY gains as a measure of value, which disadvantages (typically older) patients with less life expectancy. The National Institute of Health and Care Excellence (NICE) in the United Kingdom (UK) has taken a different approach by assessing severity on the basis of AS and PS. Conditions that have an absolute shortfall of 12 to 18 discounted QALYs *or* a proportional shortfall of 85% to 95% qualify for a 'moderate' QALY multiplier of 1.2, whilst conditions with an absolute shortfall greater than 18 QALYs or greater than 95% qualify for a 'severe' QALY multiplier of 1.7 (NICE, 2022). Unlike Norway and the Netherlands, NICE adjusts the quantity

(value) of the QALY gains rather than the acceptable CET.

FIGURE 1: NICE'S ABSOLUTE AND PROPORTIONAL SEVERITY MODIFIERS



We note that other HTA bodies, including the Canadian Drugs Agency and the Institute for Comparative-Effectiveness Research in the United States, acknowledge the importance of contextual factors such as severity in their decisions, but have not, to date, adopted explicit severity modifiers (Skedgel et al., 2022).

In general, assigning greater value to health gains to patients in more severe health conditions is consistent with evidence from studies of societal preferences. Reviews by Shah (2009) and Gu *et al.* (2015) found that public respondents in different countries were broadly supportive of giving greater value or priority to smaller gains to individuals in more severe conditions compared with larger gains in less severe conditions. Likewise, Nord and Johansen (2014) found that, *on average*, a 1-unit health gain to an individual in a more severe health state was as valuable as a 2- to 3-unit gain to an individual in a less severe state. However, the full range of their estimates was from <1 (gains to the more severely ill are worth less than gains to the less severely ill) to more than 100 times as valuable. They suggested that this large variation was due to different methodologies and samples across the studies, and that these divergent estimates may indicate cross-country variation in the strength of preferences over severity.

Richardson *et al.* (2017) formally tested what they called the "severity hypothesis," or the idea that health gains to individuals in more severe conditions are relatively more valuable, using a measure of societal willingness to pay (WTP). Consistent with their hypothesis, their estimates indicated a nonlinear value function that was steeper at lower levels of utility (greater immediate ill health) and shallower at higher levels of utility, implying greater willingness to pay for marginal health gains in more severe conditions. Notably, although this slope was relatively steep at the lowest utilities (greatest severity), it levelled off quickly, suggesting that societal willingness to prioritise severity dissipates relatively quickly, even at relatively poor levels of immediate health. Similarly, Recker-Droog *et al.* (2021) found that willingness to pay for health gains in patients nearer the end of their life expectancy was lower than for patients with a greater life expectancy.

This evidence of a societal preference for giving at least some priority to patients in more severe health states is important in how we, as societies, conduct HTA. Specifically, extra-welfarist economic theory (Culyer, 1991; Brouwer et al., 2008), as well as democratic and communitarian principles (Mooney, 1998), hold that HTA decisions should reflect the values and preferences of the society on whose behalf they are acting (and by whom they were directly or indirectly appointed).

Beyond accounting for a preference for giving some priority to patients in more severe health states, it is also important to understand how much efficiency in maximising aggregate health gains society is willing to sacrifice to achieve a fairer or more equitable allocation of resources. This is known as the acceptable 'equity-efficiency trade-off'. As noted above, for example, Nord and Johansen (2014) found that respondents were willing to forego 2 to 3 units of health gain to less severe patients to prioritise a 1-unit gain to patients in a more severe state. A severity modifier that went beyond this acceptable trade-off – say, foregoing 4 units of health in exchange for 1 unit in a more severe condition – would not be aligned with respondent preferences. Likewise, a modifier that limited this trade-off to, say, foregoing 1.5 units of gain to less severe patients to prioritise a 1-unit gain to a more severe patients, would also not be consistent with respondent preferences.

Thus, whilst the existence of *some* explicit priority for patients in more severe health states is broadly consistent with evidence of societal preferences, it is not in itself sufficient to claim that the characteristics and criteria of a value modifier reflect societal preferences, particularly in terms of the acceptable equity-efficiency trade-off. As Hausman (2024) has argued, NICE has provided little empirical justification for its specific eligibility criteria or weights.

Therefore, to address the question of whether NICE's severity modifier is consistent with UK public preferences, we conducted a preference elicitation exercise to understand the critical severity thresholds – that is, the degree of absolute and proportional shortfall at which the public believe that patients deserve some additional priority – and the relative value of health gains at specific levels of shortfall.

Our primary objectives were, first, to estimate cut-offs or thresholds for ‘severe’ and ‘very severe’ categories by absolute and proportional shortfall and compare these to NICE’s current cutoffs, and second, to estimate the relative value of QALY gains across the range of QALY shortfall severity levels, in comparison to a low-severity group, and to compare these to NICE’s current weights or ‘value multipliers’.

2 Methods

We used stated preference methods, informed by qualitative interviews, to elicit preferences for eligibility for a severity value modifier and relative value weights over future health shortfall from an age-sex representative sample of public respondents in England and Wales. We additionally collected information on their views on a series of healthcare prioritisation statements and respondent characteristics. We describe our process of survey development and administration below

2.1 Stated preference task format

Our objective was to understand the relative value of health gains to patients with different levels of health shortfall. In this context, we considered discrete choice experiment (DCE) or person trade-off (PTO) approaches.

We judged that a DCE would be useful for understanding how societal preferences over severity were attenuated by patient characteristics such as age, sex or source of health shortfall (morbidity or mortality), similar to the approach taken by Reckers-Droog, van Exel and Brouwer (2019). However, we ultimately judged that a 'context-free' estimate of preferences over severity, based on a PTO, would be more appropriate for NICE's modifier, given NICE's strict prohibitions on discriminating by personal characteristics. PTO has recently been used to understand the relative value of health gains to children and young people versus adults (Peasgood et al., 2024), to carers versus patients (Al-Janabi et al., 2022), to different socio-economic groups (McNamara, Tsuchiya and Holmes, 2021), and in rare versus more common diseases (Bourke, Plumpton and Hughes, 2018).

2.2 Survey development

To inform the wording and presentation of the PTO tasks and the larger questionnaire design, we carried out a series of qualitative interviews. The key objectives of these interviews were to test and refine:

1. The explanation and presentation of absolute and proportional future health shortfall and the wording of the questionnaire, including how well interview participants understood shortfall expressed in terms of QALYs

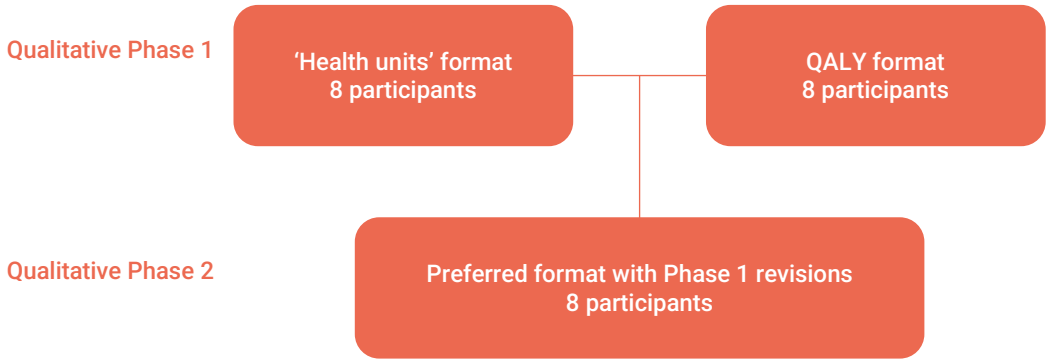
2. Implementation of the preference elicitation tasks which informed the health shortfall cut-offs and weights

The first of these objectives are discussed in section 2.2.1. The findings and key revisions made to the questionnaire relating to the second objective are discussed alongside the presentation of the final questionnaire in section 2.2.2.

2.2.1 QALY shortfall presentation and questionnaire wording

We initially piloted two versions of the survey. In one version we presented the concept of absolute and proportional future health shortfall described in terms of lost QALYs, consistent with standard health economic methods and terminology. In the other version, we described shortfall in terms of 'lost units of future health', to test whether this less-technical terminology might be more easily understood and considered by participants. Figure 2 outlines the stages of the qualitative phase with participant numbers.

FIGURE 2: STAGES OF THE QUALITATIVE PHASE



As NICE's modifier does not account for the cause of future health shortfall, we did not specify whether health was lost due to quality or length of life in tasks we presented. However, in the course of the interviews, we found that participants were more likely to think of survival losses when we described shortfall in terms of QALYs, whereas we found that they were able to think more abstractly when we described shortfall in terms of "health units". We found that participants often conflated 'QALYs' with (perfect) life years; that they often interpreted "adjusted" as specifically referring to poor health; and that they confused QALYs for a measure of quality of life.

Therefore, we proceeded to a second round of qualitative interviews with a new set of participants, presenting only 'health units', incorporating wording and presentation suggestions from both arms of the initial round of interviews. On the basis of this second set of interviews, detail was added to the information section to improve understanding of 'health units', including the graphical representations shown in section 2.2.2.1. We also changed the ordering of the tasks and added an age to the PS cutoff task for consistency with the AS task. These changes are described in more detail below.

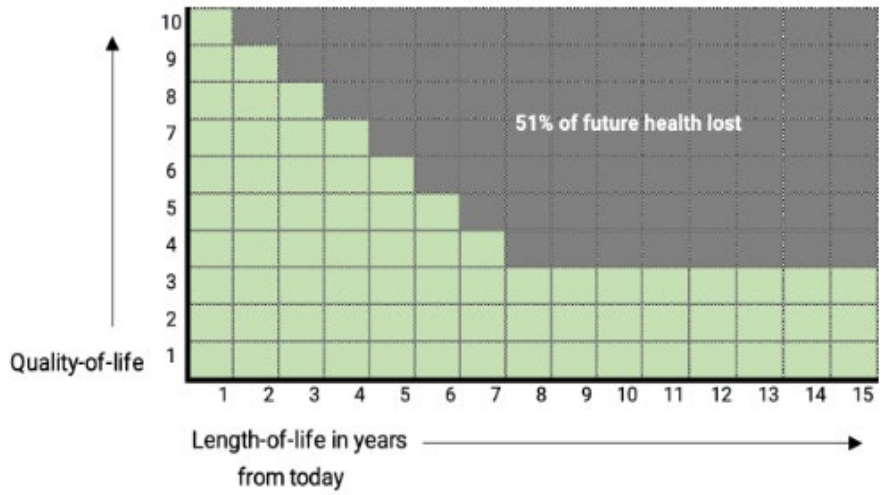
2.2.2 The final questionnaire

In the final questionnaire, we presented shortfall to survey respondents in terms of units of 'future health' lost, without mentioning the term "quality-adjusted life years" (QALYs). The main elements of the survey consisted of the introduction of the concept of health shortfall, the cut-offs elicitation, and the shortfall weights elicitation. This section presents the final questionnaire alongside any key revisions that were made as a result of the qualitative interviews.

2.2.2.1 Introduction of health shortfall

We illustrated "health shortfall" as a combination of reduced quality and reduced survival, as shown in Figure 3.

FIGURE 3: ILLUSTRATION OF 'LOST FUTURE HEALTH' IN THE INTRODUCTORY SECTION



In the final choice tasks, we did not specify the source of shortfall, consistent with NICE's 'agnostic' approach (Figure 4). However, in contrast to NICE's consideration of *discounted* shortfall, we presented *undiscounted* shortfall to participants. First, we considered that discounted health shortfall would be impractically difficult to explain to general population respondents. Second, any discounting of shortfalls would require us to define the source of the shortfall, as the same *undiscounted* shortfalls will have a different *discounted* shortfalls depending on whether future health was lost as a result of premature mortality or chronic disutility. This would have added considerable complication to the elicitation and ruled-out an 'agnostic' approach to the source of shortfall. Our use of undiscounted shortfall means that our estimates of *undiscounted* AS are not comparable to NICE's *discounted* AS. We return to the issue of discounting in the discussion.

FIGURE 4: ILLUSTRATION OF 'LOST FUTURE HEALTH' IN THE TASK INFORMING THE WEIGHTS

2.2.2.2 Reference ages in the tasks

As noted above, the qualitative interviews highlighted that respondents were making inferences about patient ages, even in the elicitation of PS cut-offs, where we had initially believed that respondents would be able to express a preference in the absence of any contextual information about patient age. Based on this insight, we added information on patient age to all tasks.

We used a reference age of 40 in the elicitation of AS and PS severity cut-off elicitations, chosen to be representative of the median age in England and Wales (40.5 per the Office of National Statistics (Office for National Statistics, 2024)). We also specified healthy life expectancy in the absence of disease, to contextualise absolute and proportional shortfalls. To align with a reference patient age of 40, we specified that patients had the potential for "350 health units", or an equivalent of 35 QALYs. Due to a mix-up in the questionnaire design, however, we specified a reference age of 45 and a potential for 300 health units (30 QALYs) in the PTO tasks. This means the reference age of patients in the PTO tasks was slightly older than the reference age in the cut-off elicitations, but all tasks implied a healthy life expectancy of 75 years.

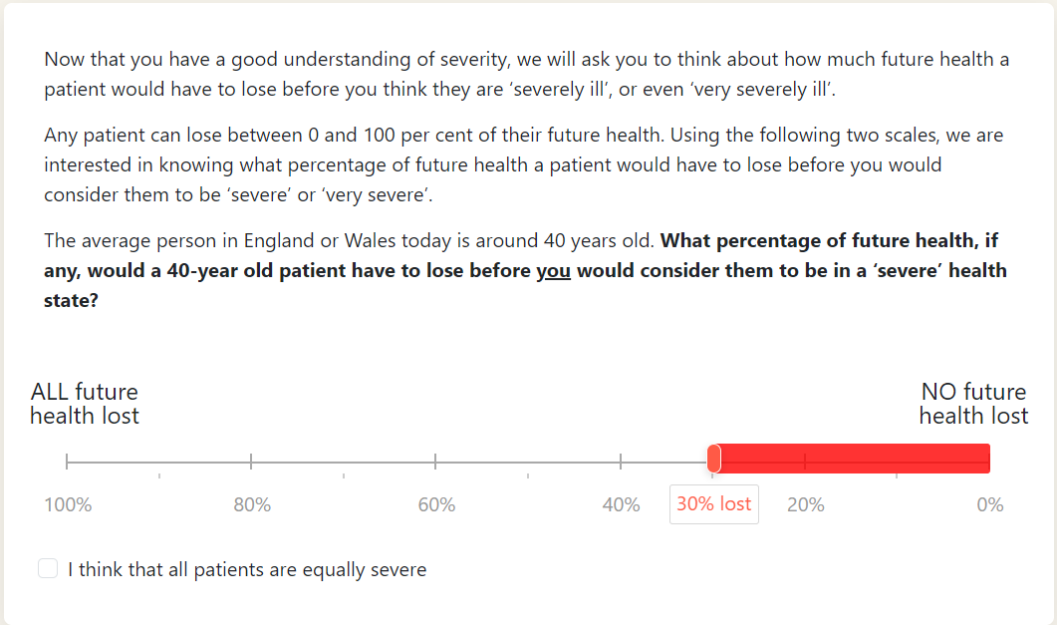
2.2.2.3 Eliciting severity cut-offs

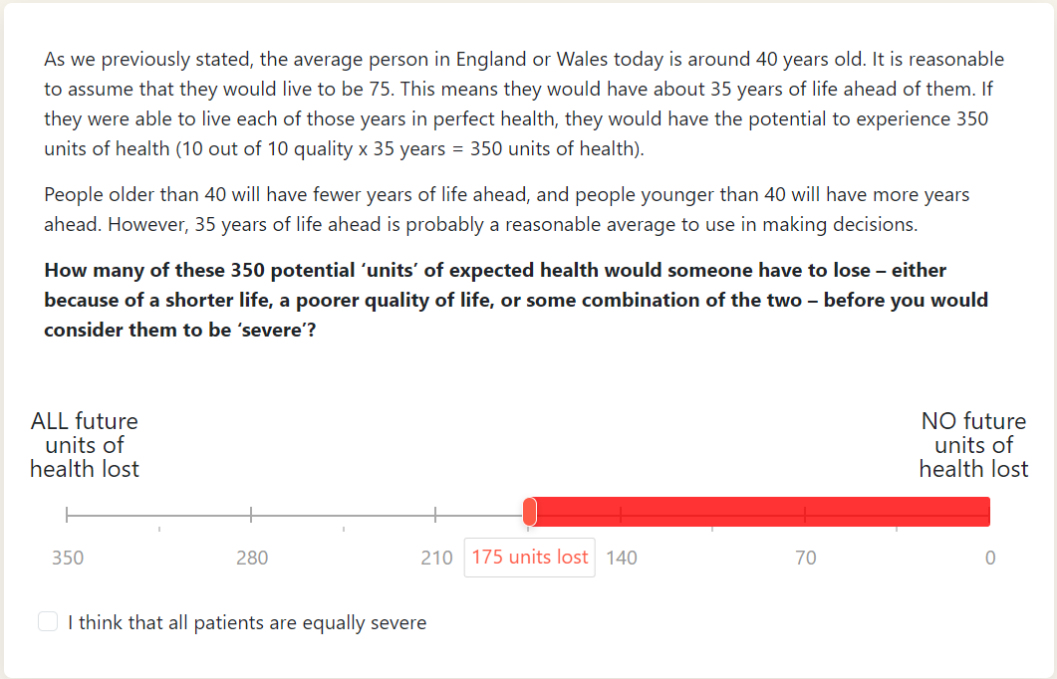
To estimate shortfall cut-offs for 'severe' and 'very severe' categories of AS and PS, respondents were presented with a series of interactive scales and asked to indicate up to four severity cut-offs.

We asked: "What percentage [PS] / How many units [AS] of future health, if any, would a 40-year-old patient have to lose before you would consider them to be in a 'severe' / 'very severe' health state?". See Figure 5 for an example of the presentation of the cut-off task.

For PS, they were asked to indicate a cut-off for 'severe' (if any) between 0 and 100% PS and, conditional on indicating a severe cut-off, a cut-off for 'very severe' (if any). The 'very severe' cut-off was constrained to be greater than the 'severe' cut-off. Respondents could indicate that all severity was equally relevant, or that there was no distinction between 'severe' and 'very severe'. The process was repeated for AS, with respondents asked to indicate absolute health shortfalls between 0 and 350 'health units' (equivalent to 35 QALYs, or a healthy life expectancy of 75 years).

FIGURE 5A & 5B: PROPORTIONAL AND ABSOLUTE SEVERITY CUT-OFFS TASKS





2.2.2.4 Eliciting shortfall weights

To elicit relative shortfall weights in our study, we administered a series of Person Trade Off (PTO) tasks. A PTO assesses the relative value of different individuals or groups by asking respondents to make prioritisation choices between them.

Respondents were presented with two groups with 100 patients initially in each: one less severe group and the other more severe. The PTO asked: "You, as the health system decision-maker have enough resources to treat one of these two groups. Which group would you treat?"

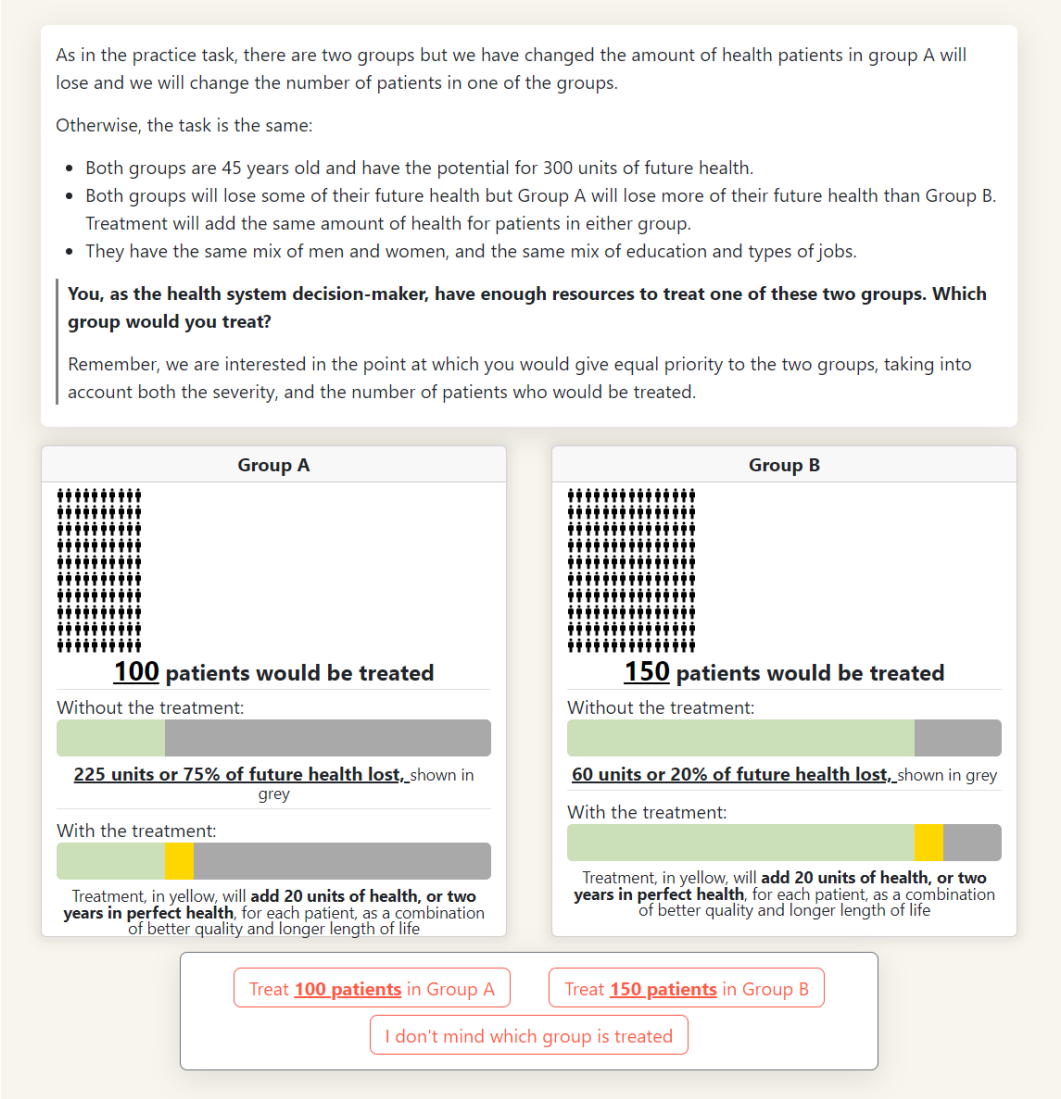
We assumed a baseline 'low severity' group of 20% PS. When we presented the PTO, we provided information on AS and the corresponding PS for a reference age of 45. Due to an oversight in the questionnaire design, this reference age is slightly different from the reference age of 40 in the severity threshold tasks. To contextualise AS and PS, we told respondents that in the absence of disease, each group had the potential for "300 units of future health" (equivalent to 30 future QALYs, or a healthy life expectancy of roughly 75 years).

After each choice, the number of patients that would be treated in the unselected group increased and they were asked to choose again. The number of patients in each severity group, at the point where they could not choose or did not mind which group to treat, informed the weight for that severity level (in comparison with the less severe group).

For example, in the PTO task in Figure 6 on the next page, if the respondent were to select 'I don't mind which group is treated', this would generate a weight of 1.5 for the 75% PS severity level, as there would have to be 1.5x as many patients treated in the less severe group for the respondent to give equal priority.

A maximum value of the number of patients in the non-selected group was limited to 1,000 patients. This would provide a relative weight of 10 if the more severe group was always chosen, and a relative weight of 0.1 if the less severity group was always chosen.

FIGURE 6: EXAMPLE PTO TASK



Appendix 2 outlines the iterative process of the PTO tasks. The algorithm shows the increased number of patients that were presented to respondents in the non-prioritised group, in each round of the PTO. If respondents did not select 'I don't mind which group is treated' for any round within a task for a given severity level, their selection of the patient groups over the course of the rounds provided a final range of patient numbers in the initially non-selected group in which their point of indifference would be found. At the end of each task, respondents who had not selected 'I don't mind which group is treated' were asked to select the number of patients within this range at which they would give equal priority to the two groups.

Each respondent completed six different PTO tasks, for different severity levels (30%, 50%, 65%, 75%, 85% and 95% PS). These tasks provided data for six relative weights for each respondent, compared to the fixed low severity group (20% PS).

We asked the PTO tasks in decreasing order of severity, starting with 95% PS. This ordering of the tasks was based on feedback from qualitative interviews. In the qualitative phase we trialled

presenting the tasks in a random order, but we found that participants were still anchoring their responses on previous severity levels even when presented randomly. Therefore, they were sometimes basing answers on unreliable recall of previous preferences, or not finding the changes in severity meaningful. Furthermore, presenting participants with the largest difference in severity helped respondents to understand the first task.

2.3 Data

We set an overall target of 1,000 individuals, with quotas based on the age and gender distribution of the population of England and Wales based on 2021 Census data (Nomis - Official Census and Labour Market Statistics, 2021). Respondents were recruited from a panel maintained by a survey firm (Prolific). Our analysis was conducted on a 'complete case' sample of 990 respondents who completed all of the mandatory severity cut-off and PTO tasks in the survey.

2.3.1 Outcome variables

The main outcome variables were the severity cut-off levels and the weights.

Cut-offs

Each respondent was asked to indicate up to four severity cut-offs. For PS, they were asked to indicate a cut-off for 'severe' (if any) between 0 and 100% PS and, conditional on indicating a severe cut-off, a cut-off for 'very severe' (if any). The 'very severe' cut-off was constrained to be greater than the 'severe' cut-off. Respondents could indicate that all severity was equally relevant, or that there was no distinction between 'severe' and 'very severe'. The process was repeated for AS, with respondents asked to indicate absolute health shortfalls between 0 and 350 'health units'. Respondents were again allowed to indicate no distinctions. Responses to the AS task were divided by 10 and are presented as QALYs in our results.

Weights

For each respondent, we calculated the relative weight of health gains at six levels of severity: 30, 50, 65, 75, 85 and 95 per cent proportional health shortfalls. Based on a reference age of 45 and a potential 300 future health units (equivalent of 30 QALYs), these proportional shortfalls correspond to absolute 'health unit' shortfalls of 90, 150, 195, 225, 255 and 285 health units (9, 15, 19.5, 22.5 and 28.5 QALYs). Each severity level was compared to a reference groups with a 20% PS (60 unit/6 QALY AS).

2.3.2 Demographic data

In demographic survey questions, we collected categorical data on age (18-29; 30-44; 45-59; 60-74; and 75+), sex (male and female), education categories (less than secondary education; secondary education; further education; higher education; prefer not to answer) and ethnicity (White; Asian; Black; Mixed; Other; and prefer not to answer).

Additionally, we collected data on the health and wellbeing status of the respondent. This included an own health quality rating between 0 (worst health quality) and 100 (best health quality). We collected data on whether the respondent self-reported as having a long-term health condition (binary yes or no). Of those who answered yes, they reported whether this condition is severe (binary yes or no). Respondents completed a series of questions relating to their overall wellbeing, providing answers on a 1-10 Likert scale: "Overall how happy did you feel yesterday?"; "Overall how anxious did you feel yesterday?"; "Overall how satisfied are you with your life nowadays?"; "Overall to what extent do you feel that the things you do in your life are worthwhile?".

Finally, each respondent answered a 3-item version of the Subjective Numeracy Scale (SNS-3) (McNaughton et al., 2015). They were asked a series of three questions which were answered on a 1-

6 Likert scale: How good are you at working with fractions?; How good are you at figuring out how much a shirt will cost if it is 25% off?; How often do you find numerical information to be useful? The total score across the three questions provided their total numeracy score between 3 (lowest numeracy level) and 18 (highest numeracy level).

2.3.3 Value statements

In addition to the elicited preference tasks, we collected data on a series of attitudinal questions regarding the prioritisation of health care services. Respondents were asked how far they agreed or disagreed (strongly agree; agree, neither agree nor disagree; disagree; strongly disagree) with a series of statements. The five value statements were: "The NHS should spend less money on patients with more severe health conditions because there is probably less the NHS can do to help them"; "The NHS should be willing to spend more on the care of patients with more severe health conditions compared to patients with less severe conditions"; "A patient who will die sooner is always more severe than someone who will live longer no matter what their quality of life is"; "Spending more on a patient with a more severe condition means there will be less money or resources available for patients with less severe conditions"; and "The goal of any health system should be to produce as many units of health as possible regardless of who receives them".

2.4 Statistical analysis

2.4.1 Sample summary

We first describe the demographic data and answers to the value statements, to describe the characteristics of our final sample, and visually compare our sample age and sex distributions to the UK population distributions.

2.4.2 Cut-offs

We conducted a descriptive analysis of the threshold values of the survey responses for severe and very severe categories. We describe the mean and median thresholds.

2.4.3 Shortfall (severity) weights

For each PTO task, we describe the mean and median of the ratios of the number of patients in the more severe group relative to the less severe group. We took a mean (median) of ratios approach, where the shortfall severity ratios (i.e. severity weights) are calculated for each individual respondent based on the person equivalent responses to the PTO. We then calculate the mean (median) of these ratios, rather than calculating the mean (median) person-equivalents across all responses for a particular shortfall and then calculating the ratio of those means/medians (ratio of mean/median).

2.4.3.1 Sensitivity checks

It is recognised that PTO can be associated with a degree of 'protest responses', where respondents provide extreme answers to emphasise their normative position on an issue, independent of the levels presented in a particular scenario (Green, 2001). To test the impact of any such protest responses, we conducted two of sensitivity analyses excluding selected categories of responses. First, we removed individuals who *selected the maximum on all tasks* in the PTO. Secondly, we removed individuals who *refused to trade on any tasks* in the PTO (immediately selected indifferent on all tasks).

We also tested the impact of removing respondents who may not have engaged with or fully-understood the tasks. Firstly, we removed individuals who failed either of the comprehension checks. Secondly, we removed individuals with a low numeracy score (lowest 5% of individuals). Finally, we

tested whether the results were sensitive to removing individuals with a low completion time, suggesting inattentive responses or speeders (lowest 5% of individuals).

3 Results

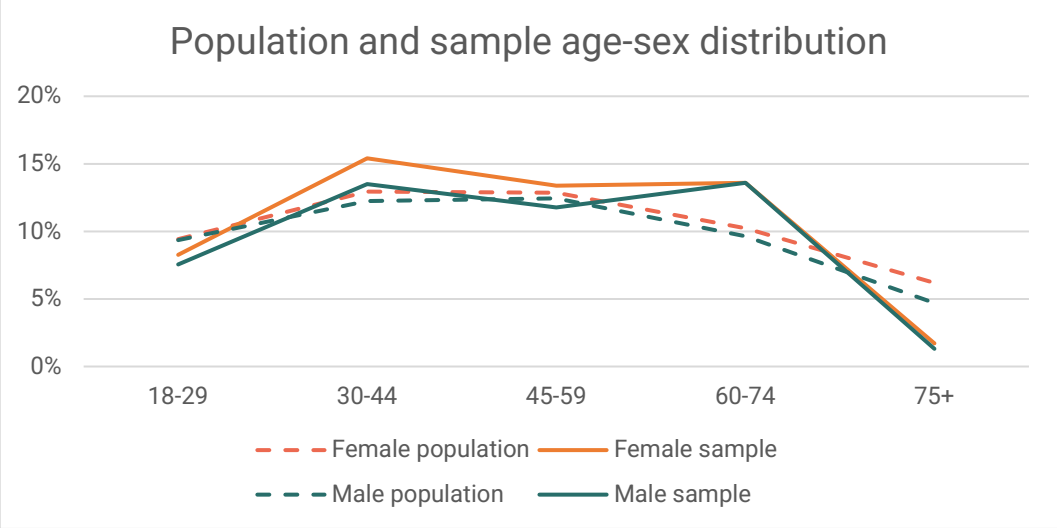
3.1 Sample summary

We received a total 997 responses. Of these, 7 were incomplete on at least one element of the cut-off or weights questions, leaving us with a sample of 990 respondents for a complete case analysis across these two key variables. There was slight variation in sample numbers on demographic questions due to non-response.

3.1.1 Demographic data

The final sample was broadly representative of the age and sex distribution of the population of England and Wales, based on Census 2021 data (Figure 7). There was slight divergence between our sample and the population age-sex distributions due to under-recruitment in some age-sex groups, primarily among the over-75 groups.

FIGURE 7: SAMPLE AND POPULATION DISTRIBUTION BY AGE AND SEX



Source for population figures : Nomis - Official Census and Labour Market Statistics (2021)

Table 1 summarises the demographic distribution of our sample. 48% of the respondents were male while 52% were female. The proportion of the sample in each category was: 16% 18-29; 29% 30-44; 25% 45-59; 27% 60-74; and 3% 75+. Our sample contains a slightly higher proportion of white respondents than population figures based on 2021 census figures, our target was 82% White, 9% Asian, 3 % Black, 2% Mixed (Office for National Statistics, 2022b) whilst our sample was 86% White, 8% Asian, 4% Black, 1% Mixed and 1% Other. Our sample also contained a much higher proportion of respondents with a higher education, and those with less than secondary education are not represented, compared to the general population. Our sample included 0% with a less than secondary education, 17% with secondary education, 22% with further education and 61% with a higher education while census figures show 34% higher educated and 18.2% with no qualifications (Office for National Statistics, 2022a).

TABLE 1: DESCRIPTIVE STATISTICS OF DEMOGRAPHIC DATA

Variable	Mean / %	Std dev	Min	Max
% Male (n=990)	47.4%			
Education (n=987)				
Less than secondary education	0.2%			
Secondary education	16.5%			
Further education	21.9%			
Higher education	61.2%			
Prefer not to answer	0.2%			
Age categories (n=990)				
18-29	15.9%			
30-44	28.8%			
45-59	25.3%			
60-74	27.0%			
75+	3.0%			
Ethnicity (n=987)				
White	86.0%			
Asian	7.7%			
Black	3.9%			
Mixed	1.3%			
Other	0.6%			
Prefer not to answer	0.5%			
Own health slider value (VAS) (n=990)	76.5	18.7	0	100
Wellbeing statements (n=990)				
1. Overall how happy did you feel yesterday?	6.7	2.4	0	10
2. Overall how satisfied are you with your life nowadays?	6.6	2.3	0	10
3. Overall to what extent do you feel that the things you do in your life are worthwhile?	7.0	2.4	0	10
4. Overall how anxious did you feel yesterday?	3.5	2.9	0	10
Long term health condition (n=987)				
Yes	36%			
No	61%			
Prefer not to answer	3%			
Severe long term health condition (n=987)				
Yes	7%			
No	89%			
Prefer not to answer	4%			
Subjective Numeracy Score (SNS3) (n=987)	13.8	3.2	3	18

Table notes: For wellbeing statements 1-3, a higher score indicates better wellbeing. For wellbeing statement 4 a lower score indicates better wellbeing.

31% of the sample had a self-assessed long term health condition, and 7% had a long-term condition which they classified as severe. The average score for health quality was 77 out of a maximum of 100. The average score for 'how happy did you feel yesterday?', 'how satisfied are you with your life nowadays?' and 'to what extent do you feel that the things you do in your life are worthwhile?' were 6.7, 6.6 and 7.0, respectively. For these three statement a higher score indicates better wellbeing. For the wellbeing statement 'how anxious did you feel yesterday?' the average score was 3.5, where a lower score indicates better wellbeing. The average SNS-3 score was 13.8 on the scale from 3 to 18.

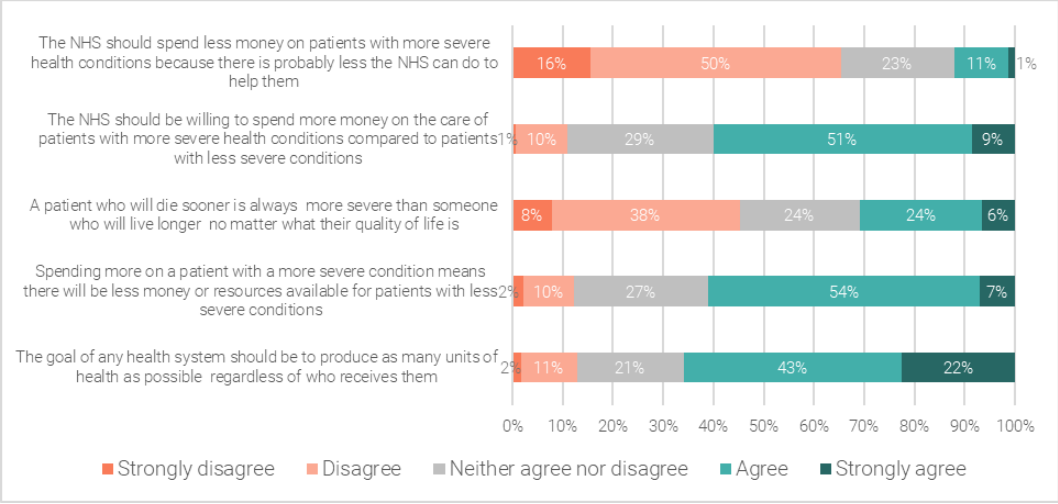
3.1.2 Value statements

Figure 8 outlines the responses to the attitudinal statements over health care prioritisation. Responses to these value statements showed little support for spending less on more severe patients, and support for spending more on more severe patients. Sixty-six per cent disagreed or strongly disagreed with spending less on more severe patients, whilst 60% agreed or strongly agreed that the NHS should be willing to spend more money on the care of patients with more severe health conditions, compared to patients with less severe conditions. This implies that at least 60% of the sample sees the relative value of health gains to more severe patients is greater than 1.0.

Respondents recognised the opportunity costs of prioritising severity. Sixty-one percent of the sample agreed or strongly agreed that spending more on a patient with a more severe condition means that there will be less money or resources available for patients with less severe conditions. (We recognise, however, that this statement implicitly imposes a fixed budget, and that disagreement may reflect a preference for easing the budget constraint rather than a rejection of the notion of opportunity cost.)

There were mixed attitudes towards the extent to which quality of life and length of life contribute to severity. In response to the statement that a patient who will die sooner is always more severely ill than someone who will live longer, no matter what their quality of life is: 32% agreed or strongly agreed, 46% disagreed or strongly disagreed, and 24% neither agreed nor disagreed.

FIGURE 8: RESPONSES TO ATTITUDINAL STATEMENTS ON HEALTHCARE PRIORITISATION



Finally, 65% of respondents agreed or strongly agreed that the goal of any health system should be to produce as much health as possible, potentially contradicting support for spending more on patients in more severe health states. Together, these suggest some support for prioritising more severe patients within a broader health maximising framework. This implies support for – but a limit

to – an equity-efficiency trade-off over severity, reinforcing the importance of identifying the limit of this trade-off.

3.2 Severity Cut-offs

Figure 9 provides histograms of respondent answers to where 'severe' and 'very severe' health shortfalls begin on proportional and absolute shortfall scales. Detailed descriptive statistics are presented in Table 2. The results suggest that the public's view of severity is substantially broader than currently assumed by NICE.

On the proportional shortfall (PS) scale, the mean cutoff for 'severe' was 48% PS and the mean cutoff for 'very severe' was 64%. Both are substantially lower than the 85% and 95% cutoffs, respectively, assumed by NICE.

On the absolute shortfall (AS) scale, the mean cutoff for 'severe' was 16 undiscounted QALYs lost and the mean cutoff for 'very severe' was 21 undiscounted QALYs lost. As NICE is using discounted QALYs, *these are not directly comparable* to NICE's AS thresholds of 12 and 18 QALYs lost.

FIGURE 9: HISTOGRAMS OF RESPONDENT ANSWERS TO 'SEVERE' AND 'VERY SEVERE' SHORTFALL CUT-OFFS

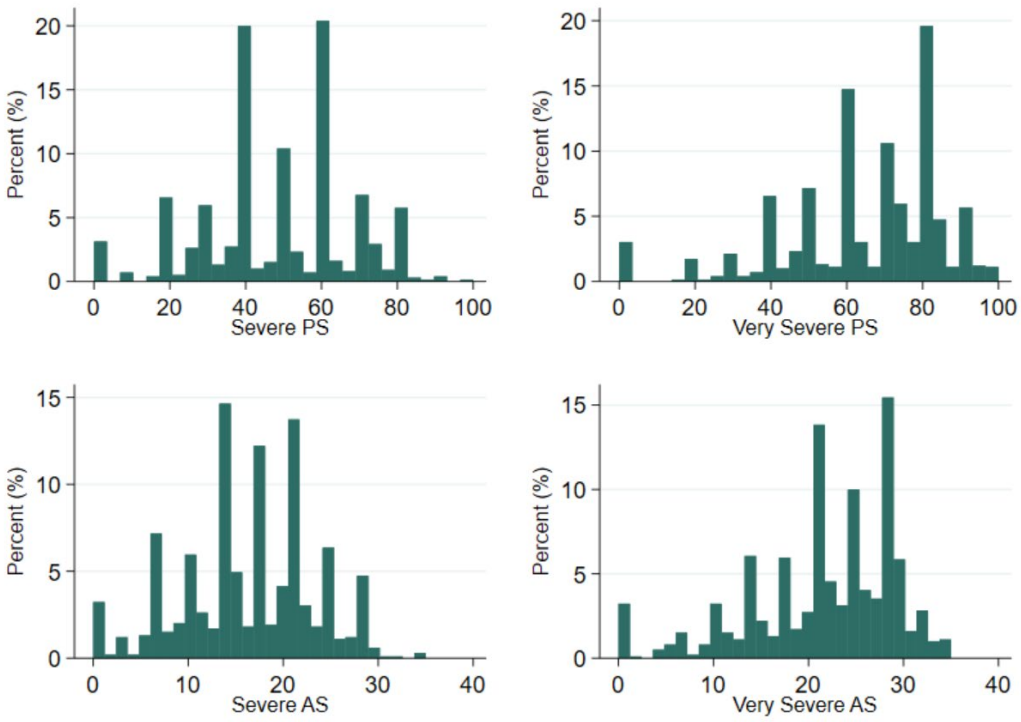


TABLE 2: DESCRIPTIVE STATICS OF RESPONDENT ANSWERS TO 'SEVERE' AND'VERY SEVERE' SHORTFALL CUT-OFFS

	Mean	Median	Standard deviation	Obs
PS (%, 0-100)				
Severe	48.3	50	19.3	990
Very severe	64.5	70	20.7	990
AS (QALYs, 0-35)				
Severe	16.3	17.4	6.9	990
Very severe	21.6	22.5	7.6	990

3.3 Shortfall Weights

We calculated mean and median weights for each shortfall severity level, relative to the low severity group. The weights data for each severity level are right-skewed, and subject to a large number of outliers. As the severity level increases, the data appear increasingly skewed, with more outliers and more respondents selecting the maximum amount on the PTO (Figure 10).

FIGURE 10: DISTRIBUTION OF WEIGHTS BY PROPORTIONAL SHORTFALL SEVERITY LEVEL, CALCULATED IN RELATION TO THE BASELINE (20% PS PATIENT GROUP)

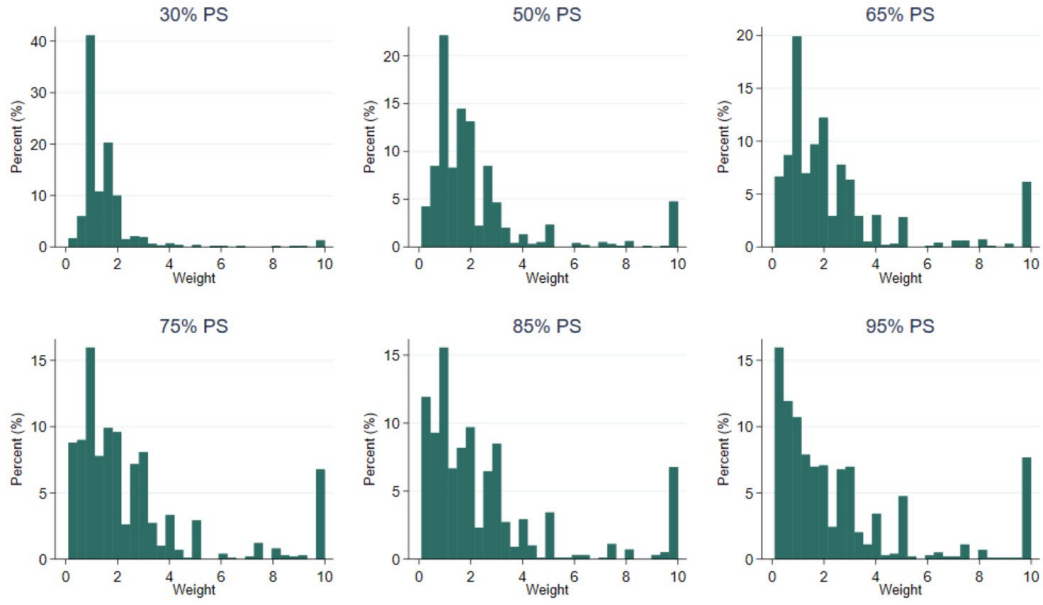


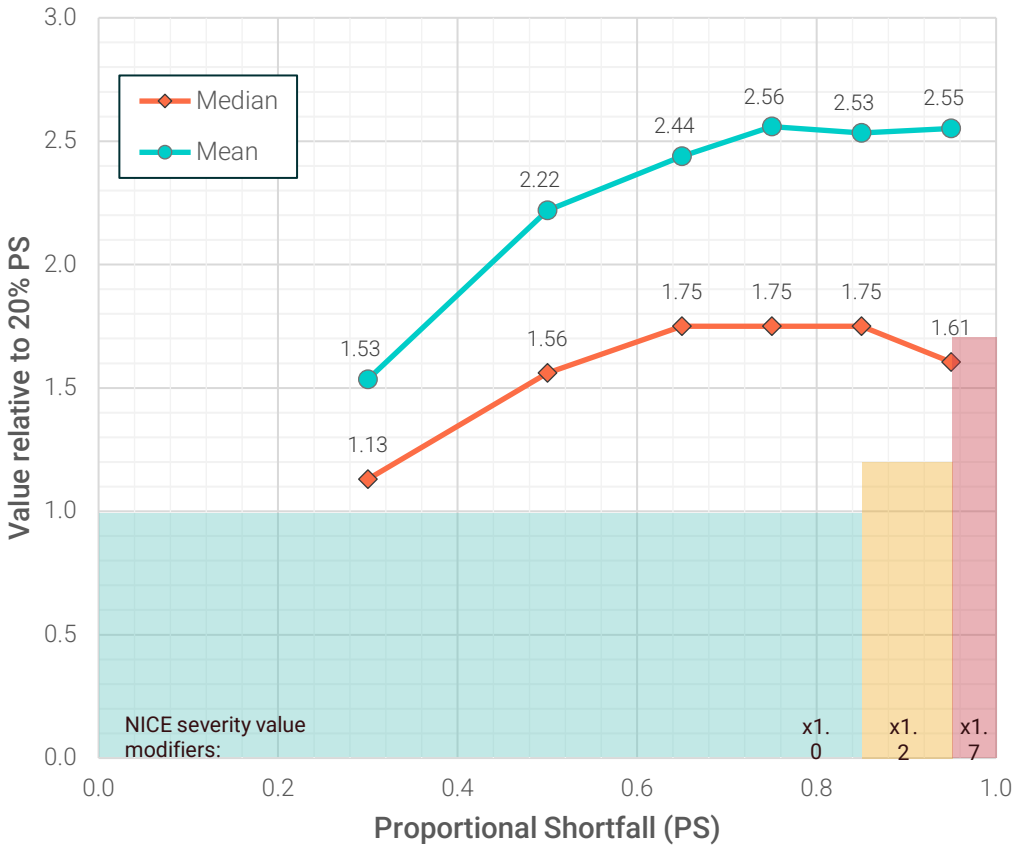
Figure 11 presents the mean and median values for relative value across different proportional shortfall severity levels, compared to NICE's current severity categories and value multipliers. We find that the relative value of health gains generally increased over severity but plateaued at higher severity levels. Appendix 2 presents detailed descriptive statistics of the mean and median values. The elicited severity weights were higher than NICE's severity modifier across the range of severity with the notable exception of the median value for the highest PS tested (95% PS). The median value

at 95% PS declined from the previous value, suggesting the possibility of a *threshold effect* in societal preferences over severity.

AS results can be found in Appendix 3. As our PTO elicited preferences for a single reference age group (45-year-olds), the absolute shortfall (AS) levels always correspond to a specific proportional shortfall (PS). Therefore, the pattern of weights over AS and PS are identical.

In practice, NICE has 'de-linked' the AS and PS thresholds to avoid disadvantaging older or younger groups with a single definition of severity. In themselves, PS is seen as more likely to favour older groups as it will tend to be higher for older individuals who have relatively fewer expected life years remaining, whilst AS is seen as more likely to favour younger groups they have more life years to potentially lose.

FIGURE 11: MEAN AND MEDIAN VALUES OF RELATIVE SEVERITY VALUE BY PROPORTIONAL SHORTFALL; NICE SEVERITY VALUE MODIFIERS SHOWN FOR REFERENCE



3.4 Sensitivity checks

Appendix 5 shows that the estimated cut-offs were not impacted by the removal of potential protest responses in the PTO. Likewise, our severity weight estimates do not appear to be impacted by protests responses. The scale of the PTO results did not substantially change with the removal of the small number of respondents who selected the maximum value of patients on all tasks, nor removing respondents who immediately selected indifference on all tasks.

Appendix 6 shows that the estimated cut-offs were not impacted by the removal of respondents with potentially low engagement. However, results indicate that respondents with low engagement or understanding provided lower PTO weights. Removing respondents who failed either of the checks, or had a low completion time, increased the median estimate of severity. Removing respondents with a low numeracy score did not appear to impact the results.

4 Discussion

The results of this preference elicitation show that the English and Welsh public is willing to forego some health gains to patients with smaller health shortfalls to prioritise gains to patients with greater health shortfalls, implying an equity-efficiency trade-off over severity, even in the mildest severity comparison (30% vs 20% proportional shortfall). As noted, we presented undiscounted shortfall to survey respondents and as such, our undiscounted AS and PS cutoffs and weights are not directly comparable to NICE's discounted AS cutoffs and weights. The impact of discounting on PS, however, is much less than on AS, and as such, we focus on our PS results and discuss the appropriateness of discounting below.

Not surprisingly, we see some divergence in the mean and the median severity weights. In the PTO tasks, we allowed responses of up to 1000 persons treated in the less preferred group, implying that the preferred group in any task could have a relative value at least 10 times greater than the anchor of 100 persons in the reference group. Given this relatively high 'top end' relative value, we were not surprised to see a long tail on the distribution of person-equivalent values and a corresponding divergence between the mean and median PTO weights. However, as this long tail means that the mean estimates are driven by some extreme responses, we believe that the median estimates are more representative of UK public preferences. Medians were also used in similar research by Reckers-Droog, van Exel and Brouwer (2019). Therefore, the remainder of our discussion refers to the median weights.

Public concern for severity appears to increase with shortfall at the milder end of the severity scale but levels out beyond 65% PS and even declines at the most severe shortfall (95% PS). Notably, the inflection point at 65% PS closely corresponds with the threshold respondents indicated as distinguishing between 'severe' and 'very severe' shortfall in a separate task in the questionnaire. To us, these corresponding findings, elicited in different questions and using different response formats, reinforce the internal consistency of the results. Together, they suggest that while respondents saw different value (or priority) in health gains to different patient groups at the lower range of shortfall, they did not discriminate between health gains to (most) patients at the more extreme end of the shortfall range, once patients had passed some critical minimum level of severity.

The one exception to the idea that respondents were not willing to discriminate between shortfalls at the upper end of the severity range was the decline in the median value weights between 90% and 95% PS (although we note that the same decline was not seen in the mean weights, which in fact increased between 90% and 95% PS). One possible explanation for this decline in the median weight is concern for the quality or functionality of the final health state, *in addition* to concern for untreated severity.

As described by Dolan and Cookson (2000), the public may prefer to give priority to patients in more severe health states out of a concern for the worse-off, but at some point they become less willing to allocate resources to patients that are likely to *remain* in a relatively poor health state following treatment. They suggest that treatment must result in some minimum, or threshold, level of post-treatment quality to justify priority, regardless of the scale of shortfall. Similarly, in interviews of the UK public, Roberts et al. (1999) found that participants were reluctant to allocate resources to patients that would remain in a severe health state following treatment, even when such allocations maximised expected QALYs.

In the context of our study, we did not specify the source of the health shortfall (i.e. morbidity or mortality) nor how the 'health units' were distributed between quality and survival, but some

respondents may have believed that the relatively small health gains associated with treatment (20 'health units', or the equivalent of 2 QALYs) would be unlikely to meaningfully improve the situation of patients with the most extreme shortfall, and therefore assigned less value (or priority) to health gains to these patients. Indeed, to the extent that patients with the greatest shortfall may be close to end-of-life, empirical evidence suggests that a majority of the UK public would prefer a 'good death', including a sense of dignity and control and good relations with family and friends, to small survival gains (c.f. Skedgel et al., 2024). At the societal level, Reckers-Droog et al. (2021) found that willingness to pay for health gains declined as patients neared the end of their natural life expectancy.

Beyond the relative priority of the most severe group, the distribution of relative value across the shortfall range may have been influenced by our presentation of the same, relatively small, health gain with treatment across all PTO tasks. This approach means that groups with the greatest shortfall received – proportional to their shortfall – smaller gains than groups with less severe shortfalls. We kept the *absolute* health gain from treatment constant across groups, but an alternative would have been to keep the *proportional* gain constant. However, we felt that this presentation could lead to a similar but reverse effect by introducing a *magnitude effect* into people's choices, if they focused on the increasing absolute gains necessary to maintain a constant proportional gain. Asking respondents to consider changing absolute health gains would also likely increase the cognitive burden of the tasks. Ultimately, we decided that the risk of bias associated with a magnitude effect from presenting increasing *absolute* gains over shortfall was greater than respondents recognising declining *proportional* gains relative to shortfall. Future research, though, could test how presentation of absolute or proportional heath gains affects respondent choices.

As noted earlier, due to an oversight in the questionnaire development, the reference age was inconsistent between the severity cut-off tasks and the PTO tasks. However, the difference was relatively small (40 and 45 years old, respectively) and the healthy life expectancy implied by the potential future health units was the same between the two sets of tasks (75 healthy life years in both sets of tasks). We saw no evidence that this discrepancy meaningfully affected the results, and we note the close agreement between the plateauing of severity weights at 65% PS and the mean 'very severe' cut-off of 64% despite the small difference in the reference age.

We also note that our choice reference age was arbitrary, and a different reference age would likely lead to different cut-offs and severity weights (especially in the context of AS). In a similar elicitation of public preferences over age and severity, Reckers-Droog, van Exel and Brouwer (2019) found a clear negative relationship between elicited severity weights and the age of the patients. Such a relationship between societal weight and patient age will be a challenge to ever identifying a single set of weights based on societal preferences. Consistent with a principle of reflecting societal characteristics in our elicitation, we used the median age in England and Wales as our reference, but a different reference age would be equally valid, such as the average or median age of an NHS patient. Future research will be required to understand the impact of different reference ages and other contextual factors on societal preferences over severity.

As we noted in Section 2.2.3, based on findings from the think-aloud interviews, we presented the PTO tasks in order of declining shortfall in the more severe group, starting from 95% vs 20% PS shortfall and finishing with 30% vs 20% PS. We intentionally adopted this approach to constrain the person-equivalent responses based on our experience in the initial round of the think-aloud pilots. In our initial PTO design, presented in order of increasing severity, many respondents reached the upper limit of the person-equivalents at relatively moderate shortfalls. As such, the results presented here are likely to be conservative relative to the same, equally valid, design presented in ascending order of shortfall. Our descending order is also likely to have imposed some internal consistency on the individual PTO responses, as a respondent's answer to the previous PTO task will have provided a logical ceiling on the person-equivalent value in the next, less-severe comparison.

Finally, we sought to understand public preferences within the conceptual framework of the current NICE modifier. Some authors, though, including Hausman (2024) and Barra *et al.* (2020), have suggested that future health shortfall may not be the most appropriate measure of disease severity from a philosophical or distributional justice perspective. We take no position here regarding the most appropriate conception of severity, but we recognise that asking about different conceptions of severity, such as *lifetime* rather than *future* health shortfall, might produce different preferences. Furthermore, we did not provide participants any information on the *past* health of the hypothetical patient groups. That is, we did not specify whether patients had lived in perfect health up to the moment presented in the PTO or whether they had accrued some health shortfall in the past. However, as Hausman (2024) acknowledges, lifetime and future health shortfalls are likely to be correlated, and therefore we would expect societal preferences to move in a broadly similar direction regardless of whether or not past shortfall is incorporated.

Returning to the issue of discounting, we have pragmatic and theoretical objections to discounted shortfall in the context of the severity modifier. First, discounting, by its nature, is not an intuitive concept. Ensuring that public respondents fully grasp the concept of discounting would most likely require an interviewer, making large-scale online preference surveys of public preferences in this area expensive and time-consuming, if not impossible. As such, adopting discounted shortfall as a criteria in the severity modifier will likely severely constrain the use of public preferences to inform the characteristics of the modifier. Second, and as we noted in the methods, the source of health shortfall (mortality or morbidity) will result in different *discounted* shortfalls for the same *undiscounted* shortfall. This means that, depending on the source of shortfall, one condition could qualify for the modifier whilst another with the same undiscounted shortfall would not. This discrimination on the basis of the *source* of health losses seems at odds with NICE's otherwise careful indifference to the source of QALY gains (or, in this case, losses). Third, and perhaps most importantly, we see no theoretical need or justification for discounting in the context of severity. Indeed, *the very fact that some health shortfalls extend far into the future is what makes them most relevant to priority-setting.* We struggle to see the logic in a position that holds individuals with greater future health shortfalls deserve greater priority while simultaneously giving increasingly less weight to the shortfalls that extend furthest into future. We understand that the use of discounted shortfalls may have been a pragmatic solution to identifying the initial parameters for the severity modifier based on an analysis of historical submissions reporting discounted QALYs, but we strongly suggest that NICE should reconsider their ongoing role.

Our results suggest that NICE's current severity modifier is not well aligned with the UK public's preference for prioritising health gains in more severe health states. We find that societal concern begins at a substantially lower shortfall threshold, and that the public assigns greater relative value to health gains at almost every level of severity, than NICE's current severity modifier. If NICE seeks to align the value and priority assigned to new medicines and technologies with societal preferences, these results suggest a need for NICE to reassess its criteria for the severity modifier.

References

Al-Janabi, H., Wittenberg, E., Donaldson, C. and Brouwer, W., 2022. The relative value of carer and patient quality of life: A person trade-off (PTO) study. *Social Science & Medicine*, 292, p.114556. 10.1016/j.socscimed.2021.114556.

Barra, M., Broqvist, M., Gustavsson, E., Henriksson, M., Juth, N., Sandman, L. and Solberg, C.T., 2020. Severity as a Priority Setting Criterion: Setting a Challenging Research Agenda. *Health Care Analysis*, 28(1), pp.25–44. 10.1007/s10728-019-00371-z.

Bourke, S.M., Plumpton, C.O. and Hughes, D.A., 2018. Societal Preferences for Funding Orphan Drugs in the United Kingdom: An Application of Person Trade-Off and Discrete Choice Experiment Methods. *Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research*, 21(5), pp.538–546. 10.1016/j.jval.2017.12.026.

Brouwer, W.B.F., Culyer, A.J., van Exel, N.J.A. and Rutten, F.F.H., 2008. Welfarism vs. extra-welfarism. *Journal of Health Economics*, 27(2), pp.325–338. 10.1016/j.jhealeco.2007.07.003.

Culyer, A.J., 1991. The normative economics of health care finance and provision'. In: A. McGuire, P. Fenn and K. Mayhew, eds., *Providing Health Care: the Economics of Alternative Systems of Finance and Delivery*. Oxford: Oxford University Press, pp.65–98.

Dolan, P. and Cookson, R., 2000. A qualitative study of the extent to which health gain matters when choosing between groups of patients. *Health Policy (Amsterdam, Netherlands)*, 51(1), pp.19–30. 10.1016/s0168-8510(99)00079-2.

Gu, Y., Lancsar, E., Ghijben, P., Butler, J.R. and Donaldson, C., 2015. Attributes and weights in health care priority setting: A systematic review of what counts and to what extent. *Social Science & Medicine*, 146, pp.41–52. 10.1016/j.socscimed.2015.10.005.

Hausman, D.M., 2024. Problems with NICE's severity weights. *Social Science & Medicine*, 348, p.116833. 10.1016/j.socscimed.2024.116833.

McNamara, S., Tsuchiya, A. and Holmes, J., 2021. Does the UK-public's aversion to inequalities in health differ by group-labelling and health-gain type? A choice-experiment. *Social Science & Medicine (1982)*, 269, p.113573. 10.1016/j.socscimed.2020.113573.

McNaughton, C.D., Cavanaugh, K.L., Kripalani, S., Rothman, R.L. and Wallston, K.A., 2015. Validation of a Short, 3-Item Version of the Subjective Numeracy Scale. *Medical Decision Making*, 35(8), pp.932–936. 10.1177/0272989X15581800.

Mooney, G., 1998. 'Communitarian claims' as an ethical basis for allocating health care resources. *Social Science & Medicine (1982)*, 47(9), pp.1171–1180. 10.1016/s0277-9536(98)00189-0.

NICE, 2022. NICE health technology evaluations: the manual.

Nomis - Official Census and Labour Market Statistics, 2021. *Sex by single year of age*. [online] Available at: https://www.nomisweb.co.uk/datasets/c2021ts009 [Accessed 16 Sep. 2024].

Nord, E. and 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), pp.281–288. 10.1016/j.healthpol.2014.02.009.

Norwegian Ministry of Health and Care Services, 2017. *Principles for priority setting in health care: Summary of a white paper on priority setting in the Norwegian health care sector*. [online] Available at: https://www.regjeringen.no/contentassets/439a420e01914a18b21f351143ccc6af/en-gb/pdfs/stm201520160034000engpdfs.pdf [Accessed 23 Jul. 2023].

Office for National Statistics, 2022a. *Education, England and Wales*. [online] Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/educationandchildcare/bulletins/educationenglandandwales/census2021 [Accessed 27 Aug. 2024].

Office for National Statistics, 2022b. *Ethnic group, England and Wales*. [online] Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/bulletins/ethnicgroupenglandandwales/census2021 [Accessed 27 Aug. 2024].

Office for National Statistics, 2024. *Population estimates for the UK, England, Wales, Scotland, and Northern Ireland: mid-2023*. [online] Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2023 [Accessed 24 Oct. 2024].

Peasgood, T., Bailey, C., Chen, G., Silva, A.D., Perera, U.D.S., Norman, R., Shah, K., Viney, R. and Devlin, N., 2024. Rationale, conceptual issues, and resultant protocol for a mixed methods Person Trade Off (PTO) and qualitative study to estimate and understand the relative value of gains in health for children and young people compared to adults. *PLOS ONE*, 19(6), p.e0302886. 10.1371/journal.pone.0302886.

Reckers-Droog, V., Exel, J. van and Brouwer, W., 2021. Willingness to Pay for Health-Related Quality of Life Gains in Relation to Disease Severity and the Age of Patients. *Value in Health*, 24(8), pp.1182–1192. 10.1016/j.jval.2021.01.012.

Reckers-Droog, V., Van Exel, J. and Brouwer, W., 2019. Equity Weights for Priority Setting in Healthcare: Severity, Age, or Both? *Value in Health*, 22(12), pp.1441–1449. 10.1016/j.jval.2019.07.012.

Richardson, J., Iezzi, A. and Maxwell, A., 2017. How important is severity for the evaluation of health services: new evidence using the relative social willingness to pay instrument. *The European journal of health economics: HEPAC: health economics in prevention and care*, 18(6), pp.671–683. 10.1007/s10198-016-0817-y.

Roberts, T., Bryan, S., Heginbotham, C. and McCallum, A., 1999. Public involvement in health care priority setting: an economic perspective. *Health Expectations*, 2(4), pp.235–244.

Shah, K.K., 2009. Severity of illness and priority setting in healthcare: a review of the literature. *Health Policy*, 93(2–3), pp.77–84. 10.1016/j.healthpol.2009.08.005.

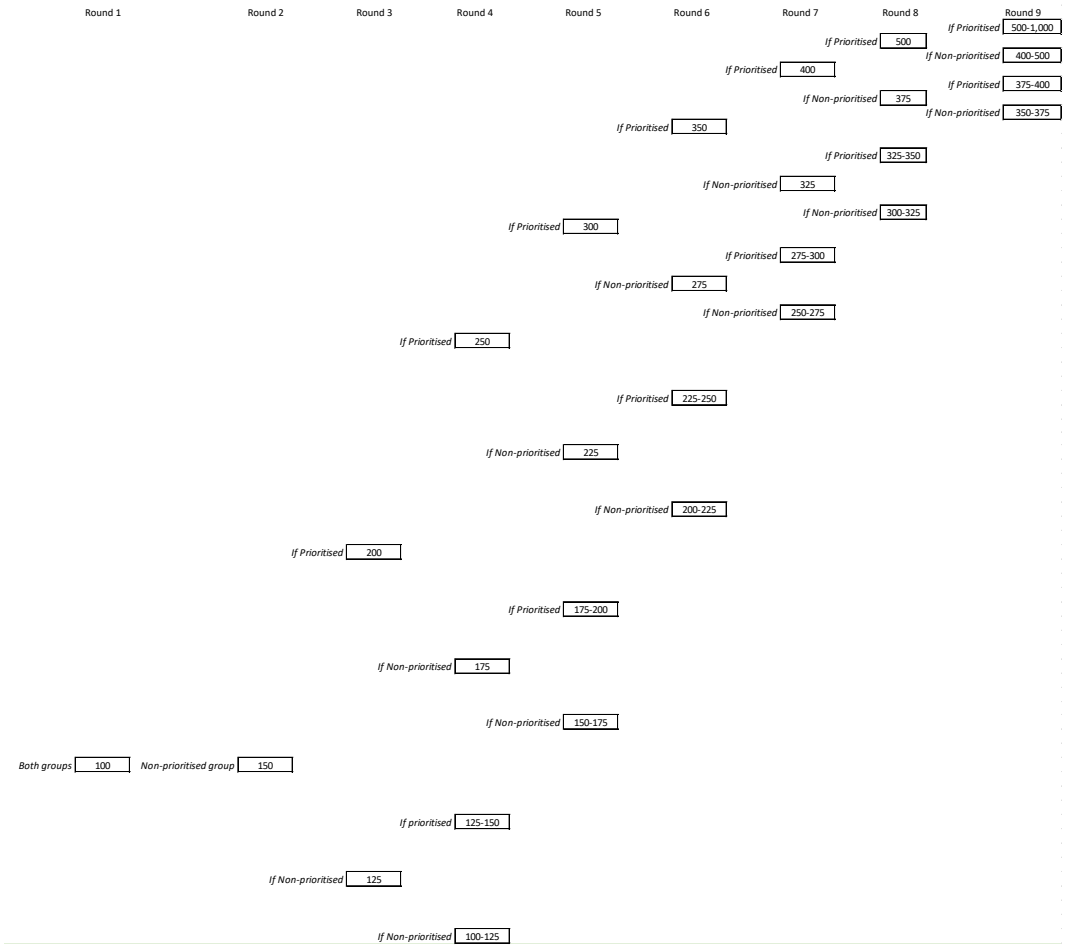
Skedgel, C., Henderson, N., Towse, A., Mott, D. and Green, C., 2022. Considering Severity in Health Technology Assessment: Can We Do Better? *Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research*, 25(8), pp.1399–1403. 10.1016/j.jval.2022.02.004.

Skedgel, C., Mott, D.J., Elayan, S. and Cramb, A., 2024. A Longer Life or a Quality Death? A Discrete Choice Experiment to Estimate the Relative Importance of Different Aspects of End-of-Life Care in the United Kingdom. *MDM Policy & Practice*, 9(1), p.23814683241252425. 10.1177/23814683241252425.

Zhang, K. and Garau, M., 2020. *International Cost-Effectiveness Thresholds and Modifiers for HTA Decision Making*. London: Office of Health Economics.

Appendices

APPENDIX 1: PTO ALGORITHM



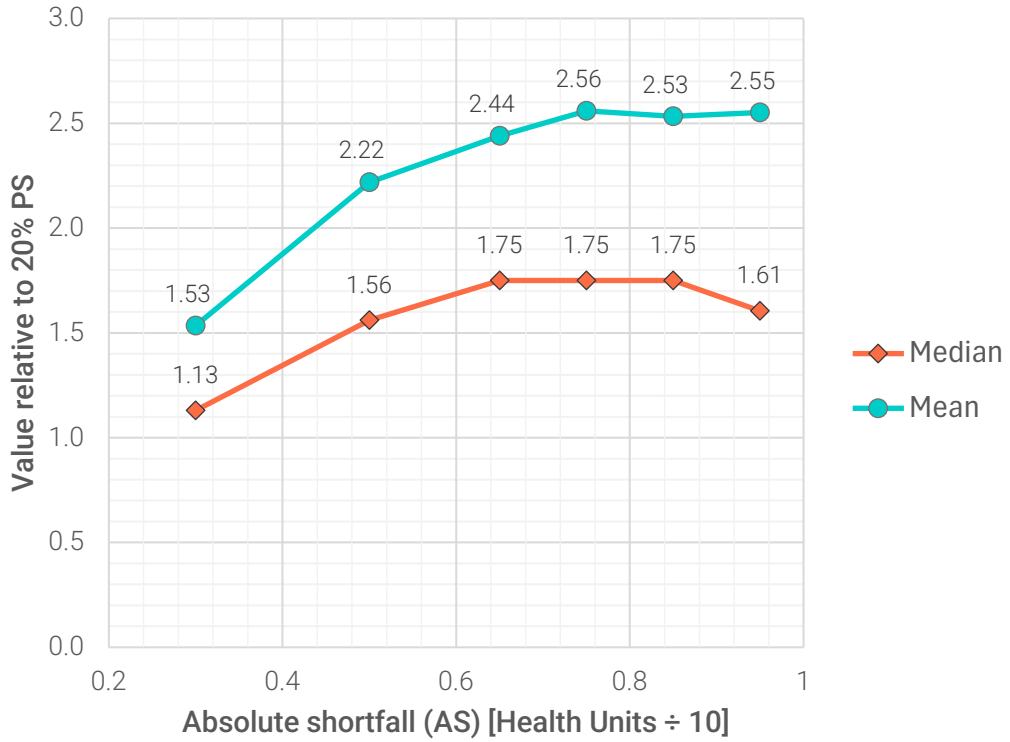
Notes: Numbers represent the number of patients presented to respondents in the prioritised or non-prioritised groups from the previous round, at each stage of the PTO. In the final rounds, respondents were presented with a range of patient numbers on a scale, and asked to select the number of patients which would make them indifferent between the two groups.

APPENDIX 2: DESCRIPTIVE STATISTICS FOR RELATIVE WEIGHTS ACROSS DIFFERENT PROPORTIONAL SHORTFALL SEVERITY LEVELS.

	Mean (Standard Error)	Median	Standard deviation	N respondents.
Weight values at each severity level				
30% PS	1.53***	1.13	1.31	990
	(0.04)			
50% PS	2.22***	1.56	2.17	990
	(0.07)			
65% PS	2.44***	1.75	2.42	990
	(0.08)			
75% PS	2.56***	1.75	2.55	990
	(0.08)			
85% PS *(NICE's severe)*	2.53***	1.75	2.58	990
	(0.08)			
95% PS *(NICE's very severe)*	2.55***	1.61	2.71	990
	(0.09)			

Notes: Heteroskedasticity-robust standard errors are reported in parentheses. Statistical significance of the difference in group means, compared to the null hypothesis of a mean value of 0, is presented. * p<0.05, ** p<0.01, *** p<0.001.

APPENDIX 3: MEAN AND MEDIAN VALUES FOR RELATIVE WEIGHTS ACROSS DIFFERENT ABSOLUTE SHORTFALL SEVERITY LEVELS.



Notes: We are unable to overlay NICE's AS thresholds as these are based on discounted AS and are not directly comparable to our elicitation of undiscounted AS. Furthermore, as noted, our PTO elicited preferences for a single reference age group (45-year-olds), and therefore AS levels always correspond with the same PS. Therefore, the pattern of severity weights over AS and PS are identical.

APPENDIX 4: SENSITIVITY ANALYSIS REMOVING POTENTIAL PROTEST RESPONSES

	Full sample (for comparison)	Removing respondents who selected the maximum value of patients on all PTO tasks	Removing respondents who refused to trade on any tasks in the PTO
Cut-offs (means)			
Severe (PS), %	48.30	48.33	48.40
Very severe (PS) %	64.48	64.54	64.59
Severe (AS)	16.34	16.36	16.39
Very severe (AS)	21.57	21.60	21.63
Weights (medians)			
PS (%)			
30	1.13	1.13	1.13
50	1.56	1.50	1.60
65	1.75	1.75	1.75
75	1.75	1.75	1.75
85	1.75	1.75	1.75
95	1.61	1.60	1.63
Respondents (N)	990	980	975
% of respondents excluded		1.01%	1.52%

Notes: Means are presented for the cut-offs. Due to the skewed nature of the open-ended PTO data, PTO estimates are presented in terms of medians. Consistent grey shading indicates a similar estimate compared with the full sample.

APPENDIX 5: SENSITIVITY ANALYSIS REMOVING POTENTIALLY INATTENTIVE OR LOW ENGAGEMENT RESPONSES

	Full analysis (for comparison)	Removing respondents who failed either of the understanding checks	Removing respondents with a low numeracy score (lowest 5%)	Removing respondents with a low completion time (lowest 5%)
Cut-offs (means)				
Severe (PS), %	48.30	48.39	48.30	48.82
Very severe (PS) %	64.48	64.77	64.54	65.06
Severe (AS)	16.34	16.45	16.39	16.53
Very severe (AS)	21.57	21.66	21.67	21.78
Weights (medians) PS (%)				
30	1.13	1.13	1.13	1.15
50	1.56	1.63	1.60	1.63
65	1.75	1.80	1.75	1.80
75	1.75	1.85	1.75	1.85
85	1.75	1.80	1.75	1.80
95	1.61	1.67	1.62	1.75
Respondents (N)	990	926	937	941
% of respondents excluded		6.46%	5.35%	4.95%

Notes: Means are presented for the cut-offs. Due to the skewed nature of the open-ended PTO data, PTO estimates are presented in terms of medians. Consistent grey shading indicates a similar estimate compared with the full sample. Red highlighted cells indicate higher estimates than the full sample.

About us
With over 60 years of expertise, the Office of Health Economics (OHE) is the world's oldest independent health economics research organisation. Every day we work to improve health care through pioneering and innovative research, analysis, and education.

As a global thought leader and publisher in the economics of health, health care, and life sciences, we partner with Universities, Government, health systems and the pharmaceutical industry to research and respond to global health challenges.

As a government-recognised Independent Research Organisation and not-for-profit, our international reputation for the quality and independence of our research is at the forefront of all we do. OHE provides independent and pioneering resources, research and analyses in health economics, health policy and health statistics. Our work informs decision-making about health care and pharmaceutical issues at a global level.

All of our work is available for free online at www.ohe.org.

Areas of expertise

- Evaluation of health policy
- The economics of health care systems
- Health technology assessment (HTA) methodology and approaches
- HTA's impact on decision making, health care spending and the delivery of care
- Pricing and reimbursement for biologics and pharmaceuticals, including value-based pricing, risk sharing and biosimilars market competition
- The costs of treating, or failing to treat, specific diseases and conditions
- Drivers of, and incentives for, the uptake of pharmaceuticals and prescription medicines
- Competition and incentives for improving the quality and efficiency of health care
- Incentives, disincentives, regulation and the costs of R&D for pharmaceuticals and innovation in medicine
- Capturing preferences using patient-reported outcomes measures (PROMs) and time trade-off (TTO) methodology
- Roles of the private and charity sectors in health care and research
- Health and health care statistics

CONTRACT RESEARCH REPORT
NOVEMBER 2024

ohe.org

The Office of Health Economics A Company Limited by Guarantee of Registered No.09848965
OHE Consulting Ltd Registered Company No.09853113
OHE is a Charity Registration No.1170829
Registered Office 2nd Floor Goldings House, Hay's Galleria, 2 Hay's Lane, London, SE1 2HB