Aversion to Health Inequality – Pure, Income-Related and Income-Caused

Matthew Robson
Owen O'Donnell
Tom Van Ourti

EsCHER

ERASMUS CENTRE
FOR HEALTH ECONOMICS
ROTTERDAM

Erasmus University Rotterdam
Making Minds Matter

Title
Aversion to Health Inequality – Pure, Income-Related and Income-Caused

Authors
Matthew Robson. Erasmus School of Economics and Erasmus Centre for Health Economics Rotterdam, Erasmus University Rotterdam; Tinbergen Institute.
Owen O'Donnell. Erasmus Centre for Health Economics Rotterdam, Erasmus School of Health Policy and Management and Erasmus School of Economics, Erasmus University Rotterdam; Tinbergen Institute.
Tom Van Ourti. Erasmus Centre for Health Economics Rotterdam, Erasmus School of Health Policy and Management and Erasmus School of Economics, Erasmus University Rotterdam; Tinbergen Institute.

Corresponding author and contact details:
Matthew Robson: robson@ese.eur.nl

Aversion to Health Inequality –
Pure, Income-Related and Income-Caused

Matthew Robson [†]   Owen O'Donnell [‡]   Tom Van Ourti [§¶]

April 12, 2023

Abstract

We design a novel experiment to identify aversion to pure (univariate) health inequality separately from aversion to income-related and income-caused health inequality. Participants allocate resources to determine health of individuals. Identification comes from random variation in resource productivity and in information on income and its causal effect. We gather data (26,286 observations) from a UK representative sample (n=337) and estimate pooled and participant-specific social preferences while accounting for noise. The median person has strong aversion to pure health inequality, challenging the health maximisation objective of economic evaluation. Aversion to health inequality is even stronger when it is related to income. However, the median person prioritises health of poorer individuals less than is assumed in the standard measure of income-related health inequality. On average, aversion to that inequality does not become stronger when low income is known to cause ill-health. There is substantial heterogeneity in all three types of inequality aversion.

Keywords: Inequality Aversion, Social Preferences, Health, Income, Experiment

JEL: C90, D30, D63, I14

1 Introduction

Aversion to health inequality is multifaceted. Concern about differences in health between individuals irrespective of their non-health characteristics motivates prioritisation of health gains to the least healthy. Concern about systematic differences in health between individuals distinguished by income motivates prioritisation of health gains to the poor. Elicitation of social preferences over the distribution of health usually fails to separate aversion to pure health inequality (the first case) from aversion to income-related health inequality (the second case). This risks confounding one type of aversion with the other and so biasing distributionally sensitive evaluation of health programmes that use the elicited preferences. We design a novel experiment to separately identify the two facets of health inequality aversion.

In the experiment, participants play the role of social decision makers allocating resources to determine the health of individuals. By varying resource productivity between individuals, we force trade-offs between increasing aggregate health and decreasing health inequality. This identifies aversion to inequality. In one treatment, individuals are anonymous. This identifies aversion to pure health inequality. In a second treatment, we label individuals by randomly assigned incomes. Responses to this information, as well as the trade-offs made between health maximisation and equalisation, simultaneously identify both aversion to income-related health inequality and aversion to pure health inequality.

Aversion to income-related health inequality may derive from the consequences of that inequality for the distribution of well-being defined over health and income. A given marginal distribution of health generates greater inequality in well-being when health and income are positively correlated. This may motivate prioritisation of poorer individuals in the distribution of health resources in order to compensate material disadvantage with improved health.

1

Aversion to differences in health by income could also derive from a perception that lower income causes worse health and a belief that this is unfair. Many may see injustice in the poor being in worse health because they cannot afford medicine or nutritious food. This would motivate prioritisation of health gains to the poor beyond that deemed appropriate to compensate for poverty through improved health.

To identify the extent to which belief in a causal effect of income on health strengthens aversion to income-related health inequality, we tell participants, in a third treatment, the proportion of differences in resource productivity that is caused by differences in incomes. This induces exogenous variation in beliefs about income-health causality that is used to identify the effect of these beliefs on prioritisation of the health of poorer (or richer) individuals. Any systematic shift in allocations implies aversion to income-caused health inequality.

We minimise restrictions on the social preferences we can identify by asking participants to allocate resources when richer individuals are both advantaged and disadvantaged in the production of health. We allow preference for a negative, positive, or zero association between health and income. We can identify preference for prioritisation by income that weakens, not only strengthens, when income is known to cause health.

We ran the experiment online and recruited a broadly representative sample of the UK adult population. Repeated choices by 337 participants give 26,286 observations that we use to obtain both pooled and participant-specific estimates. The latter is potentially important given the substantial heterogeneity found in other estimates of social preferences (Cappelen et al., 2007; Fisman, Kariv, and Markovits, 2007; Hurley, Mentzakis, and Walli-Attaei, 2020). A within-subject design further increases statistical power.

We estimate parameters of a social welfare function that aggregates over a population health profile through a social utility function of each individual's health that is potentially weighted by a function of their income (Makdissi and Yazbeck, 2016). The revealed will-

ingness to sacrifice health maximisation for less inequality identifies the degree of concavity of the utility function that reflects aversion to pure health inequality. If the allocations are independent of income, then the weights are constant and the model collapses to an Atkinson (1970) welfare function that others have used to estimate health inequality aversion (Dolan, 1998; Dolan and Tsuchiya, 2011; Robson et al., 2017). Prioritisation by income identifies weights that decrease or increase with rising income to reflect aversion to pro-rich and pro-poor health inequality, respectively. We use a general weighting function that nests one that underpins the most common measure of income-related health inequality, the (extended) concentration index (Wagstaff, Paci, and Van Doorslaer, 1991; Wagstaff, 2002; O'Donnell et al., 2008). Consequently, the model (Makdissi and Yazbeck, 2016) also encompasses Wagstaff's achievement index that penalises mean health for pro-rich health inequality without allowing for aversion to pure health inequality (Wagstaff, 2002). We extend the weighting function to include two parameters that represent aversion to income-related health inequality and the extent to which it intensifies (or weakens) with beliefs about causality. We use responses to the causal information treatment to separately identify these parameters and so the extent to which the income weights shift with causality beliefs.

Non-parametric analysis reveals that no participant maximises aggregate health. All are averse to pure health inequality. This is consistent with evidence that most people state a willingness to sacrifice health maximisation in order to reduce health inequality (Abellan-Perpiñan and Pinto-Prades, 1999; Ubel and Loewenstein, 1996; Ratcliffe, 2000; Gyrd-Hansen, 2004; Dolan et al., 2005; Richardson et al., 2012). We find that more than three quarters of participants are prepared to sacrifice health maximisation to an extent that involves giving more (not just any) resources to individuals who benefit less from them. On average, priority is given to the health of the poor. This is driven by less than a quarter (23%) of participants who allocate to ensure that the poorest individual gets more health than the richest. On average, information about the extent to which income causally determines

health has no impact on the allocations, which suggests that aversion to income-related health inequality is not contingent on beliefs about causality.

Parametric analysis gives a pooled estimate of constant relative health inequality aversion of 1.4, which indicates moderate prioritisation of the least healthy. However, there is substantial heterogeneity, as in related studies (Cropper, Krupnick, and Raich, 2016; Hurley, Mentzakis, and Walli-Attaei, 2020). The median participant-specific estimate of this parameter is 3.2, indicating that a majority displays substantial aversion to pure health inequality. The estimate increases slightly to 3.5 when income weights are estimated simultaneously.

Both the pooled and median estimates indicate weak prioritisation of the health of the poor after taking account of aversion to pure health inequality. Weights decline with rising income rank but less rapidly than those implicit in the concentration index. That index forces aversion to health inequality and prioritisation of health by income into a single parameter that determines the weights. We separate these two dimensions of social preferences and so avoid limitations of the concentration and achievement indices. One is a lack of aversion to health differences that are not related to income. Another is a restriction on the (abbreviated) social welfare function that, paradoxically, implies that welfare increases with health inequality provided it is pro-poor (Makdissi and Yazbeck, 2016). In fact, our parametric estimates indicate that a little less than a quarter of the UK population is pro-rich — all else equal, they prioritise the health of richer over poorer individuals. Both pooled and median estimates of the degree of aversion to income-related health inequality are insensitive to information on causality.

Specifying social preferences in the form of an Atkinson social welfare function based on a relative conception of inequality combined with rank-dependent income weights fits the data slightly better than both a Kolm-Pollak social welfare function based on absolute inequality (Pollak, 1971; Kolm, 1976) and share-dependent income weights. Neither the Atkinson welfare function nor Wagstaff's achievement index is sufficiently flexible by itself

to capture median preferences. There is clear evidence of consequentialism — participants allocate resources to optimise the distribution of health.

Beyond our empirical findings, we make three methodological contributions that can improve the reliability and extend the scope of evidence on health inequality aversion. This is the first study to simultaneously estimate aversion to pure health inequality and income weights that together determine aversion to income-related health inequality. Many studies elicit aversion to only one type of health inequality (Abasolo and Tsuchiya, 2004; Dolan and Tsuchiya, 2011; Cookson et al., 2018; Robson et al., 2017; Hardardottir, Gerdtham, and Wengström, 2021). Some elicit aversion to pure and income-related health inequality separately, with each of these social preferences represented by a different single-parameter welfare function (Dolan and Tsuchiya, 2009; Hurley, Mentzakis, and Walli-Attaei, 2020; McNamara, Tsuchiya, and Holmes, 2021). One study estimates a two-parameter welfare function but imposes independence between the aversion to pure health inequality and the socioeconomic weights that the parameters represent (Pinho and Botelho, 2018). Simultaneous estimation of the two parameters is important because restricting attention to one fails to distinguish between concern for the poor and concern for the less healthy. This confounding will upwardly bias willingness to prioritise health of poorer individuals when health and income are restricted to be positively correlated, which, as far as we know, is the case in all previous studies. In our set up, income is orthogonal to (potential) health.

This is also the first study to allow aversion to income-related health inequality to depend on the extent to which income causes better (or even worse) health. An outcome-focused concern about poor people also being unhealthy can be distinguished from a procedural concern about people being unhealthy because they are poor (Schokkaert and Devooght, 2003). A presumption of greater aversion to income-related health inequality that arises from a causal effect of income on health has not been tested until now. If aversion were to depend

on causality, then causal evidence would be required not only to design policies that reduce income-related health inequality but also to assess the normative rationale for such policies.

A third innovation is that our design and empirical approach make it possible to estimate participant-specific inequality aversion while allowing for noise in the participant's choices. Most studies elicit participant-specific preferences without allowing for noise (Abasolo and Tsuchiya, 2004; Dolan and Tsuchiya, 2009; Dolan and Tsuchiya, 2011; Robson et al., 2017; Cookson et al., 2018; Hardardottir, Gerdtham, and Wengström, 2021; McNamara, Tsuchiya, and Holmes, 2021). This risks the mistaken inference of preferences from choices that are simply random errors. These studies often drop a substantial proportion responses that appear irrational. Some studies allow for noise, but only when pooling data across participants (Edlin, Tsuchiya, and Dolan, 2012; Hurley, Mentzakis, and Walli-Attaei, 2020), and so losing the opportunity to estimate participant-specific preferences. We avoid these limitations by estimating inequality aversion from allocations of resources in a series of constrained optimisation problems that present equity-efficiency trade-offs. We allow for the mistakes each participant inevitably makes by estimating a random behavioural model to infer preferences from error-prone allocations that are assumed to be optimal only on average (Harless and Camerer, 1994; Conte and Moffatt, 2014; Robson, 2021).

The next section describes the experiment. Section 3 presents models of social welfare that we fit to the data. Section 4 explains estimation of the model parameters allowing for noise. Section 5 presents results from non-parametric and parametric analyses and compares the data fit of alternative models. Section 6 gives an illustrative application to policy evaluation that demonstrates the value of obtaining participant-specific estimates and the importance of estimating and using aversion to pure health inequality and to income-related weights simultaneously. Section 7 compares our estimates to previous evidence, considers explanations for findings, and acknowledges limitations. The final section concludes.

2 Experiment Design

2.1 General Setup

We ran an online interactive experiment to elicit social preferences over the distribution of health. Participants were recruited through Prolific and are broadly representative of the UK adult population.

Preferences are inferred from choices in a series of constrained optimisation problems that pose equity-efficiency trade-offs. In each round, a participant was assigned a randomly-generated *budget* and asked to allocate *resources* to three hypothetical *individuals*. They were forced to exhaust the budget. The participant was told that the *health* of each individual would be the product of the resources allocated to that individual and an individual-specific productivity factor referred to as a *multiplier* (Arrow, 1971). These varied from round to round.

Participants were told that health is the number of years an individual lives adjusted for illness or disability. They were given an example to encourage them to interpret this as equivalent years lived in full health — quality-adjusted life years (QALYs) (Appendix A.2).

The task was completed using an online screen interface designed in R Shiny that is shown in Figure 1. The participant allocates resources using sliders, and can use arrow keys to refine allocations. Resources and the resulting health outcomes are shown graphically by the blue and black bars, respectively, and numerically in the table. The (remaining) budget is shown on the left of the screen. Summary measures are on the right. The *Resource Gap* is the largest absolute difference between resources allocated to two individuals. The *Health Gap* is the equivalent for health. *Total Health* is the sum of the health outcomes. To encourage deliberation, minimum timers were placed on each round.

Figure 1: Experiment Interface
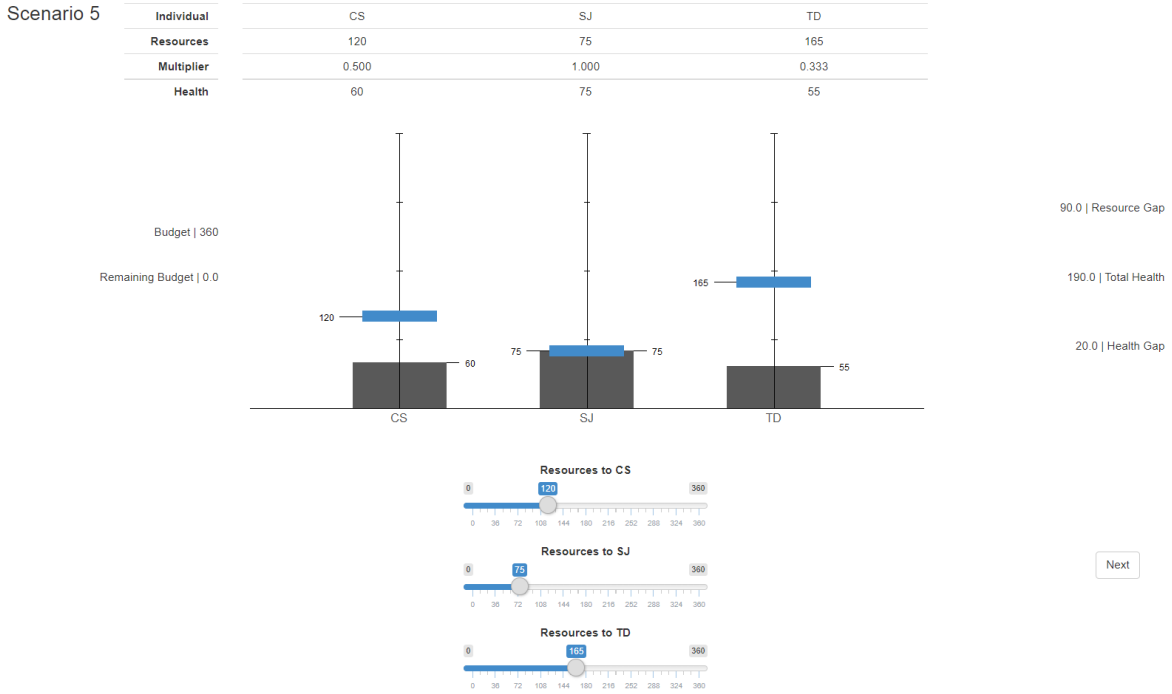
2.2 Treatments

The experiment had three within-subject treatments that are used to identify aversion to A) pure health inequality, B) income-related health inequality, and C) income-caused health inequality. The treatments differ with respect to information participants were given about a) the identity of the three individuals and b) how the multipliers are determined. All participants faced the same chronology of treatments — A, B, and, finally, C. We could not randomise the order as each treatment reveals more information.

2.2.1 Treatment A: Anonymous

In Treatment A, participants were given no information about the identity of the individuals, who were labeled by randomly drawn initials (e.g. *CS*, *SJ*, and *TD*). The multipliers (p_i) changed across 10 rounds (see Appendix A.3 Table A1). Between individual differences in

p_i force trade-offs between health maximisation and equalisation that are used to identify aversion to health inequality. The distribution of p_i over 10 rounds is the same irrespective of the screen position of the individuals. This ensures that comparisons between the positions can be made without bias. The order of the 10 rounds were randomised across participants and the set of initials was drawn randomly for each round and participant.

2.2.2 Treatment B: Income

In Treatment B anonymity was lifted. Participants were told the income of each individual. Each participant again allocated resources in 10 randomly ordered rounds. The distribution of multipliers over the rounds is identical to that used in Treatment A, allowing direct comparison between the treatments. In each round and for each participant, we randomly selected (without replacement) three incomes from $\{£5,000, £10,000, £25,000, £50,000, £100,000\}$ and used them to label the three individuals, e.g. *HD: £10,000*.[1] Since different sets of three random draws give the same income ranks but different income shares, we can distinguish between aversion to income-rank and income-share related health inequality.

2.2.3 Treatment C: Income-Causation

In Treatment C, participants were told not only the income (x_i) of each of the three individuals but also the extent to which income differences cause differences in multipliers. In the two scenarios presented, the set of multipliers was $\{0.33, 0.5, 1\}$.[2] In one scenario, these multipliers monotonically increase with income — health interventions are more effective at higher incomes. In the other, they monotonically decrease as income rises. For each scenario, the participant was told, in three different rounds, that the percentage of the

---

[1]The order of the income-labelled individuals on the screen was also randomised between participants, either increasing or decreasing with income. The set of incomes was chosen to give sufficient variation for identification.

[2]To facilitate comparison, this is the same set used in rounds 7 and 8 of Treatments A and B (Appendix A.3 Table A1).

multiplier differences caused by income is 0%, 100%, and $X\%$, where X was randomly drawn from $\{20, 40, 60, 80\}$. We refer to this as *causal information* (I). See Appendix A.5 for the script and interface.

Each participant completed six rounds in this treatment (Appendix A.3 Table A2). The design ensures that correlation between the multipliers and health is orthogonal to the randomly assigned causal information.[3] Orthogonality allows aversion to income-caused health inequality to be separately identified from aversion to non-causal income-health association. This is done by leveraging the variation in resource allocations made from round to round in response to the causal information.

2.3 Timing and sample

After conducting two pilot experiments (Appendix A.6), the experiment was conducted in two sessions over a three-week period. In the first session (December 14-17, 2021), participants received instructions, followed an interactive tutorial, answered follow-up questions of comprehension (Appendix A.2), did Treatment A (10 rounds), and completed a questionnaire about sociodemographic characteristics and beliefs. In the second session (December 18, 2021 - January 5, 2022), the participants followed a shorter tutorial, did Treatment B (10 rounds), completed a belief elicitation exercise (Appendix A.4), did Treatment C (6 rounds), and completed a final questionnaire. The median completion time was 29.0 minutes in the first session and 27.6 minutes in the second.[4] See Appendix A.1 for a graphical overview of the experiment.

---

[3]In addition, each of a) the incomes of the three individuals, b) their screen order (increasing or decreasing with income), c) the order of the scenarios distinguished by whether multipliers are increasing or decreasing with income, and d) the order of presenting the causal information as 0%, 100%, and $X\%$ was (separately) randomised between participants.

[4]Participants were paid £3.50 for the first session and £5 for the second. The average payment was £9.01 per hour over the two sessions.

Of the 402 participants who completed the first session, 21 (5.2%) did not complete the second session. We drop an additional 44 participants who incorrectly answered 3 or more of 5 comprehension questions after the tutorial in the first session. We test robustness to this exclusion (Appendix D.4). This leaves an analysis sample of 337 participants with complete data from both sessions. The sample is younger than the UK adult population but representative with respect to sex and ethnicity (see Appendix B Table B1).

The 337 participants made choices in 26 rounds across three treatments, with each round involving allocations to three individuals. This gives a total of 26,286 ($= 337 \times 26 \times 3$) observations. Participants were only shown information on income of the individuals in treatments B and C (16 rounds), and so there are 16,176 observations for income-related allocations. Analyses conducted only with Treatment A or Treatment B data use 10,110 ($= 337 \times 10 \times 3$) observations in each case. Table 1 summarises the data used for estimation.

Table 1: Summary of data

Variable	Notation	Definition	Mean	Range	Obs.
Budget	m	Total resources per round	261.5	$[180, 360]$	26,286
Resources	y_i	Resources to individual	87.2	$[0, 360]$	26,286
Resource Share	$\tilde{y}_i$	y_i/m	0.333	$[0, 1]$	26,286
Multiplier	p_i	Marginal effect of y_i on h_i	0.679	$[0.333, 1]$	26,286
Relative Multiplier	$\tilde{p}_i$	$p_i/\sum p_i$	0.333	$[0.14, 0.54]$	26,286
Health	h_i	$p_i \times y_i$	53.7	$[0, 360]$	26,286
Health Share	$\tilde{h}_i$	$h_i/\sum h_i$	0.333	$[0, 1]$	26,286
Income	x_i	Income of individual	£38k	$[£5k, £100k]$	16,176
Income Rank	$r(x_i)$	i/N with i ordered by x_i	0.666	$[0.333, 1]$	16,176
Causal Information	I	% Δp_i caused by Δx_i	49.7	$[0, 100]$	6,066

Note: Data from 337 participants, across 26 rounds. $N = 3$ is the number of individuals in each round.

3 Theory

3.1 Social Welfare Function

Participants take the role of a social decision maker (SDM) who is assumed to maximise social welfare (W) that is an aggregation of health (h_i) (QALYs) of individuals (i) in a population of size N ($= 3$ in the experiment). We use participants' resource allocations, and how they respond to variation in multipliers, incomes, and causal information, to identify parameters of a weighted utilitarian social welfare function (SWF) (Fleming, 1952; Harsanyi, 1955; Vickrey, 1960),

$$W = \sum_{i=1}^{N} \omega_i \, U\left(h_i\right),$$ (1)

where $0 \leq \omega_i \leq 1 \; \forall i$ and $\sum_{i=1}^{N} \omega_i = 1$ are weights. The social utility function, $U(.)$, is common across individuals and represents the social preference for health of the SDM (Wagstaff, 1991; Bleichrodt, 1997; Dolan, 1998; Bleichrodt, Doctor, and Stolk, 2005). It is assumed to be concave and could be linear. If it is strictly concave and the weights are constant ($\omega_i = \omega \; \forall i$), then there is aversion to pure health inequality in the sense that a transfer of health from a healthier to a less healthy individual increases welfare (Wagstaff, 1991; Dolan, 1998).

We restrict attention to the iso-elastic function, $U\left(h_i\right) = \left(h_i^{1-\varepsilon} - 1\right) / \left(1 - \varepsilon\right)$ for $\varepsilon \geq 0$ and $\varepsilon \neq 1$, and $U\left(h_i\right) = ln\left(h_i\right)$ for $\varepsilon = 1$ (Atkinson, 1970). Welfare can be measured by the equally distributed equivalent (EDE) level of health,

$$h_{EDE} = \left(\sum_{i=1}^{N} \omega_i \, h_i^{1-\varepsilon}\right)^{\frac{1}{1-\varepsilon}}.$$ (2)

With constant weights, the parameter ε captures the trade-off the SDM is willing to make between maximising aggregate health and equalising the distribution of health. With $\varepsilon > 0$,

there is willingness to forgo health maximisation in order to reduce inequality: aversion to pure health inequality. As ε increases, this inequality aversion intensifies and social welfare becomes more sensitive to the lowest levels of health. As $\varepsilon \to \infty$, the SDM's preferences approach the Rawlsian maximin — only health improvements experienced by the least healthy raise welfare.

The weights can be a function of non-health characteristics that are possibly correlated with health and may even determine health (Wagstaff, 1991; Dolan and Tsuchiya, 2009; Makdissi and Yazbeck, 2016).[5] This allows the social value of an individual's health to depend on their non-health characteristics.

We use the experiment to ascertain whether the weights depend on income.[6] If they do, it may be because income is considered to be a causal determinant of health and this is judged to be an unfair source of health differences. Alternatively, priority may be given to the health of poorer people to compensate for their material disadvantage. The experiment is designed to distinguish between these two motivations for income-dependent weights. We do not restrict the weights to be decreasing in income. Some may prioritise the health of the economically better off due to a belief that the marginal social value of health is increasing in income.[7]

---

[5]We assume that health does not directly determine the weights. This distinguishes eq.(1) from the (nonlinear) rank-dependent QALY model (Bleichrodt, Diecidue, and Quiggin, 2004; Bleichrodt, Doctor, and Stolk, 2005). Our experiment set-up does not permit a formal test of eq.(1) with $\omega_i = \omega \ \forall i$ against the rank-dependent QALY model consisting of eq.(1) with $\omega_i = v(i/N) - v((i-1)/N)$, $v()$ non-decreasing, and $h_i \leq h_{i-1} \ \forall i$. However, in Appendix E we show that a model consisting of concave $U()$ and health-dependent weights does not fit the data substantially better than a restricted version that imposes constant weights.

[6]The income-related distribution of health can be evaluated without considering the direct effect of the income distribution on welfare provided the SDM's evaluation of individual well-being is additively separable in health and income (Makdissi and Yazbeck, 2016). Conditional on this restriction, income-dependent weights allow welfare evaluation of interventions that impact health differently depending on income without affecting the income distribution. With our experimental setup, even with additive separability, it would not be possible to identify parameters of a SWF over well-being determined by both health and income. Such an extension might demand too much cognitive effort from participants.

[7]This is not inconsistent with the assumed additive separability of social utility because the marginal effect of health on social welfare depends not only on the marginal utility of health but also on the social weight, which can depend on income.

We elicit preferences by asking participants to allocate resources (y_i) while revealing the consequences for the distribution of health, which are generated by the health production function, $h_i = p_i \times y_i$.[8] Maximisation of social welfare (h_{EDE}) subject to a binding budget constraint, $\sum_{i=1}^{N} y_i = m$, gives the optimal allocations,

$$
y_i^* = m \left[1 + \sum_{\substack{j \neq i}}^{N} \frac{p_i}{p_j} \left(\frac{\omega_j \, p_j}{\omega_i \, p_i} \right)^{\frac{1}{\varepsilon}} \right]^{-1} \forall_i.
\tag{3}
$$

3.2 Equity Weights

Specification of the weights distinguishes aversion to pure health inequality — differences in the health anonymous individuals — from aversion to income-related health inequality and the latter from aversion to income-caused health inequality.

3.2.1 Constant Weights

If a SDM considers health to be the only characteristic that is relevant to the allocation of resources, then they will give all individuals equal weight:

$$
\omega_i^A = 1/N.
\tag{4}
$$

3.2.2 Income-Dependent Weights

Let x_i be income and order individuals from poorest to richest, $x_{i-1} \leq x_i$. We capture aversion to income-related health inequality by weights that are a function of income ranks, $r(x_i) = i/N$. The assumption that income rank, and not income level, affects the social value of health is consistent with the predominant use of rank-dependent (concentration) in-

---

[8]As we discuss in section 7, a concave production function would be more realistic but would complicate the elicitation task.

dices to measure income-related health inequality (Wagstaff, Paci, and Van Doorslaer, 1991; O'Donnell et al., 2008). We specify the weights in a way that is standard for these indices (Donaldson and Weymark, 1980; Donaldson and Weymark, 1983; Yitzhaki, 1983; Wagstaff, 2002):

$$\omega_i^B(\beta) = \int_{r(x_{i-1})}^{r(x_i)} \beta (1-q)^{\beta-1} dq, \tag{5}$$

where $\beta > 0$ reflects the direction and degree of health prioritisation by income.[9] With $\beta = 1$, the weights are constant — there is no aversion to income-related health inequality. There can still be aversion to health inequality. But, if there is, the welfare loss generated by that inequality is not larger when part of it is related to income. With $\beta > 1$, the weights decrease monotonically as income rank increases — there is aversion to pro-rich health inequality. With $1 < \beta < 2$, the weight-income rank function is negatively sloped and concave. At $\beta = 2$, it is the linear weighting function of the standard concentration index: $\omega_i^B(2) = (2N - 2i + 1)/N^2$ (Wagstaff, Paci, and Van Doorslaer, 1991). With $\beta > 2$, the function is convex and relative weights on poorer individuals increase. As $\beta \to \infty$, the weights approach zero for all but the poorest individual.

While measurement of income-related health inequality usually imposes weights that decline with income (Wagstaff, 2002; Bleichrodt and van Doorslaer, 2006; Erreygers, Clarke, and Van Ourti, 2012), eq.(5) can accommodate aversion to pro-poor inequality. With $0 < \beta < 1$, the weights increase monotonically and convexly with income rank. Values closer to 0 give greater weight to the very rich.[10]

---

[9]We calculate the weight for individual i over the interval $[r(x_{i-1}), r(x_i)]$ to account for the small-sample bias that arises when $\beta \neq 2$ (Erreygers, Clarke, and Van Ourti, 2012). Solving the definite integral simplifies eq.(5) to $\omega_i^B(\beta) = \left(\frac{N-i+1}{N}\right)^\beta - \left(\frac{N-i}{N}\right)^\beta$.

[10]One could allow for an even broader array of social preferences with the tractable, yet flexible, beta density function. This would involve specifying, $\omega_i^B(\alpha, \beta) = \int_{r(x_{i-1})}^{r(x_i)} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} q^{\alpha-1}(1-q)^{\beta-1} dq$, with $\Gamma(.)$ representing the gamma function and $\alpha, \beta > 0$. This weighting function collapses to eq.(5) when $\alpha = 1$. In

The model of social preferences given by eq.(2) and eq.(5) nests: a) an Atkinson SWF that allows for aversion to relative inequality in the univariate distribution of health but with no aversion to income-related health inequality ($\beta = 1$) and b) the SWF of Wagstaff's (2002) achievement index that allows for aversion to income-related health inequality through the extended concentration index but with no aversion to pure health inequality ($\varepsilon = 0$).

3.2.3   Causality-Dependent Weights

We allow the strength of aversion to income-related health inequality to depend on beliefs about the extent to which variation in income causes variation in health by generalising eq.(5) to

$$\omega_i^C (\beta_1, \beta_2, \lambda) = \int_{r(x_{i-1})}^{r(x_i)} \beta_1 \beta_2^\lambda (1 - q)^{\beta_1 \beta_2^\lambda - 1} \, dq, \tag{6}$$

where $\beta_1 > 0$ and $\beta_2 > 0$ capture aversion to income-related and income-caused health inequality, respectively, and $\lambda \in [-1, 1]$ represents beliefs about causality.[11] The sign of λ indicates the perceived direction of any causal effect of income on health. The magnitude of this parameter is the perceived proportion of health differences caused by income differences.

If the SDM believes that none of the observed health inequality is caused by income differences, then $\lambda = 0$ and eq.(6) collapses to eq.(5) with $\beta = \beta_1$. In that case, while the SDM may be averse to income-caused health inequality ($\beta_2 \neq 1$), this would not affect their resource allocations because they believe there is no such inequality. If such a SDM favours either poorer or richer individuals when allocating resources, this behaviour must be motivated by a concern about non-causal income-related health inequality, which will be reflected in the parameter β_1.

---

its unrestricted form, it allows for concavely increasing weights with income, weights that center around the median income rank, and weights with positive or negative skewness.

[11]Solving the definite integral, eq.(6) simplifies to $\omega_i^C (\beta_1, \beta_2, \lambda) = \left(\frac{N-i+1}{N}\right)^{\beta_1 \beta_2^\lambda} - \left(\frac{N-i}{N}\right)^{\beta_1 \beta_2^\lambda}$.

If the SDM believes that income differences cause health differences, at least to some extent, then $\lambda \neq 0$ and β_2 plays a role in the determination of social welfare and in resource allocations. The weight (and allocations) to the income poor, which is determined by $\beta_1\beta_2^{\lambda}$, will increase if either a) the SDM believes there is a positive causal effect of income on health ($\lambda > 0$) and is averse to the resulting pro-rich health inequality ($\beta_2 > 1$), or b) the SDM believes there is a negative causal effect of income on health ($\lambda < 0$) and likes the resulting pro-poor health inequality ($\beta_2 < 1$). The weight to the income poor will decrease if either c) income is believed to have a negative causal effect on health ($\lambda < 0$) and there is aversion to the resulting pro-poor inequality ($\beta_2 > 1$), or d) income is believed to have a positive causal effect on health ($\lambda > 0$) and there is preference for the resulting pro-rich inequality ($\beta_2 < 1$). If the SDM is indifferent to income-caused health inequality ($\beta_2 = 1$), then beliefs about the direction and magnitude of that inequality (λ) do not affect the weights.[12]

To separately identify β_1 and β_2, we use induced random variation in beliefs about the extent to which income-related health inequality (in each direction) is caused by income differences. We assume that participants believe entirely the causal information (I) provided in Treatment C on the percentage of multiplier (potential health) differences between individuals that is caused by income differences (section 2.2.3). Under this assumption, we define $\lambda = I\left(\mathbb{1}(Cov\left(p_i, x_i\right) > 0) - \mathbb{1}(Cov\left(p_i, x_i\right) < 0)\right)/100$. This fixes λ at zero and at two positive and two negative values for each participant. We can then identify β_1 and β_2 from resource allocations made in these different cases.

---

[12]See Appendix C for a summary of all cases.

3.3 Alternative Social Welfare and Weighting Functions

3.3.1 Non-Consequentialism

Some participants may allocate resources without considering consequences for the distribution of health. We can accommodate such non-consequentialist ethics by substituting y_i for h_i as the argument of the iso-elastic social utility function in eq.(1) and solving for the EDE allocation of resources,[13]

$$y_{EDE}^N = \left(\sum_{i=1}^N \omega_i y_i^{1-\varepsilon} \right)^{\frac{1}{1-\varepsilon}}. \tag{7}$$

In this case, the optimal allocations are a function of the budget, the weights, and inequality aversion, but not the health production function parameters:

$$y_i^{N*} = m \left[1 + \sum_{j \neq i}^N \left(\frac{\omega_j}{\omega_i} \right)^{\frac{1}{\varepsilon}} \right]^{-1} \forall_i. \tag{8}$$

3.3.2 Absolute Invariance

Equation (2) captures willingness to sacrifice health maximisation for less *relative* health inequality. To accommodate SDMs who are concerned about *absolute* health inequality, we also consider the Kolm-Pollak family of SWFs (Pollak, 1971; Kolm, 1976), extended to allow aggregation of health to depend on income. We do this by specifying $U(h_i) = -exp(\theta h_i), \theta < 0$ in eq.(1). The respective EDE health is

$$h_{EDE}^K = \frac{1}{\theta} \ln \left[\sum_{i=1}^N \omega_i \exp(\theta h_i) \right]. \tag{9}$$

The allocation of resources that maximises this indicator of welfare is

---

[13]While this ignores consequences for the distribution of health, it is not entirely non-consequentialist since the SDM evaluates the distribution of resources by taking a concave aggregation to get welfare.

18

$$y_i^{K^*} = \left[m - \sum_{j \neq i}^{N} \frac{1}{\theta p_j} ln \left(\frac{p_i \omega_i}{p_j \omega_j} \right) \right] \left[1 + \sum_{j \neq i}^{N} \frac{p_i}{p_j} \right]^{-1} \forall_i. \tag{10}$$

The non-consequentialist SDM who is concerned about absolute resource inequality would make an optimal allocation that is given by eq.(10) with multipliers (p_i, p_j) set to 1.

3.3.3 Share-dependent Weights

While specifying weights as a function of income ranks is consistent with the predominance of rank-dependent measurement of income-related health inequality (Wagstaff, Paci, and Van Doorslaer, 1991; O'Donnell et al., 2008), it does not allow for the possibility a SDM pays attention to cardinal incomes in prioritising the health of individuals. To accommodate social preferences of this kind, we specify weights that are a function of income shares, $\tilde{x}_i = x_i / \sum_{j=1}^{N} x_j$, and normalize them to sum to 1:[14]

$$\omega_i^D(\gamma) = \frac{(1 - \tilde{x}_i)^{\gamma - 1}}{\sum_{j=1}^{N} (1 - \tilde{x}_j)^{\gamma - 1}}. \tag{11}$$

With $\gamma = 1$, the weights are constant and there is no aversion to income-share-related health inequality. With $\gamma > 1$, the weights decline with increasing income share, reflecting aversion to pro-rich inequality. With $\gamma < 1$, the weights increase with income share.

---

[14]These weights can be derived from eq.(5) by replacing income ranks $(r(x_i))$ with income shares $(\tilde{x}_i)$, removing the definite integral — as the width of the interval $[\tilde{x}_{i-1}, \tilde{x}_i]$ is non-constant — and normalising to ensure the weights sum to 1 and are free of small sample bias (Erreygers, Clarke, and Van Ourti, 2012). These weights are non-negative and so the SWF satisfies Pareto, unlike in Erreygers and Kessels (2017). Replacing income ranks with *cumulative* income shares would fix the weight given to the richest individual, which would be inconsistent with taking account of relative income levels.

4 Estimation

We estimate parameters that reflect aversion to pure health inequality (ε or θ) and income-related/caused health inequality (β, β_1 and β_2, or γ) through weights specified by eq.(5), eq.(6), or eq.(11). Within each treatment, the decision problem varies across rounds with the budget (m), multipliers (p_i) and, for Treatments B and C, incomes (x_i) of individuals. We use this variation and a random behavioural model to estimate participant-specific parameters. Orthogonality of the multipliers to the income ranks of individuals allows us to estimate the response of resource allocations to these ranks and so to identify the β parameter and income-dependent weights. Exogenous variation in causal information on differences in multipliers caused by income allows estimation of the response of allocations to this information, which identifies the β_1 and β_2 parameters, and so causality-dependent weights.

Estimation involves maximising the likelihood of observing the resource allocations a participant chooses under the assumption that these allocations are optimal on average but are subject to error. Depending on the specification of the SWF, the optimal resource allocations are given by eq.(3), eq.(8), or eq.(10). The observed resource shares, $\tilde{y}_i = y_i/m$, are assumed to be drawn from the distribution of a random variable, $\tilde{Y}_i$, that equals the optimal resource share in expectation, $E[\tilde{Y}_i] = \tilde{y}_i^*$, where $\tilde{y}_i^* = y_i^*/m$. We assume that the vector of observed resource shares allocated over the three individuals in each round is Dirichlet (1839) distributed with probability density function $f\left(\tilde{y}_1, \tilde{y}_2, \tilde{y}_3; \alpha_1, \alpha_2, \alpha_3\right) = \frac{1}{B(\boldsymbol{\alpha})} \prod_{i=1}^{3} \tilde{y}_i^{\alpha_i-1}$, where $B(\boldsymbol{\alpha}) = \frac{\prod_{i=1}^{3} \Gamma(\alpha_i)}{\Gamma\left(\sum_{i=1}^{3} \alpha_i\right)}$ and $\Gamma()$ is the gamma function (Robson, 2021).[15] The α_i parameters determine the shape of the distribution. Here, they capture the relative weight given to each individual.

---

[15]The Dirichlet distribution is a (flexible) multinomial Beta distribution, bounded between 0 and 1.

Define $\alpha_0 = \sum_{i=1}^{3} \alpha_i$. From the properties of the Dirichlet distribution and the assumption that the actual resource shares are equal to the optimal shares in expectation, we have

$$E[\tilde{Y}_i] = \frac{\alpha_i}{\alpha_0} = \tilde{y}_i^* . \tag{12}$$

We assume that $Var(\tilde{Y}_i) = \frac{\tilde{y}_i^*(1-\tilde{y}_i^*)}{\sigma}$, where $\sigma > 0$ is a precision parameter that reflects noise in the choices that generate the observed allocations. The larger is σ, the lower is the variance of each observed allocation for any given vector of optimal allocations. From the distributional assumption, we have

$$Var(\tilde{Y}_i) = \frac{\alpha_i(\alpha_0 - \alpha_i)}{\alpha_0^2(\alpha_0 + 1)} = \frac{\tilde{y}_i^*(1 - \tilde{y}_i^*)}{\sigma} . \tag{13}$$

It follows that

$$\tilde{y}_i^*(\sigma - 1) = \alpha_i \qquad \forall i . \tag{14}$$

The preference parameters determine the optimal allocation of resource shares to individuals, $\tilde{y}_i^*$, and together with the shape parameters, α_i, these determine the observed resource shares $\tilde{y}_i$. For each participant, k, the estimated parameters are those that maximise the log-likelihood function defined over all the rounds $t \in T$ of a treatment,[16]

$$LL_k = \sum_{t=1}^{T} \log \left(\frac{\Gamma\left(\sum_{i=1}^{3} \alpha_{ik}\right)}{\prod_{i=1}^{3} \Gamma(\alpha_{ik})} \prod_{i=1}^{3} \tilde{y}_{ikt}^{\alpha_{ik}-1} \right) . \tag{15}$$

We also estimate population averaged parameters by pooling the data over all participants and all rounds within a treatment, defining the log-likelihood as the sum of the participant-specific contributions $\left(\sum_{k=1}^{K} LL_k\right)$, and estimating one set of parameters that capture the preferences and weights of a representative SDM.

[16] T is 10 for Treatments A and B, and 6 for Treatment C.

5 Results

5.1 Non-Parametric Analyses

5.1.1 Equity-Efficiency Trade-Off

Using data from Treatment A, Figure 2 plots cumulative density functions (CDFs) of resource shares ($\tilde{y}_i$) and health shares ($\tilde{h}_i$) conditional on relative multipliers ($\tilde{p}_i = p_i/\sum p_i$).[17] When $\tilde{p}_i = 1/3$ for any one individual, the absolute multipliers are equal across the three individuals in a round (see Table A1) and there is no equity-efficiency trade-off. In that case, almost all participants share resources, and therefore health, equally ($\tilde{y}_i = \tilde{h}_i = 1/3$). At other values of $\tilde{p}_i$, the resource share CDF has substantial density on either side of $1/3$. Participants vary allocations in response to between-individual differences in the productivity of resources. The direction and strength of the response varies between participants.

Figure 2: Distributions of resource and health shares by relative multipliers
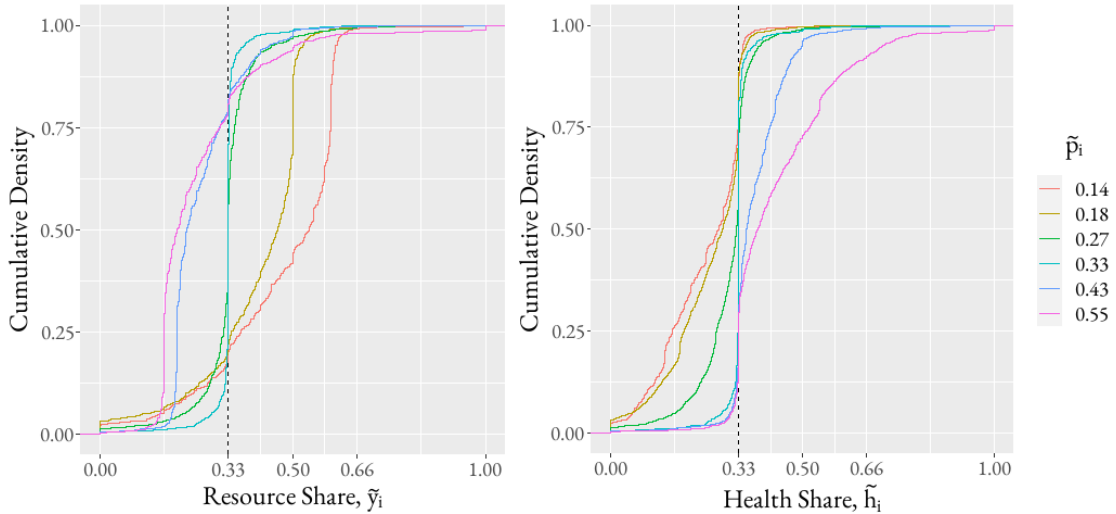
Note: Empirical cumulative density functions of resource shares ($\tilde{y}_i = y_i/\sum y_i$) and health shares ($\tilde{h}_i = h_i/\sum h_i$) to individuals distinguished by relative multipliers ($\tilde{p}_i = p_i/\sum p_i$). Data from all participants and rounds in Treatment A (n=10,110).

---

[17]Since $h_i = p_i \times y_i$, for given ($\tilde{p}_i$), the distribution of ($\tilde{y}_i$) determines distribution of ($\tilde{h}_i$).

Individuals with $\tilde{p}_i > 1/3$ are more productive than average. At less than equal resource shares ($\tilde{y}_i < 1/3$), the CDFs for these individuals are to the left of the CDF for the average productivity individuals. This indicates inefficient allocations that do not maximise health. The above average productivity individuals get less than an equal share of resources in more than three quarters (78.5%) of the allocations to them. In more than one third (34.6%) of allocations, resources are reduced to such an extent that the above average productivity individuals get an approximately equal share of health. In these cases, health maximisation is sacrificed to an extent sufficient to reach equality.

As $\tilde{p}_i$ increases (further) above $1/3$, the health share CDF shifts to the right at above equal shares. In almost two thirds of the respective cases, participants choose allocations that leave individuals who are more productive than average with better than average health. In a substantial minority (17%) of the allocations to individuals who are more productive than average, they are given more resources. In these cases, efficiency is pursued despite the inequality it generates. But not to the extent of giving all resources to the most productive individual and so maximising aggregate health.

For individuals who are less productive than average ($\tilde{p}_i < 1/3$), the resource share CDFs lie to the left of the CDF for average productivity ($\tilde{p}_i = 1/3$) at less than equal shares. This indicates that some participants give less resources to less productive individuals. Very few give no resources to the less productive, which would be required to maximise aggregate health. Most are prepared to trade efficiency for less inequality by giving more resources to less productive individuals. As the relative multiplier falls further below ($1/3$), the resource share CDF shifts further to the right above equal shares. This priorisation of the less productive is, in many cases, insufficient to prevent them from ending up with less than a one-third share of health — the health share CDFs for $\tilde{p}_i < 1/3$ have substantial density below a one-third share.

Pooled data regressions confirm that resource shares fall and health shares rise with increases in the relative multiplier (Appendix D.1). On average, participants compensate for lower productivity by allocating more resources, but not by enough to fully offset the productivity disadvantage. Participant-specific regressions reveal substantial heterogeneity (Appendix D.1). Approximately 14.8% prioritise efficiency by giving more resources, and therefore health, to individuals with higher multipliers. Around 6.2% of participants do not adjust resource allocations in response to the multiplier and so give more health to the more productive. Around a half (49.3%) sacrifice efficiency for less inequality by giving fewer resources to individuals with higher multipliers, while ensuring that these individuals end up with better than average health. Around 27.9% allocate resources to equalise health.[18]

5.1.2 Prioritisation by Income

Using data from all choices in Treatment B, the left panel of Figure 3 shows the mean health shares for individuals ranked by income within each round. On average, the poorest individual receives the largest share (0.3511), while the richest gets the smallest share (0.3141). The null of equal shares is rejected (p-value < 0.01).[19]

The right panel of Figure 3 shows the distribution participant-level estimates of the difference between the health shares given to the poorest and the richest individuals. A little less than a quarter (23.4%) of participants give a significantly larger share of health to the poorest, which appears in the figure as a positive difference. These participants drive the difference in the mean shares in the left panel. A large majority (71.5%) of participants do not discriminate significantly by income. A minority (5.0%) of participants are pro-rich — they give a significantly larger share of health to the richest individual.

---

[18]Additionally, 1.8% of participants give less health to those who have higher multipliers.

[19]This pro-poor result is robust to the specification of the health share-income relationship (Appendix D.2 Table D2.)

24

Figure 3: Health shares by income - pooled (left) and heterogeneous (right)
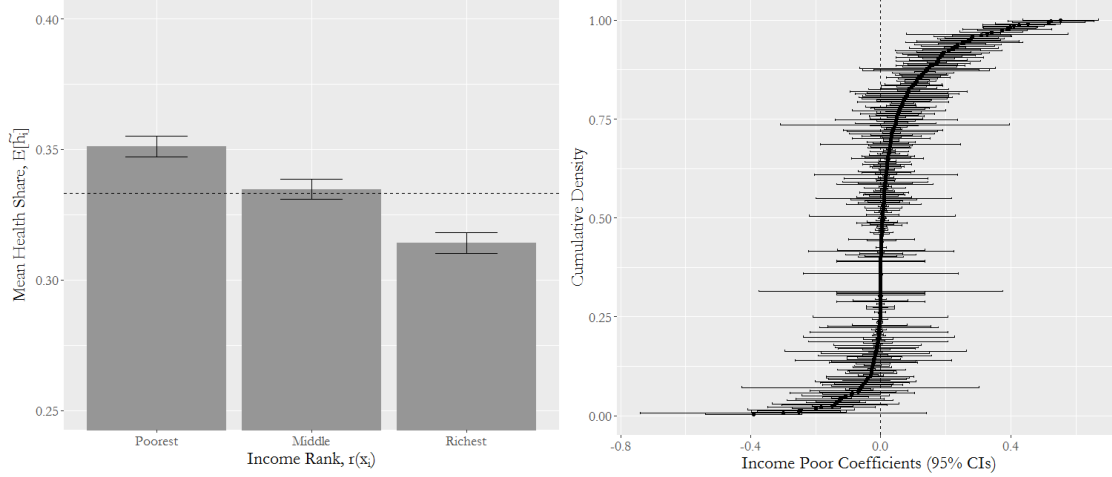

Note: Left panel shows mean of health share ($\tilde{h}_i$) by within round income rank of individual. In this panel, data are pooled and averaged over all participants (n=337) and rounds (10) in Treatment B (n=10,110). Right panel shows distribution of participant-specific regression estimates of the health share difference between the poorest and richest individuals within each round. In each panel, interval lines show 95% confidence intervals

5.1.3 Sensitivity to Causality

Using the data from Treatment C, Figure 4 shows CDFs of the health share to the poorest individual within each round stratified by values of the exogenously varying causal information and whether the poorest individual is the least (left) or most (right) productive.

With each panel, the lack of any substantial differences between the CDFs indicates that participants generally do no adjust allocations, and so health shares, in response to information about the causal effect of income on the productivity of resources. On average across all participants, increasing the percentage of the multiplier differences that participants are told is caused by income from 0% to 100% has no impact whatsoever on the mean health share to the poorest when that individual is the least productive and it raises the poorest individual's health share by 1.65% when it is the most productive (Appendix D.3 Table D3). Around 45% of participants do not change the health share given to the poorest at all when the causal information is increased from 0% to 100% (Appendix D.3 Figure D2).

Figure 4: Distribution of health share to poorest by causal information
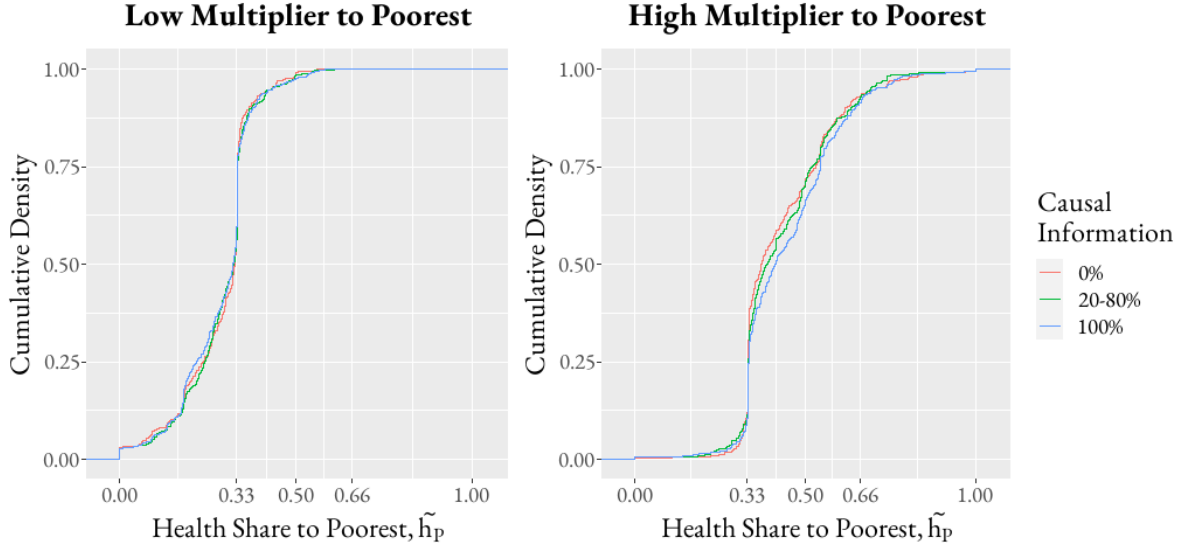
Note: CDFs of health share to poorest individual within each round ($\tilde{h}_p$, $p = i$ with $x_i < x_j \forall j \neq i$) stratified by exogenously varying values of causal information on percentage of multiplier differences caused by income differences (section 2.2.3). In left (right) panel, poorest individual has smallest (largest) multiplier. n=1011 in each panel.

5.2 Pooled Estimates of Preference Parameters

Table 2 shows parameter estimates obtained from data that are pooled over allocations made by all participants in all rounds within each treatment. Using data from Treatment A only and imposing constant weights, we obtain $\hat{\varepsilon} = 1.391$ with a 95% confidence interval well above 1. Hence, the representative SDM is willing to sacrifice health maximisation for less inequality.[20] The point estimates of ε are marginally and significantly larger when weights are allowed to depend on income (Treatment B) and its causal effect (Treatment C).

The Treatment B estimate of β is significantly larger than 1, implying that, on average, participants put greater weight on the health of poorer individuals. The magnitude of the estimate implies that the representative SDM would give the poorest individual in a popula-

---

[20]The welfare loss generated by inequality is the difference between the mean and EDE health. For example, for three individuals with QALYs of 40, 60, and 80, the mean is 60, the EDE at $\varepsilon = 1.391$ is 56.78, and the welfare loss is 3.22 QALYs.

Table 2: Pooled estimates of social welfare function parameters

Treatment	Health Inequality Aversion	Income Weight	Causal Income Weight	Precision
	ε	β (β_1 in C)	β_2	σ
A	1.391			8.047
	[1.323 - 1.470]			[7.354 - 8.910]
B	1.563	1.105		8.677
	[1.474 - 1.663]	[1.075 - 1.136]		[7.938 - 9.599]
C	1.670	1.102	0.956	9.044
	[1.477-1.793]	[1.060 - 1.147]	[0.892 - 1.048]	[7.996 - 10.189]

Note: Estimates in each row are obtained from pooling data over all participants (n=337) and rounds within the respective treatment. Estimates maximise the total log-likelihood, $\sum_k LL_k$, with LL_k defined in eq.(15) and the parameters defined in eq.(2) for ε, eq.(5) for β, and eq.(6) for β_1 and β_2. The total log-likelihoods are 3308.2, 3626.5, and 2231.3 for treatments A, B, and C, respectively. Number of observations is 10,110, 10,110, and 6,066 in treatments A, B, and C, respectively. In brackets are 95% bootstrap confidence intervals obtained with the percentile method.

tion of three a weight that is 21.5% larger than the weight given to the richest individual.[21]

This indicates substantial preference for a pro-poor distribution of health and implies greater aversion to health inequality when health differences are positively associated with income differences. The same inferences can be made from the fact that the Treatment C point estimate of β_1 is also significantly greater than 1.

Treatment B estimates of the general model consisting of eq.(2) and eq.(5) imply rejection of both an Atkinson SWF that does not allow aversion to income-related health inequality ($H_0 : \beta = 1$) and Wagstaff's (2002) achievement index that does not allow aversion to pure health inequality ($H_0 : \varepsilon = 0$). Allowing the latter type of aversion gives the greater improvement in data fit.[22]

Using Treatment C data, the point estimate of β_2 is very close to 1 and is not significantly different from this value. This implies that, on average, the weight given to poorer individuals

---

[21]With β=1.105, $\boldsymbol{\omega}^{\boldsymbol{B}}(\beta) = [0.361, 0.341, 0.297]$ for the poorest, middle, and richest individuals, respectively.

[22]With $\varepsilon = 0$ imposed, we get $\hat{\beta} = 3.22$ and a mean squared error (MSE) of 0.688, which is substantially larger than the MSE of the general model (0.040) and the MSE of the Atkinson SWF with $\beta = 1$ (0.041).

is not dependent on whether the health-income association is causal. The representative SDM is not more averse to income-related health inequality when low income causes poor health.[23]

5.3 Heterogeneous Estimates of Preference Parameters

Table 3 shows percentiles of the distribution of participant-specific estimates of each preference parameter.[24] It also shows estimates of the precision parameter, which are much larger than the respective pooled estimates, indicating that the participant-specific estimates are substantially more precise. This signals the importance of preference heterogeneity, which is evident for all parameters and is particularly marked for pure health inequality aversion.

The median estimates of ε imply greater aversion to pure health inequality than the respective pooled estimates.[25] Irrespective of the treatment data used, the median estimate is well above 1, indicating that a majority is substantially averse to inequality. The 10-90 percentile ranges and Figure 5, which plots the distribution of $\hat{\varepsilon}$ from Treatment A, show extensive heterogeneity. While there are no health maximisers ($\hat{\varepsilon} = 0$), around one sixth (16.0%) of Treatment A participants have $0 < \hat{\varepsilon} \leq 0.9$. These Efficiency Seekers have only weak aversion to pure health inequality. About a tenth (9.8%) have approximately Cobb-Douglas preferences (Dolan, 1998): $0.9 < \hat{\varepsilon} < 1.1$. More than two fifths (42.7%) display

---

[23]To identify β_2, we use the randomly generated casual information to fix λ in eq.(6). We confirm robustness to an alternative identification strategy that fixes λ with elicited beliefs. We asked each participant to express, on a [-100, 100] scale, the strength of their belief that income causally raises or lowers the multiplier (Appendix A.4 and Appendix G). After re-scaling to [-1,1], this provides an estimate of λ for each participant that can be used with the Treatment B data to identify β_2. This gives a pooled estimate of $\hat{\beta}_2 = 1.001$ [0.952-1.050]. The respective estimates of ε and β_1 are almost identical to the Treatment B estimates.

[24]See Appendix D.5 for distributions of estimates and correlations between estimates: both within and between treatments.

[25]For example, using data from Treatment A and the 3-person scenario from fn. 20, the welfare loss from inequality calculated with the median participant's $\hat{\varepsilon}$ is 7.12, which is more than twice the loss obtained using the pooled estimate of ε (3.22).

Table 3: Heterogeneous estimates of social welfare function parameters

Treatment	Percentile	Health Inequality Aversion ε	Income Weight β (β_1 in C)	Causal Income Weight β_2	Precision σ
A	50th	3.170			171.58
	10th, 90th	0.731, 104.20			13.069, 172.61
B	50th	3.498	1.047		150.00
	10th, 90th	0.954, 45.186	0.798, 2.482		19.203, 172.61
C	50th	4.413	1.056	0.999	150.00
	10th, 90th	0.932, 81.785	0.561, 3.363	0.283, 1.416	36.397, 172.61

Note: Estimates maximise participant-specific log-likelihoods, eq.(15), with parameters defined in eq.(2) for ε, eq.(5) for β, and eq.(6) for β_1 and β_2. Within each panel (A, B, C), top row gives median estimates and next row gives 10th and 90th percentiles in the respective distribution of participant-specific estimates. Each distribution has 337 estimates — one for each participant. In Treatments A and B, each estimate is obtained from 30 data points (10 rounds × 3 allocations per round). In Treatment C, each estimate is from 18 data points (6 rounds × 3 allocations per round).

more strongly Prioritarian preferences (Parfit, 2000): $1.1 \leq \hat{\varepsilon} < 15$. A little less than a third (31.5%) exhibit preferences that approach Maximin: $\hat{\varepsilon} \geq 15$.[26]

The median estimates given in Table 3 of the income weight parameters that reflect aversion to income-related inequality (β in Treatment B and β_1 in Treatment C) are slightly smaller than the respective pooled estimates (Table 2). This indicates that the median participant is a little less pro-poor than the representative SDM captured by the pooled estimates. The 10-90 percentile ranges and the left panel of Figure 6, which plots the distribution of β estimates obtained from Treatment B data, again show preference heterogeneity. Over a quarter of participants (27.0%) are approximately income neutral: $0.95 \leq \hat{\beta} \leq 1.05$. They do not discriminate strongly in favour of either poorer or richer individuals. Around half (49.3%) are clearly pro-poor: $\hat{\beta} > 1.05$. Just under a quarter (23.7%) are clearly pro-rich: $\hat{\beta} < 0.95$.

---

[26]The 44 participants who performed poorly on the tutorial questions and are excluded from the analysis sample do not display significantly different health inequality aversion. The estimates obtained for them have lower precision and worse goodness-of-fit (Appendix D.4).

Figure 5: Distribution of pure health inequality aversion estimates, $\hat{\varepsilon}$
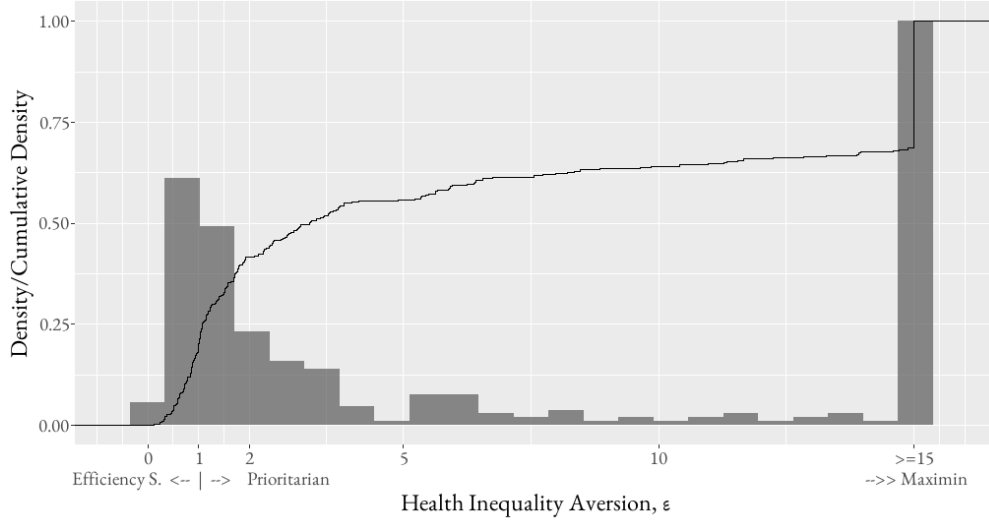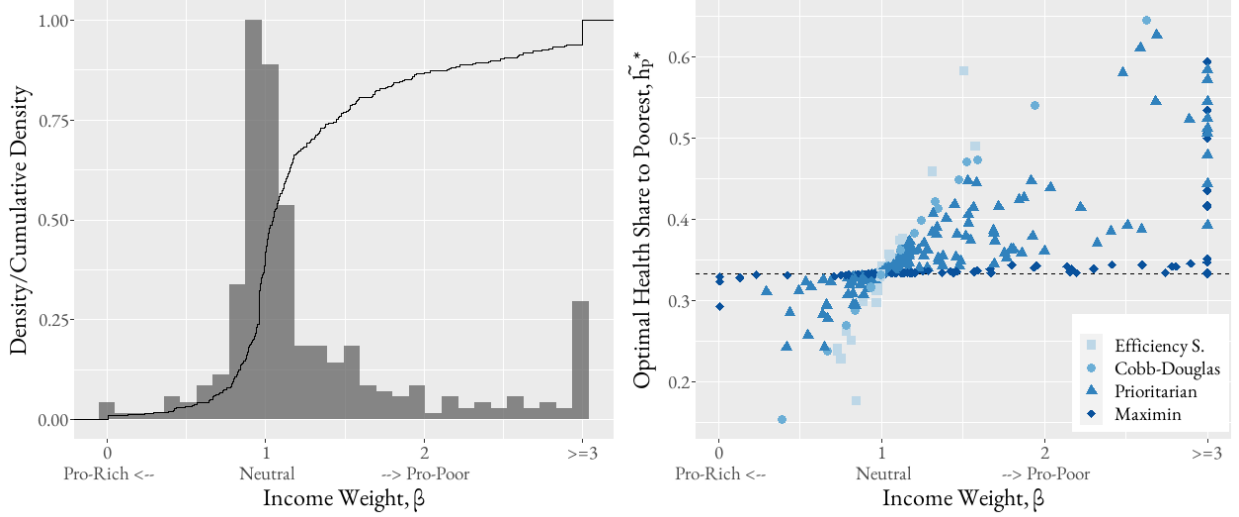

Notes: Distribution of participant-specific estimates of ε obtained from Treatment A data. There are 337 estimates. 30 data points used to obtain each estimate. Distribution is shown as both a histogram, with density normalised to 1, and an empirical cumulative density plot. Values of $\hat{\varepsilon} > 15$ censored at 15.

For a majority of the sample, estimates of $\varepsilon > 0$ and $\beta > 1$ imply rejection of restrictions on the general model (eq.(2) and eq.(5)) that would give an Atkinson SWF with no aversion to income-related health inequality ($\beta = 1$) and an achievement index with no aversion to health inequality ($\varepsilon = 0$).

Prioritisation of the poor, or the rich, is constrained by aversion to pure health inequality. When this aversion is stronger, an increase in pro-poor weights has a smaller impact on the optimal allocation of health to the poor. This is illustrated in the right panel of Figure 6, which uses Treatment B estimates to plot participants' optimal health shares to the poorest individual ($\tilde{h}_p^*$) against their $\hat{\beta}$. Symbols and colours distinguish between participant categories defined by ($\hat{\varepsilon}$) as Efficiency Seeking, Cobb-Douglas, Prioritarian, and Maximin. We consider a case with constant multipliers to ensure that there is no equity-efficiency trade-off. Hence, when $\hat{\beta} \approx 1$, $\tilde{h}_p^* \approx 1/3$. That is, with income-neutral weights, the poorest individual optimally gets an equal share of health. As $\hat{\beta}$ increases above 1, $\tilde{h}_p^*$ increases above 1/3, but

Figure 6: Distribution of income weight parameter estimates, $\hat{\beta}$

Notes: The left panel shows the distribution of participant-specific maximum likelihood estimates of β (eq.(5)) obtained from Treatment B data. There are 337 estimates, with 30 data points used to obtain each estimate. The distribution is shown as both a histogram, with density normalised to 1, and an empirical cumulative density plot. Values of $\hat{\beta} > 3$ censored at 3. Pro-poor is $\hat{\beta} < 1$. Pro-rich is $\hat{\beta} > 1$. The right panel plots the optimal health share to the poorest (assuming equal multipliers) against ($\hat{\beta}$), with symbols for categories of health inequality aversion. Efficiency Seeking, Cobb-Douglas, Prioritarian, and Maximin correspond to $\hat{\varepsilon} \leq 0.9$, $0.9 < \hat{\varepsilon} < 1.1$, $1.1 \leq \hat{\varepsilon} < 15$, and $\hat{\varepsilon} \geq 15$, respectively.

clearly to a much greater extent for the Efficiency Seekers and Cobb-Douglas types. For the Maximin types, the optimal share hardly moves from 1/3 because their extreme aversion to pure health inequality constrains them from giving more health to the poor even when they are strongly inclined toward the poor. Efficiency Seekers are relatively unconcerned by pure health inequality, and since with equal multipliers there is no efficiency motivation to consider, more pro-poor income weights (higher $\hat{\beta}$) strongly increase the optimal allocation to the poor.

In Table 3, the median estimate of the causal income weight parameter β_2 is even closer to 1 than the pooled estimate in Table 2. Both estimates, consistent with the non-parametric analysis, indicate that, on average, aversion to income-related health inequality does not strengthen when low income is known to cause poor health. However, the 10-90 percentile

range for $\hat{\beta}_2$ implies that there are participants who give causality-dependent income weights. For a little less than a third (30.6%), $\hat{\beta}_2 > 1.05$ (Appendix D.5, Figure D3). These participants would appear to place greater weight on the health of poorer individuals after learning that lower income causes worse health (Table C1). Such causal information would seem to lead almost two fifths (38.9%) of participants with $\hat{\beta}_2 < 0.95$ to revise the weights in the opposite direction. Against such interpretations based on point estimates, a restricted model that imposes $\beta_2 = 1$ fits the Treatment C data better than the unrestricted model using either pooled or heterogeneous estimates (Appendix E.2).

Participant-specific estimates of each of ε and σ are strongly and significantly correlated between the treatments (Appendix D.5 Table D5). This demonstration of within participant consistency across the three treatments lends face validity to the analysis. Further evidence of consistency is a strong, positive correlation between estimates of β from Treatment A and β_1 from Treatment C (Table D5). Interestingly, estimates of ε from Treatment A are positively and significantly correlated with estimates of β from Treatment B, while within Treatment B the correlation between the estimates of these two parameters is weaker and not significant. This supports the contention that there is confounding of the two types of health inequality aversion when they are not estimated simultaneously. Within each treatment, there is a positive correlation between ε and σ, which is partly due to low noise in the allocations of maximin types who always opt for an equal distribution of health.

5.4   Model Comparisons

Table 4 shows pooled and heterogeneous parameter estimates and goodness-of-fit (GOF) statistics for our main specification (Atkinson SWF, eq.(2) and income-rank-dependent weights, eq.(5)) and for two alternative SWFs (Kolm, eq.(9) and non-consequentialist, eq.(7))

and weighting functions (income-share-dependent, eq.(11) and constant, eq.(4)). We obtain all estimates with Treatment B data and show medians of the heterogeneous estimates.

Table 4: Parameter estimates and goodness of fit for alternative models

		Parameters			Goodness-of-Fit		
		Inequality Aversion ε or θ	Income Weight β	Precision σ	MPL	MSE	AIC
Pooled	*Main*	1.563	1.105	8.677	0.687	0.040	-7662.1
	SWF						
	Kolm	-0.035	1.102	8.671	0.684	0.041	-7584.2
	Non-Consq.	0.270	1.017	7.812	0.631	0.057	-6881.6
	Weights						
	Share	1.560	1.154	8.673	0.687	0.040	-7659.0
	Constant	1.560	N/A	8.576	0.683	0.041	-7577.3
Heterogeneous	*Main*	3.498	1.047	150.0	0.909	0.011	-23974.4
	SWF						
	Kolm	-0.079	1.063	146.6	0.896	0.016	-22801.6
	Non-Consq.	4.066	1.001	14.23	0.727	0.051	-10789.7
	Weights						
	Share	3.451	1.067	149.2	0.906	0.012	-23772.5
	Constant	3.422	N/A	117.7	0.876	0.018	-22782.0

Note: Estimates from Treatment B data. Medians of heterogeneous parameters estimates shown. *Main* is Atkinson SWF (eq.2) and income-rank-dependent weights (eq.5). Inequality aversion is ε in eq.(2) and eq.(7) for Main and Non-consequentialist, respectively, and θ in eq.(9) for Kolm. Income weight is β in eq.(5) for Main, Kolm, and Non-consequentialist, and is γ in eq.(11) for Share. Constant is eq.(4) weight function. MPL is what we call the Mean Proportional Likelihood. Define $PL_t = L_t/(L_t + L_t^U)$, where L_t is the likelihood in round t for the data and estimates and L_t^U is a likelihood for a uniform distribution draw. $MPL = 1/T \sum_t^T (PL_t)$. If $MPL = 0.5$, model fit to data is no better than the fit to uniform distribution draws. As $MPL \to 1$, data fit improves. $MSE = 1/T \sum_t^T (\tilde{y}_i - \tilde{y}_i^*)^2$ is the Mean Square Error. $AIC = 2k - 2\ln(L)$ is Akaike Information Criterion, with k the number of parameters. GOF is increasing with MPL and decreasing with MSE and AIC.

The heterogeneous estimates give a much better fit to the data than the pooled estimates irrespective of the GOF measure and the specification (see also Appendix E). For both pooled and heterogeneous estimates, all GOF measures indicate that the Atkinson SWF is strongly preferred to the non-consequentialist SWF and is slightly preferred to the Kolm SWF (see

33

also Appendix E). The estimates, particularly the pooled ones, indicate that aversion to resource inequality, which is captured by the estimate of ε with a non-consequentialist SWF, is weaker than the main specification estimate of health inequality aversion. Using the pooled estimates, we reject the null of no aversion to pure health inequality for both the Atkinson SWF ($\varepsilon = 0$) and the Kolm SWF ($\theta = 0$).[27]

Specification of the income weights as rank-dependent versus share-dependent does not affect the GOF as much as the specification of the SWF. However, allowing for some form of income-related weights improves the GOF compared with the restricted model that imposes constant weights (see also Appendix E). Specification of the SWF as Atkinson versus Kolm has little effect on the pooled and median estimates of the income-dependent weight parameter. In both cases, using the pooled estimates, we reject the null of constant weights ($\beta = 1$) in favour of weights that decrease with rising income ($\beta > 1$).[28]

In sum, the evidence supports health consequentialism, pure health inequality aversion, and pro-poor income-dependent weights. The data are not definitively more consistent with aversion to relative or absolute health inequality and the evidence is not decisively in favour of rank- or share-dependent income weights. Our main specification of an Atkinson SWF with rank-dependent income weights fits the data at least as well as all others considered.

6 Illustrative Application

Our estimates of heterogeneous SWF parameters obtained from a UK representative sample can be used in policy evaluation. They can be applied to an estimated policy-specific distri-

---

[27]Bootstrapped 95% confidence intervals are $[1.474 - 1.663]$ for ε and $[-0.038 - -0.033]$ for θ. The magnitudes of ε and θ are not comparable.

[28]Bootstrapped 95% confidence intervals for β are $[1.075 - 1.136]$ and $[1.072 - 1.134]$ with the Atkinson and Kolm SWFs, respectively. We also reject the null of constant weights ($\gamma = 1$) against pro-poor income-share-dependent weights based on a bootstrapped 95% CI for $\gamma = 1$ of $[1.113 - 1.195]$.

bution of health over individuals or groups, which may ordered by income, to simulate the distribution of support for the respective policy.[29]

To illustrate this potential, we simulate a population of 100,000 individuals characterised by income and health. The annual income of each individual is a random draw from a (rescaled) beta distribution with the mean (£34,281) and standard deviation (£30,052) set to be broadly consistent with the respective values for the UK income distribution. Health (QALYs) is a positive and stochastic function of log income.[30] We evaluate two policies that have the same cost and do not change the distribution of income. Policy A produces more health for poorer individuals and so results in less pure health inequality and less income-related health inequality than Policy B, which gives a higher mean level of health.[31]

The top panel of Table 5 shows, for each policy, the mean and standard deviation of health and its correlation with income.[32] Policy B is preferred by standard economic evaluation that only considers the impact on mean health.

The bottom panel shows EDE health using our estimates of the Atkinson SWF without and with income weights. In the first case, Policy B is still preferred if we use the pooled estimate. The health inequality aversion of the representative SDM is insufficient for the greater inequality generated by Policy B to outweigh the higher mean it achieves and so tilt the balance in favour of Policy A. However, the median of the heterogeneous EDE estimates is larger with A. For more than half (51.3%) of the representative sample, Policy A gives the larger EDE health and so this policy would be chosen under simple majority voting.

---

[29]Appendix F explains how to access and use our estimates, at https://doi.org/10.17632/9vy6f6g5k3.1. If grouped (by income) data are used, then the weighting parameters within the SWF would be applied to the proportion of the population in each group.

[30]We use $x_i \sim Beta(1.1, 10.5) \times 360000$ to generate the distribution of income and derive from it a baseline distribution of health by setting $h_i = 45 + 2log(x_i) + v_i$,, where $v_i \sim N(0,3)$.

[31]For Policy A, $h_i^A = h_i + 0.75log(x_i)(1 - r(x_i))$, where $r(x_i) \in [0,1]$ is income rank that is increasing with income. For Policy B, $h_i^B = h_i + 0.75log(x_i)r(x_i)$. Policy A allocates proportionately less resources to individuals with higher income rank, while Policy B allocates proportionately more. The productivity of these resources in determining health is a positive function of log income.

[32]Appendix F Figure F1 shows the simulated marginal and joint distributions of income and health for each policy.

Table 5: Hypothetical policy evaluation

	Policy		Preferred	Majority
	A	B		
Health				
Mean	68.46	68.93	B	
Std. Dev.	3.37	5.64		
Income Correlation	0.08	0.76		
EDE Health				
Without Income Weights				
Pooled	68.35	68.60	B	
Heterogeneous	68.21	68.17	A	51.3%
With Income Weights				
Pooled	68.31	68.16	A	
Heterogeneous	68.03	67.14	A	70.9%

Note: See footnotes 30 and 31 for details of policy simulations. EDE = equally distributed equivalent. Median EDE are given for heterogeneous estimates. **Preferred** indicates the policy that generates the largest EDE for the representative SDM (Pooled) or median voter (Heterogeneous). **Majority** gives the percentage of the representative sample with a larger EDE from the **Preferred** policy. Under *Without Income Weights*, EDE are calculated using estimates of ε obtained from Treatment A data. Under *With Income Weights*, EDE are calculated using estimates of ε and β obtained from Treatment B data.

The second scenario presented in the bottom panel allows for aversion to income-related health inequality through income-rank-dependent weights. Using the pooled estimates of ε and β from Treatment B, we infer that the representative SDM would prefer Policy A. Adding aversion to positive health-income correlation to even moderate pure health inequality aversion is sufficient to tilt the balance in favour of A for the representative SDM, despite the higher mean generated by B. Using the heterogeneous estimates, preference for Policy A is even more emphatic. It would be the choice of 70.9% of the representative sample.

To take account of policy impacts on pure health inequality, the pooled or median estimate of ε is all that is needed to add distributional sensitivity to standard economic evaluation. As the above example demonstrates, the consequence of this extension for the choice of policy can depend on whether a pooled or median estimate is used. Our approach allows examination of variation in support for a policy along the distribution of estimates.

When attention is paid to income-related health inequality aversion, the pooled estimates of ε and β remain sufficient to rank any set of health outcome distributions generated by alternative policies provided the preferences of a representative SDM are considered relevant. When opting to use heterogeneous estimates, the medians of two parameters are not enough. In that case, the analyst must use the entire joint distribution of $\hat{\varepsilon}$ and $\hat{\beta}$ that we provide.

7 Discussion

Standard economic evaluation of healthcare pursues an objective — health maximisation — that is inconsistent with the social preferences we elicit from a representative sample of the UK population. On average, people are willing to sacrifice efficiency in health production for less inequality. They also prioritise the health of poorer individuals.

There is substantial heterogeneity in social preferences over the distribution of health. A pooled estimate understates the extent to which most people would sacrifice maximisation of aggregate health to reduce inequality. Our median estimate of aversion to pure health inequality ($\hat{\varepsilon} = 3.5$), which is estimated simultaneously with income weights, is smaller than previous UK estimates that potentially confound this aversion with aversion to income-related health inequality (Dolan and Tsuchiya, 2011; Robson et al., 2017; McNamara et al., 2020). Our median estimate is larger than the median interval estimate ($\hat{\varepsilon} = 1.0-1.5$) obtained from a representative sample in Ontario (Hurley, Mentzakis, and Walli-Attaei, 2020), although $\hat{\varepsilon} > 3$ for 48% of that sample. Our median estimate is also within the range of median estimates ($2.24 < \hat{\varepsilon} < 4.85$) identified from a sample of Portuguese college students (Pinho and Botelho, 2018).

In addition to aversion to pure health inequality, we find that, on average, there is prioritisation of the health of poorer individuals. Consequently, aversion to pro-rich health inequality is greater than aversion to pure health inequality. However, both pooled and

median estimates indicate only slightly larger weights on the health of poorer individuals (β slightly above 1). The weights are less pro-poor than those imposed by the standard concentration index measure of income-related health inequality ($\beta = 2$) (Wagstaff, Paci, and Van Doorslaer, 1991; O'Donnell et al., 2008). This appears somewhat inconsistent with the Ontario study that finds a median degree of aversion to income-related health inequality closer to that built into the concentration index ($1.5 < \hat{\beta} < 2$) (Hurley, Mentzakis, and Walli-Attaei, 2020). Another study finds that, if anything, the degree of aversion implicit in the concentration index understates that of the median person in Sweden ($2 < \hat{\beta} < 3$) (Hardardottir, Gerdtham, and Wengström, 2021). The discrepancy between our estimates and these others is consistent with our hypothesis that studies that impose a positive correlation between health and income and do not elicit income weights simultaneously with aversion to pure health inequality will obtain upwardly biased estimates of willingness to prioritise the health of poorer individuals. In these studies, elicited aversion to differences in health by income also reflects aversion to differences in health per se.

In our approach, income weights have less impact on the allocation of health resources when there is stronger aversion to pure health inequality. A social decision maker who is less tolerant of that inequality allocates more resources to the less healthy. Indirectly, this increases the allocation to poorer individuals when health and income are positively correlated, as typically they are. This reduces the need for and marginal effect of pro-poor weights. Effectively, aversion to pure health inequality substitutes for the weights in raising the socially preferred health of poorer individuals. This explains the discrepancy between our and other estimates of the income-weight parameter.

The income-weight and pure health inequality aversion parameters jointly determine aversion to income-related health inequality. To illustrate, consider the marginal rate of substitution (MRS) of a poor individual's health (h_P) for a rich individual's health (h_R) with social welfare given by eq.(2): $\frac{\partial h_{EDE}/\partial h_P}{\partial h_{EDE}/\partial h_R} = \frac{\omega_P}{\omega_R}\left(\frac{h_P}{h_R}\right)^{-\varepsilon}$. The relative amount of QALYs

a rich individual must gain in order to offset a reduction in the QALYs of a poor individual, such that social welfare is constant, depends not only on the relative income weights, $\frac{\omega_P}{\omega_R}$, and so the parameter β in eq.(5), but also on the relative health inequality, $\frac{h_P}{h_R}$, and the pure health inequality aversion parameter, ε.

Table 6 shows the MRS for a two-person society and for configurations of social preferences and five distributions of health (QALYs). The top row shows the preferences of a health maximiser with no aversion to pure or income-related health inequality. In that case, the health of rich and poor individuals are always perfect 1:1 substitutes. The second row corresponds to the case in which there is no aversion to pure health inequality ($\varepsilon = 0$) and so aversion to income-related health inequality is entirely determined by the income weight parameter, which we set to the value imposed in the standard concentration and achievement indices ($\beta = 2$). In this case, the rich individual must *always* gain 3 times the number of QALYs to compensate for the poor individual's loss of QALYs, irrespective of their levels of health.

Table 6: Marginal rates of substitution between health of poor (h_P) and rich (h_R) individuals

Inequality aversion			Health distribution, (h_P, h_R)				
Type	Pure (ε)	Income (β)	(62, 74)	(70, 74)	(74, 74)	(78, 74)	(86 ,74)
None	0	1	1	1	1	1	1
Income only	0	2	3	3	3	3	3
Pure only	$\hat{\varepsilon}_{0.5}$	1	1.75	1.19	1	0.85	0.62
Both	$\hat{\varepsilon}$	$\hat{\beta}$	2.40	1.42	1.02	0.87	0.67

Note: Two person society with health measured in QALYs. Marginal rate of substitution (MRS) calculated for various social preferences and health distributions from the equation given in text. "None" shows the MRS of a health maximiser. "Income only" shows MRS for social welfare given by eq.(2), with $\varepsilon = 0$, and income weights from eq.(5), with $\beta = 2$. "Pure only" shows MRS derived from eq.(2) with ε set to the median participant-specific estimate from Treatment A, $\hat{\varepsilon}_{0.5} = 3.2$, and constant weights, eq.(4). "Both" gives the sample median MRS derived from eq.(2) and eq.(5) using participant-specific estimates $\hat{\varepsilon}$ and $\hat{\beta}$ from Treatment B.

In the third row, there is no explicit prioritisation of the poor person's health ($\beta = 1$). Aversion to income-related health inequality arises indirectly through aversion to pure health inequality (the median participant-specific estimate from Treatment A, $\hat{\varepsilon}_{0.5}$) and any association between health and income. Even at the most extreme pro-rich health inequality considered, the MRS is less than three fifths of that implied by the achievement index scenario ($\varepsilon = 0$, $\beta = 2$). As pro-rich health inequality falls in magnitude and then turns to pro-poor inequality, the MRS diverges further from that of the achievement index case.

In the bottom row, aversion to income-related health inequality arises directly through non-constant income weights and aversion to pure health inequality. For this case, we show the median MRS obtained from the distributions of $\hat{\varepsilon}$ and $\hat{\beta}$ estimated from Treatment B. At the most extreme pro-rich inequality, the MRS is very close to that implied by the achievement index scenario. This illustrates that despite our median estimate of the income weight parameter (β) being smaller than respective estimates obtained by others (Hardardottir, Gerdtham, and Wengström, 2021; Hurley, Mentzakis, and Walli-Attaei, 2020), after taking all parameters into account, our (median) estimates do not necessarily imply less aversion to income-related health inequality. The limitation of capturing this aversion through a single parameter, as is done with the concentration/achievement indices, is apparent from comparisons of the MRS as health inequality becomes less pro-rich. When there is no inequality, we estimate that the median person would require 1.02 times the QALY gain to a rich individual to offset a poor individual's loss of QALYs. In the achievement index scenario, the rich individual would still have to gain 3 times the QALYs lost. Such linear preferences are implausible. Further, in this scenario, the MRS between the health of the poorest and richest individuals increases with population size and becomes preposterously large even for small real-world populations. This does not happen with the two-parameter approach.

There is heterogeneity in the prioritisation of health by income. Our non-parametric and parametric analyses suggest that a quarter to a half of the UK population is pro-poor, while

somewhere between a twentieth and a little less than a quarter is pro-rich. The preferences of the latter group are entirely inconsistent with the normative foundation of concentration and achievement indices (Wagstaff, 2002; Bleichrodt and van Doorslaer, 2006; Erreygers, Clarke, and Van Ourti, 2012). We are not the first to estimate that a sizeable proportion of a population would prioritise the health of richer people. Hardardottir, Gerdtham, and Wengström (2021) find that slightly more than one quarter of a representative Swedish sample displays a pro-rich bias, while Hurley, Mentzakis, and Walli-Attaei (2020) estimate that a little less than one fifth of Ontarians are pro rich. These preferences are consistent with the marginal utility of health increasing with income, which, in turn, is implied by positive dependence of the marginal utility of income (consumption) on health. There is some empirical support for the latter (Finkelstein, Luttmer, and Notowidigdo, 2013), although the evidence is mixed (De Nardi, French, and Jones, 2010). Some may choose to allocate more health resources to richer individuals because higher income is perceived to offer greater opportunity to get the most from good health. This would be consistent with maximisation of aggregate well-being defined over health and income, with positive interaction between these two arguments. Our set up does not allow for such interdependence.

Another limitation is that to keep the experiment task cognitively feasible for a general population sample, we used a linear health production function. This sharpens the trade-off between efficiency and equity. It may also increase the estimated aversion to health inequality. Without diminishing marginal product, maximisation of aggregate health requires the allocation of all resources to the most productive individual, which would also maximise health inequality. If there were diminishing returns to health resources, less aggregate health would need to be sacrificed to satisfy preference for lower inequality. The design of an experiment that allows diminishing returns and yet remains feasible remains a challenge.

Our experimental manipulation of income-health causality did not deliver clear evidence that causality strengthens aversion to income-related health inequality. However, prioritisa-

tion of the health of poorer individuals is associated with beliefs about causality. Additional analysis reported in Appendix G shows that participants tend to give the poorest individual a larger share of health when they believe that a larger fraction of that individual's low potential health is caused by low income. This is merely descriptive evidence because the beliefs, unlike the causal information given in Treatment C, were not randomly assigned. Nonetheless, it is consistent with beliefs about causality conditioning aversion to income-related health inequality. The lack of strong support for this hypothesis from Treatment C could possibly be because participants perceive distributive justice through the lens of equality of opportunity (Roemer, 2002) and view high income, and any health advantage it bestows, as a just reward for effort exerted to increase income. Another possible explanation is that aversion to income-related health inequality arises from concern about deprivation in multiple dimensions of well-being irrespective of whether one dimension (income) has a causal effect on another (health).

8 Conclusion

Our novel experiment and estimation strategy make it possible to disentangle aversion to pure health inequality from aversion to income-related health inequality. The approach could be used to estimate aversion to health inequality related to any non-health characteristic. Our findings cast doubt on the normative principles that underpin standard practice in health economic evaluation and the measurement of income-related health inequality. Strengthening of the normative foundations of these health economics methods needs to take account of the substantial heterogeneity we reveal in social preferences over the distribution of health. This is feasible with the distributions of estimated social preference parameters we provide.

References

Abasolo, I. & Tsuchiya, A. (2004). Exploring social welfare functions and violation of monotonicity: An example from inequalities in health. *Journal of Health Economics*, *23*(2), 313–329. doi:10.1016/j.jhealeco.2003.08.003

Abellan-Perpiñan, J. & Pinto-Prades, J. (1999). Health state after treatment: A reason for discrimination? *Health Economics*, *8*(8), 701–707. doi:10 . 1002 / (SICI) 1099 - 1050(199912)8:8%3C701::AID-HEC473%3E3.0.CO;2-M

Arrow, K. J. (1971). A utilitarian approach to the concept of equality in public expenditures. *The Quarterly Journal of Economics*, *85*(3), 409–415. doi:10.2307/1885930

Atkinson, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory*, *2*(3), 244–263. doi:10.1016/0022-0531(70)90039-6

Bleichrodt, H. (1997). Health utility indices and equity considerations. *Journal of Health Economics*, *16*(1), 65–91. doi:10.1016/S0167-6296(96)00508-5

Bleichrodt, H., Diecidue, E., & Quiggin, J. (2004). Equity weights in the allocation of health care: The rank-dependent qaly model. *Journal of Health Economics*, *23*(1), 157–171. doi:10.1016/j.jhealeco.2003.08.002

Bleichrodt, H., Doctor, J., & Stolk, E. (2005). A nonparametric elicitation of the equity-efficiency trade-off in cost-utility analysis. *Journal of Health Economics*, *24*(4), 655–678. doi:10.1016/j.jhealeco.2004.10.001

Bleichrodt, H. & van Doorslaer, E. (2006). A welfare economics foundation for health inequality measurement. *Journal of Health Economics*, *25*(5), 945–957. doi:10.1016/j.jhealeco.2006.01.002

Cappelen, A. W., Hole, A. D., Sørensen, E. Ø., & Tungodden, B. (2007). The pluralism of fairness ideals: An experimental approach. *American Economic Review*, *97*(3), 818–827. doi:10.1257/aer.97.3.818

Conte, A. & Moffatt, P. G. (2014). The econometric modelling of social preferences. *Theory and Decision*, *76*, 119–145. doi:10.1007/s11238-012-9309-4

Cookson, R., Ali, S., Tsuchiya, A., & Asaria, M. (2018). E-learning and health inequality aversion: A questionnaire experiment. *Health Economics*, *27*(11), 1754–1771. doi:10.1002/hec.3799

Cropper, M., Krupnick, A., & Raich, W. (2016). *Preferences for equality in environmental outcomes* (Working Paper No. 25447). National Bureau of Economic Research. doi:10.3386/w22644

De Nardi, M., French, E., & Jones, J. B. (2010). Why do the elderly save? the role of medical expenses. *Journal of Political Economy*, *118*(1), 39–75. doi:10.1086/651674

Dirichlet, P. G. L. (1839). Sur une nouvelle méthode pour la détermination des intégrales multiples. *Journal de Mathématiques Pures et Appliquées*, *4*, 164–168.

Dolan, P. (1998). The measurement of individual utility and social welfare. *Journal of Health Economics*, *17*(1), 39–52. doi:10.1016/S0167-6296(97)00022-2

Dolan, P., Shaw, R., Tsuchiya, A., & Williams, A. (2005). Qaly maximisation and people's preferences: A methodological review of the literature. *Health Economics*, *14*(2), 197–208. doi:10.1002/hec.924

Dolan, P. & Tsuchiya, A. (2009). The social welfare function and individual responsibility: Some theoretical issues and empirical evidence. *Journal of Health Economics*, *28*(1), 210–220. doi:10.1016/j.jhealeco.2008.10.003

Dolan, P. & Tsuchiya, A. (2011). Determining the parameters in a social welfare function using stated preference data: An application to health. *Applied Economics*, *43*(18), 2241–2250. doi:10.1080/00036840903166244

Donaldson, D. & Weymark, J. A. (1980). A single-parameter generalization of the gini indices of inequality. *Journal of Economic Theory*, *22*(1), 67–86. doi:10.1016/0022-0531(80)90065-4

Donaldson, D. & Weymark, J. A. (1983). Ethically flexible gini indices for income distributions in the continuum. *Journal of Economic Theory*, *29*(2), 353–358. doi:10.1016/0022-0531(83)90053-4

Edlin, R., Tsuchiya, A., & Dolan, P. (2012). Public preferences for responsibility versus public preferences for reducing inequalities. *Health Economics*, *21*(12), 1416–1426. doi:10. 1002/hec.1799

Erreygers, G., Clarke, P., & Van Ourti, T. (2012). "mirror, mirror, on the wall, who in this land is fairest of all?"—distributional sensitivity in the measurement of socioeconomic inequality of health. *Journal of Health Economics*, *31*(1), 257–270. doi:10. 1016 / j. jhealeco.2011.10.009

Erreygers, G. & Kessels, R. (2017). Socioeconomic status and health: A new approach to the measurement of bivariate inequality. *International Journal of Environmental Research and Public Health*, *14*(7). doi:10.3390/ijerph14070673

Finkelstein, A., Luttmer, E. F., & Notowidigdo, M. J. (2013). What good is wealth without health? the effect of health on the marginal utility of consumption. *Journal of the European Economic Association*, *11*(1), 221–258.

Fisman, R., Kariv, S., & Markovits, D. (2007). Individual preferences for giving. *American Economic Review*, *97*(5), 1858–1876.

Fleming, M. (1952). A cardinal concept of welfare. *The Quarterly Journal of Economics*, *66*(3), 366–384. doi:10.2307/1885309

Gyrd-Hansen, D. (2004). Investigating the social value of health changes. *Journal of Health Economics*, *23*(6), 1101–1116. doi:10.1016/j.jhealeco.2004.02.002

Hardardottir, H., Gerdtham, U.-G., & Wengström, E. (2021). Parameterizing standard measures of income and health inequality using choice experiments. *Health Economics*, *30*(10), 2531–2546. doi:10.1002/hec.4395

Harless, D. W. & Camerer, C. F. (1994). The predictive utility of generalized expected utility theories. *Econometrica*, 1251–1289. doi:10.2307/2951749

Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, *63*(4), 309–321. doi:10.1086/257678

Hurley, J., Mentzakis, E., & Walli-Attaei, M. (2020). Inequality aversion in income, health, and income-related health. *Journal of Health Economics*, 102276. doi:10.1016/j.jhealeco.2019.102276

Kolm, S.-C. (1976). Unequal inequalities. i. *Journal of Economic Theory*, *12*(3), 416–442. doi:10.1016/0022-0531(76)90037-5

Makdissi, P. & Yazbeck, M. (2016). Avoiding blindness to health status in health achievement and health inequality measurement. *Social Science & Medicine*, *171*(1), 39–47. doi:10.1016/j.socscimed.2016.10.027

McNamara, S., Holmes, J., Stevely, A. K., & Tsuchiya, A. (2020). How averse are the uk general public to inequalities in health between socioeconomic groups? a systematic review. *European Journal of Health Economics*. doi:10.1007/s10198-019-01126-2

McNamara, S., Tsuchiya, A., & 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*, *269*, 113573. doi:10.1016/j.socscimed.2020.113573

O'Donnell, O., van Doorslaer, E., Wagstaff, A., & Lindelow, M. (2008). Analyzing health equity using household survey data : A guide to techniques and their implementation.

Parfit, D. (2000). The ideal of equality. In M. Clayton & A. Williams (Eds.), (Chap. Equality or Priority?, pp. 81–125). Basingstoke: Palgrave MacMillian.

Pinho, M. & Botelho, A. (2018). Inference procedures to quantify the efficiency–equality trade-off in health from stated preferences: A case study in portugal. *Applied Health Economics and Health Policy*, *16*, 503–513. doi:10.1007/s40258-018-0394-6

Pollak, R. A. (1971). Additive utility functions and linear engel curves. *The Review of Economic Studies*, *38*(4), 401–414. doi:10.2307/2296686

Ratcliffe, J. (2000). Public preferences for the allocation of donor liver grafts for transplantation. *Health Economics*, *9*(2), 137–148. doi:10.1002/(SICI)1099-1050(200003)9:2%3C137::AID-HEC489%3E3.0.CO;2-1

Richardson, J., Sinha, K., Iezzi, A., & Maxwell, A. (2012). Maximising health versus sharing: Measuring preferences for the allocation of the health budget. *Social Science & Medicine*, *75*(8), 1351–1361. doi:10.1016/j.socscimed.2012.05.036

Robson, M. (2021). Inequality aversion, self-interest and social connectedness. *Journal of Economic Behavior and Organization*, *183*, 744–772. doi:10.1016/j.jebo.2020.12.029

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

Roemer, J. (2002). Equality of opportunity: A progress report. *Social Choice and Welfare*, *19*, 455–471. doi:10.1007/s003550100123

Schokkaert, E. & Devooght, K. (2003). Responsibility-sensitive fair compensation in different cultures. *Social Choice and Welfare*, *21*, 207–242. doi:10.1007/s00355-003-0257-3

Ubel, P. A. & Loewenstein, G. (1996). Distributing scarce livers: The moral reasoning of the general public. *Social Science & Medicine*, *42*(7), 1049–1055. doi:10.1016/0277-9536(95)00216-2

Vickrey, W. (1960). Utility, strategy, and social decision rules. *The Quarterly Journal of Economics*, *74*(4), 507–535. doi:10.2307/1884349

Wagstaff, A. (1991). Qalys and the equity-efficiency trade-off. *Journal of Health Economics*, *10*(1), 21–41. doi:10.1016/0167-6296(91)90015-F

Wagstaff, A. (2002). Inequality aversion, health inequalities and health achievement. *Journal of Health Economics*, *21*(4), 627–641. doi:10.1016/S0167-6296(02)00006-1

Wagstaff, A., Paci, P., & Van Doorslaer, E. (1991). On the measurement of inequalities in health. *Social Science & Medicine*, *33*(5), 545–557. doi:10.1016/0277-9536(91)90212-U

Yitzhaki, S. (1983). On an extension of the gini inequality index. *International Economic Review*, *24*(3), 617–628. doi:10.2307/2526814

APPENDICES

Appendix A   Experiment Details

Appendix A.1   Experiment Overview

An overview of the experiment is shown in Figure A1. The experiment was run over two sessions, with instructions, three treatments, a belief elicitation and two questionnaires. The order participants went through the experiment is indicated by the arrows. The median times the participants completed each section are shown in minutes, in the bottom right corners. All participants went through all sections.

Figure A1: Experiment Overview
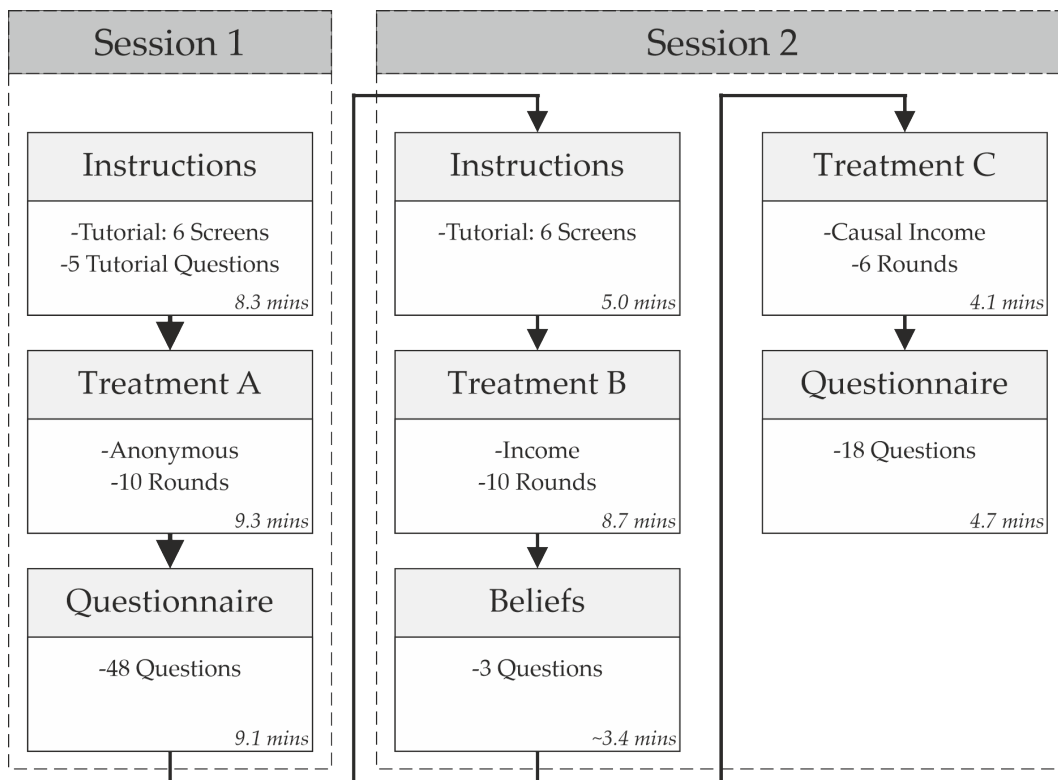
Appendix A.2   Instructions and Tutorial Script

The text for the instructions, tutorial and tutorial questions below are shown to all participants in Session 1, on screen within the experiment. The instructions give an overview of the experiment to come. The six stages of a tutorial explain how to use the on-screen interface; each of the scripts are followed by an interactive on-screen tutorial. Finally, five tutorial questions are presented to check and reinforce understanding.

Instructions

Welcome. Thank you for taking part today.

Please Read These Instructions Carefully.

You will be asked to make decisions which determine the health of hypothetical individuals in society.

You will be given a "Budget" that you must divide between these individuals. The Budget is the total amount of "Resources" available to spend.

Resources determine "Health". Health is the number of years a person lives, adjusted for illness or disability. For example, consider someone who reached the age of 70 without any illness or disability, who then lived for a further 10 years with an illness which reduced their quality of life to half of what it was before. That person might be said to have lived for the equivalent of 75 years in full health. For shorthand, we refer to this as "Health".

Giving more Resources to an individual increases their Health. The impact of Resources on Health is determined by a number referred to as the "Multiplier". The higher the Multiplier, the higher the level of Health achieved from a given number of Resources.

On the screen, you will distribute Resources between three Individuals. You will do this a number of times. Each screen will show a different scenario. The choices you make on one screen will not affect the scenarios that follow.

There are no right or wrong answers. We are interested in the choices you make, whatever they are.

You will now go through a tutorial, which will explain how to use the computer interface and the exact nature of the experiment.

Please click Next to continue.

Tutorial 1

This tutorial will show you how to use the on-screen interface.

You will first get practice in giving Resources to only one individual, who is identified by initials (e.g. MR). Drag the horizontal slider at the bottom of the next screen to the right to give more Resources to the individual.

The amount of Resources you give is shown by the height of the blue bar in the chart above the slider. Resources are also shown by the number to the left of the blue bar and by the number in the table at the top of the page.

Once you have dragged the slider, you can use the left and right arrow keys to make precise changes to the amount. Press the arrow key for a change of 0.1 and hold the arrow key for changes of 1.

The Resources you give to an individual are taken from the Budget, which is shown on the left of the screen. As you increase the Resources, the Remaining Budget will decrease.

You must always use all of the Budget, so that the Remaining Budget is zero. When there is only one individual, this means dragging the slider all the way to the right. Later you will have to distribute the Budget between individuals.

Press Next to try out the slider. When you are done, allocate all of the Budget (100) and press Next.

Tutorial 2

The Resources you give to an individual determines their "Health". Health is the number of years a person lives, adjusted for illness or disability. Health is equal to the Resources multiplied by a number we call the "Multiplier".

The Multiplier is shown in the table at the top of the screen. When you give Resources by moving the slider, the resulting Health is shown by the height of the grey bar. The number to the right of this bar is the amount of Health, and this is also shown in the table at the top of the screen.

For this first individual the Multiplier is 1. So, if you give all of the Budget of 100 to the individual, their Health will be 100. They will live 100 years in full health.

Next and see how Health changes as you adjust the Resources given to the individual. When you are done, allocate all of the Budget and press Next.

Tutorial 3

The Multipliers can vary from individual to individual. In the previous scenario, the Multiplier was 1. In the next scenario, it is 0.5.

Press Next and see how Health changes as you give more Resources to this individual. Notice that there is now a gap between the Resources given (blue bar) and the Health achieved (grey

bar). If you give all the Budget of 100 to this individual, their Health will be 50. They will live 50 years in full health.

When you are done, allocate all of the Budget and press Next.

Tutorial 4

In each round of the experiment, there are three individuals. Individuals are identified by their initials (e.g. MR, TO and OD) and change between rounds.

On the next screen, there are three sliders at the bottom of the screen that you can use to give Resources to each individual and so determine their Health.

The Resources and Health of each individual are shown by the blue and grey bars. The table at the top also shows the Resources, Multiplier and Health for each individual.

Now you must allocate the Budget across the three individuals. In doing so, you determine the Health of each one. In the example on the next screen, the Multipliers are the same for all three Individuals.

You must use all of the Budget, so that the Remaining Budget (on the left) equals zero.

Press Next and then give Resources to the three individuals. Remember: you can use the left and right arrow keys to make small changes to the amount of Resources. When you have used all of the Budget on the next screen, press Next.

Tutorial 5

The three individuals change from round to round.

On the previous screen, all three individuals had a Multiplier of 1. But the Multipliers can differ between individuals, as on the next screen.

Move the sliders to give Resources to the three individuals and notice how the Health achieved depends on the Multiplier of each individual. If you give the three individuals the same Resources, their Health will differ.

Take note of the size of the Budget, which can change from screen to screen.

Take note of the size of the Budget, which can change from screen to screen.

If you are having difficulty seeing both the table and the graph on your screen, zoom out on your web browser by holding "Ctrl" and pressing "-". Hold "Ctrl" and press "+" to zoom in.

Press Next and then give Resources to the three individuals. When you the Remaining Budget is zero, press Next.

Tutorial 6

The right of the screen shows further information.

"Resource Gap" is the gap between the largest and smallest amounts of Resources you give to the individuals.

"Total Health" is the total amount Health of the three individuals (e.g. Health to AC + Health to TD + Health to RC).

"Health Gap" is the gap between the largest and smallest amounts of Health achieved by the individuals.

If you have used the whole Budget, you will not be able to move any slider to the right. If you want to give more Resources to one individual, you will need to give less to another individual first.

Press Next and then distribute Resources across the three individuals. When you have allocated all of the Budget, press Next.

Tutorial Questions

Following the tutorial, participants answered five questions to reinforce and check understanding. The questions are shown below, with the correct response in bold. After submitting answers participants were given feedback about the correct response for each question.

1. On each screen, you will give Resources to how many individuals? - Options: 2; **3**; 4; Not Sure.

2. You can make and adjust the Resources you give by (tick all that apply): **Clicking and Dragging the Allocation Sliders**; **Using the Arrow Keys**; Moving the Vertical Bar; Not Sure.

3. If you give 100 Resources to an individual with a Multiplier of 1, then the Health of that individual will be? - Options: 25; 50; **100**; Not Sure.

4. If you give 100 Resources to an Individual with a Multiplier of 0.5, then the Health of that individual will be? - Options: 25; **50**; 100; Not Sure.

5. Once you have finished giving the Resources, you proceed to the next screen by: - Options: Clicking Next; **Ensuring the Remaining Budget = 0, then Clicking Next**; Waiting; Not Sure.

In Session 2, a modified version of the above instructions and tutorial are shown. First, to remind participants of the experiment, and second, to highlight the additional information on the income of each individual. They are told that income is an individual's "annual personal income (before tax) in pounds", which is shown in the label for each individual.

Appendix A.3 Multipliers

Table A1 shows the multipliers and relative multipliers used for Treatment A and Treatment B. There are 10 rounds within each treatment, the order is randomised between participants. Table A2 shows the multipliers and causal information shown to participants in Treatment C. The multipliers are orthogonal to the screen position, income rank and causal information.

Table A1: Multipliers in Treatment A and B

Round	Absolute, p_i			Relative, $p_i / \sum p_i$		
	Left	Middle	Right	Left	Middle	Right
1	1	1	1	0.33	0.33	0.33
2	1	1	0.33	0.43	0.43	0.14
3	1	0.33	1	0.43	0.14	0.43
4	0.33	1	1	0.14	0.43	0.43
5	1	0.33	0.5	0.55	0.18	0.27
6	0.5	1	0.33	0.27	0.55	0.18
7	0.33	0.5	1	0.18	0.27	0.55
8	1	0.5	0.33	0.55	0.27	0.18
9	0.33	1	0.5	0.18	0.55	0.27
10	0.5	0.33	1	0.27	0.18	0.55
Mean	0.7	0.7	0.7	0.33	0.33	0.33

Note. Columns give multipliers assigned to individuals labeled with initials in Treatment A, and incomes in Treatment B; distinguished by screen position: left, middle, and right.

Table A2: Information provided in Treatment C

Round	Incomes			Multipliers			Casual Information
	x_1	x_2	x_3	p_1	p_2	p_3	
1	D	E	F	0.33	0.5	1	*0%*
2	D	E	F	0.33	0.5	1	*20-80%*
3	D	E	F	0.33	0.5	1	*100%*
4	X	Y	Z	1	0.5	0.33	*0%*
5	X	Y	Z	1	0.5	0.33	*20-80%*
6	X	Y	Z	1	0.5	0.33	*100%*

Note: The randomly assigned incomes D, E, F and X, Y, Z are identical to those randomly drawn for the corresponding rounds (7 & 8) in Treatment B, either increasing or decreasing. Causal information shows the information provided on the percentage of the differences in the multipliers caused by income differences. 20-80% is a random draw from {20, 40, 60, 80}.

Appendix A.4 Belief Elicitation

Before participants begun Treatment C, three questions relating to their beliefs on the causal relationship between *Income* and *Health* (via the *Multipliers*) were asked. Question 1 asks whether, generally, participants believe that Multipliers would be Pro-Poor, Neutral or Pro-Rich. Question 2, split into 2a and 2b, gives two specific sets of Multipliers 1, 0.5, 0.33 and 0.33, 0.5 and 1 and asks the extent to which they believe the differences in Multipliers are caused by Income differences. The order of Pro-Poor, and Pro-Rich multipliers is randomised between participants, as is the screen order of poor to rich income individuals. For both questions interactive sliders are used. The script, alongside example screenshots, is below.

Question 1

The Health gained from an amount of Resources is determined by the value of the Multiplier. The Multipliers can differ between individuals. Resources can improve the Health of some people a lot, while improving the Health of other people much less. These differences can arise by chance. They may also be caused by characteristics of individuals.

Consider three individuals with different incomes. If income did not affect the Multipliers, they might look like this.
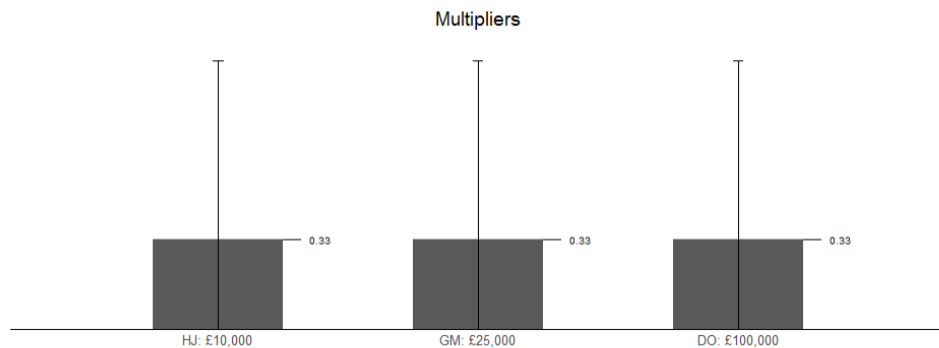

Do you expect that an individual's income would also affect the size of the Multipliers?

If you would expect higher income to increase the Multipliers, then move the slider to the right. Observe that the richer the individual, the greater is the increase in the multiplier. Move the slider further to the right if you expect income to cause larger differences in the Multipliers.

If you would expect lower income to increase the Multiplier, then move the slider to the left. Observe that the poorer the individual, the greater is the increase in multiplier. Move the slider further to the left if you expect income to cause larger differences in the Multipliers.

If you would expect income to have no effect on the Multiplier, then leave the slider in the middle.

Once you have made you choice, please press Next.
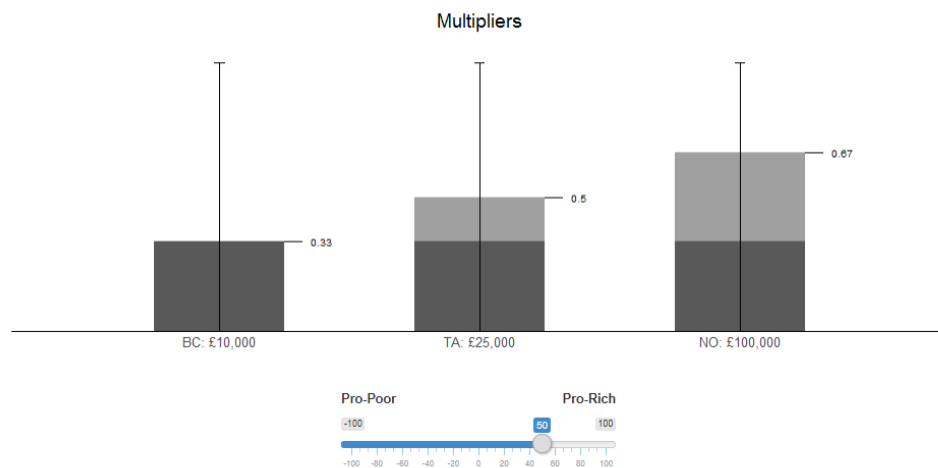
Question 2a/b

Now, imagine that the Multipliers of the three individuals are as shown below.

The differences between the Multipliers may be partly caused by the differences in incomes. However, the Multipliers may also differ by chance, or because of other characteristics of these individuals.

Multipliers

BC: £10,000   TA: £25,000   NO: £100,000
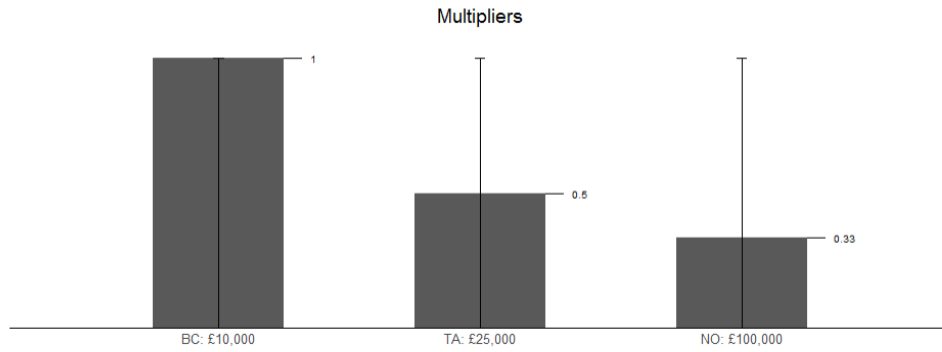
Move the slider below to show how much of the differences between the Multipliers you would expect to be caused by the differences in incomes.

Move the slider all the way to the left if you would expect NONE (0%) of the differences in the Multipliers to be caused by the differences in incomes.

Move the slider all the way to the right if you expect ALL (100%) of the differences in the Multipliers to be caused by the differences in incomes.

Move the slider to a point between these extremes to show the percentage of the differences in the Multipliers that you would expect to be caused by the differences in incomes.

Once you have made your choice, please press Next.

Multipliers

BC: £10,000   TA: £25,000   NO: £100,000

No Difference   All Difference
0   100
0 10 20 30 40 50 60 70 80 90 100
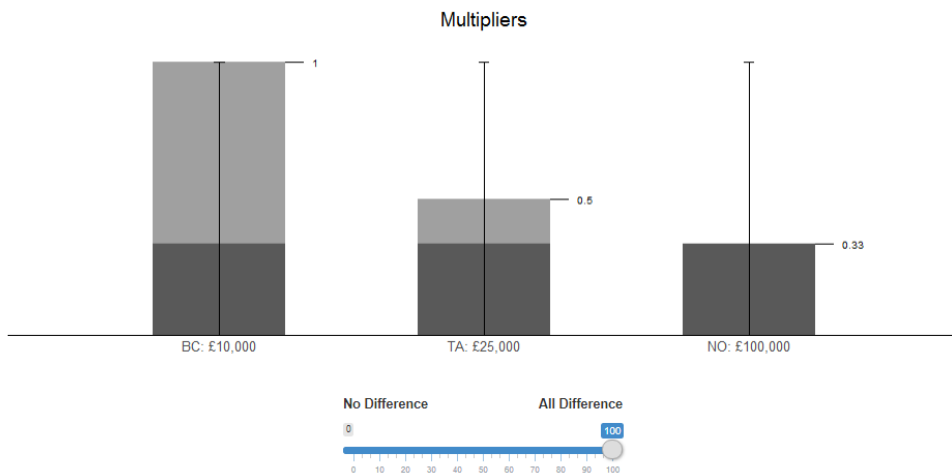
Appendix A.5 Treatment C

An example script and set of screenshots for Treatment C are shown below. Here, the participant is shown the Health allocations they made, in the round where the Multipliers were 0.33, 0.5 and 1, for Individuals who earn £5,000, £50,000 and £100,000, respectively. They are asked to imagine that these income differences caused 20% of the differences in the Multipliers. With this new information they are asked if (and how) they would reallocate resources differently.

Question 4

In a different previous scenario, the Multipliers were: 0.33, 0.5 and 1 to individuals BL: £ 5,000, EB: £50,000 and KR: £100,000, respectively.

In that scenario you allocated resources and health as follows:
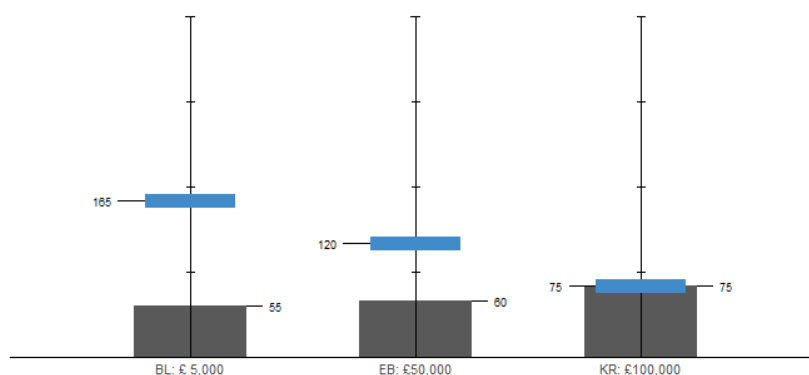
Imagine that the differences in income caused 20% of the differences in the Multipliers. Any remainder of the differences were caused by chance and the other characteristics of these individuals.

Then, the Multipliers would be as below, where the light grey part of the bars represents the part of the Multiplier caused by income differences. If there is no light grey bar, this means that none of the differences are caused by differences in income.
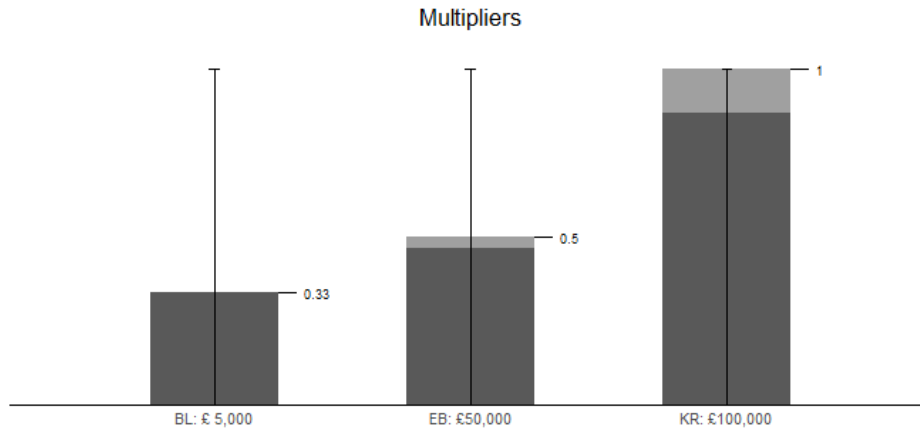
Multipliers

Given this new information, would you change how you distributed the Resources between the three individuals?

Press Next and make any changes on the next screen.

Appendix A.6   Pilot Experiments

Prior to our main experiment two pilot experiments were run. The first, a laboratory experiment with a student sample (n=32) at Erasmus University Rotterdam; which was held over two sessions on the 28th of November 2019 at the ESE-econlab. The second, was an online experiment, with a UK adult sample (n=27) recruited via Prolific, which ran on the 25th of October 2021. As with the main experiment, the experimental interface was designed in R Shiny.

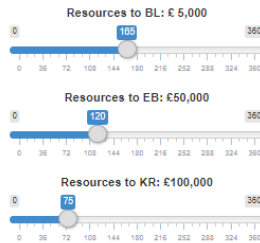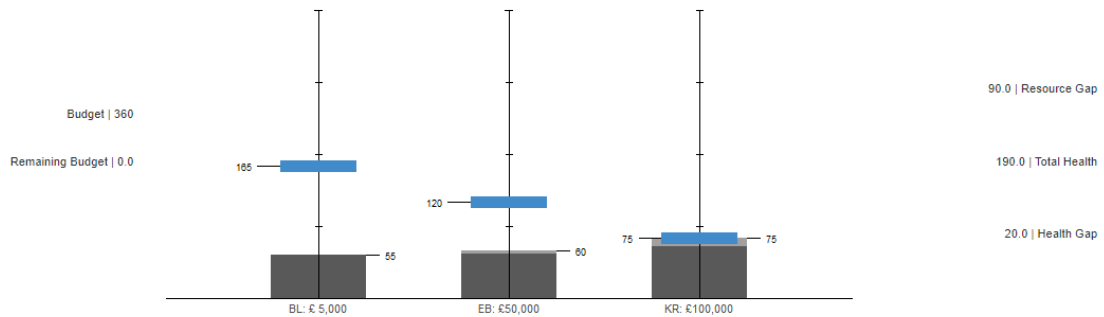The first pilot was, primarily, run to test the experimental design. To test comprehension and allow for the improvement of the experimental design, survey questions were included to ask: 1) how difficult participants found the experiment, 2) what was most difficult, and 3) what could be improved. As a result of these comments, changes were made to the instructions, interactive tutorial, experimental display, error messages and the questionnaire. To test the experimental design parameters, participant-level preference parameters were estimated. Only minor changes were made to increase the variation in multipliers in Treatment A and B. More major changes were needed for Treatment C. First, the addition of the causal belief questions and, second, an overhaul of Treatment C to ensure the causal λ parameters were orthogonal to the individual multipliers, p_i.

The second pilot aided the transition from a laboratory experiment, with a student sample, to an online experiment with a general population. The experiment was changed from a one one-hour session to two-30 minute sessions, the language in the tutorial was further simplified and an additional battery of demographic questions were added. Further changes were also needed to run the experiment through Prolific.

Appendix B   Descriptive Statistics

Table B1 shows the characteristics of the 337 participants included in our main analysis. Extensive information from our questionnaire is included, relating to: demographic characteristics, socioeconomic status, health, COVID-19, economic preferences, views and beliefs.

Table B1: Descriptive Statistics: Main Sample

	Mean	Min	Max	Obs
Demographics				
Female	0.50	0	1	337
Age	40.08	18	76	337
Married	0.50	0	1	337
Born in UK	0.78	0	1	334
White	0.79	0	1	337
Household Size	2.74	0	8	336
Children in HH	0.50	0	6	335
Highest Education				
- Postgraduate	0.24	0	1	336
- Undergraduate	0.35	0	1	336
- A-Level	0.25	0	1	336
- Secondary/Primary	0.16	0	1	336
Labour Market Status				
- Employed	0.61	0	1	335
- Unemployed	0.13	0	1	335
- Retired	0.08	0	1	335
- Student	0.11	0	1	335
- Other	0.06	0	1	335
Occupation				
- Manager/Professional	0.38	0	1	326
- Intermediate	0.31	0	1	326
- Never Worked	0.12	0	1	326
- Other	0.19	0	1	326
Income and Wealth				
Income (£k)	29.14	3	175	304
Wealth: House (£k)	195.77	0	1250	278
Wealth: Savings (£k)	26.22	0	125	274
Own Home	0.19	0	1	324
Subjective Socioeconomic Position	5.41	1	10	335
Health				
Health: Likert 0-100	68.91	10	100	334
Self-Assessed Health	2.18	1	5	336
Subjective Life Expectancy	79.61	32	165	330
Health Condition	0.20	0	1	330
Health Condition: Family	0.27	0	1	326
Treated for COVID-19	0.14	0	1	331
Cigarettes Smoked Per Day	1.10	0	25	336
Alcohol Units Per Week	5.16	0	50	337
Exercise Hours Per Week	4.90	0	23	337
Views				
Left-Right	3.93	0	10	319
Libertarian-Authoritarian	3.75	0	10	315
Health Ineq. vs Total Health	5.32	0	10	335

The characteristics of the English and Welsh adult population, according to the Office of National Statistics 2021 census are: 51.0% female, 81.7% white, with an approximate mean adult age of 49.1. We use this as a proxy for the characteristics of the UK adult population. Our sample is, therefore, younger than the adult population, but relatively representative in terms of ethnicity and sex. It is important to note, that the preferences estimated in the experiment appear to depend little on these demographic characteristics. OLS Regressions with inequality aversion, income weights or precision parameters as dependent variables which use age, age squared, female and white as explanatory variables all have R^2 values below 0.033. It is more important, given this is an online sample, that we have extensive variation across all socio-demographic characteristics, health states, preferences, political views and beliefs, as is shown in Table B1.

Appendix C Weights, Causality Beliefs and Attitudes

Equation (6) allows for a range of beliefs and attitudes concerning income-caused health inequality. Table C1 provides a summary of the consequences of combinations of beliefs, represented by λ, and attitudes, represented by β_2, for the weights given to the income-poor, which are determined, in part, by $\beta_1\beta_2^\lambda$. $\lambda > 0$, $\lambda < 0$, and $\lambda = 0$ indicate beliefs that the causal effect of income on health is positive, negative, and zero, respectively. $\beta_2 > 1$, $\beta_2 < 1$, and $\beta_2 = 1$ indicate aversion, inclination, and indifference to income-caused health inequality.

Table C1: Effect of causality beliefs and attitudes on weights to income-poor

		β_2		
		< 1	1	> 1
	< 0	Increase	Unchanged	Decrease
λ	0	Unchanged	Unchanged	Unchanged
	> 0	Decrease	Unchanged	Increase

Note: *Increase* (*Decrease*) indicates cases in which allowing for causality in the income-health relationship increases (decreases) weights on the health of the income-poor. *Unchanged* indicates cases in which allowing for causality has no impact on the weights on the income-poor.

Appendix D   Additional Results

Appendix D.1   Response of Allocations to Multipliers

We pool the Treatment A data over participants (k) and rounds (t) and estimate the following regressions for relative multipliers on resource shares and health shares, respectively,

$$\tilde{y}_{ikt} = a_y + b_y \tilde{p}_{ikt} + \nu_k + \epsilon_{ikt}, \tag{D1}$$

$$\tilde{h}_{ikt} = a_h + b_h \tilde{p}_{ikt} + \mu_k + \upsilon_{ikt}, \tag{D2}$$

where i indicates individuals and ν_k and μ_k are random effects at the participant level.

Since $h_{ikt} = p_{ikt} y_{ikt}$, $b_y = 0$ implies $b_h \approx 1$ and $b_h = 0$ implies $b_y \approx -1$. When $b_y \approx -1$ and $b_h = 0$, resources are allocated to fully offset productivity differences and keep health shares equal. Priority is given to reducing health inequality. As both b_y and b_h increase more resources and health are given to individuals with higher multipliers, which increases total health and health inequality.

Table D1 shows estimates of b_y and b_h. A negative estimate of $\hat{b}_y = -0.482$ and a positive estimate of $\hat{b}_h = 0.480$ indicates that, on average, participants allocate fewer resources to less productive individuals but not by enough to leave these individuals with a smaller share of health.

To explore heterogeneity in responses to the multiplier, we estimate participant-level regressions like eq.(D1) and eq.(D2) (without the random effects) using data from all rounds of Treatment A for each participant.[33] Figure D1 plots the distributions of the participant-level estimates of b_y and b_h, with the respective 95% confidence intervals.

---

[33]Goodness-of-fit is better for the participant-level regressions. The mean squared error (MSE) for the resource (health) share model is 0.0146 (0.0109) with pooled data and 0.0068 (0.0047) with participant-level estimation.

Table D1: Responses of resource and health shares to relative multiplier

	(1) Resource Coef./(S.E.)	(2) Health Coef./(S.E.)
Relative Multiplier	-0.4821***	0.4801***
	(0.034)	(0.029)
Constant	0.3333***	0.3333***
	(0.000)	(0.000)
Participants	337	337
Observations	10110	10110
R^2	0.2455	0.3123

Note: Table shows estimates of eq.(D1) and eq.(D2). The relative multiplier is re-centered around $1/3$. Random effects model. Robust standard errors in parentheses. p-values: $^*\, p < 0.10$, $^{**}\, p < 0.05$, $^{***}\, p < 0.01$.

Figure D1: Responses of resource and health shares to relative multiplier - distributions of participant-level estimates

Note: Figure shows cumulative densities of participant-level estimates of b_y and b_h from eq.(D1) and eq.(D2), respectively (without random effects). Thick line trace point estimates. Horizontal lines show robust 95% confidence intervals. The relative multiplier is re-centered around $1/3$.

There is extensive and significant heterogeneity in responsiveness to the relative multipliers. For a substantial percentage of participants ($\approx 27.9\%$), the null of $b_h = 0$ is not rejected (at 5% significance level). These participants allocate health equally between individuals.

66

Since this is a relatively easy rule to follow, the confidence intervals are narrow for those with a point estimate close to $b_h = 0$. For a small minority ($\approx 6.2\%$) we do not reject that null of $b_y = 0$ (and therefore $b_h = 1$). These participants do not change the resource shares as the multiplier changes, and so they give more health to individuals for whom resources are more productive. Approximately 14.8% of the participants prioritise efficiency and so give more resources, and therefore health, to individuals with higher multipliers (i.e. $b_h > 1$). Around half ($\approx 49.3\%$) of participants give less resources to individuals with higher multipliers, whilst still ensuring that these individuals end up with larger health shares (ie. $0 < b_h < 1$). These participants sacrifice some efficiency to get less inequality.

Appendix D.2   Response of Allocations to Income

We pool the Treatment B data over participants (k) and rounds (t) and estimate the following regressions for health shares,

$$\tilde{h}_{ikt} = a + f\left(x_{ikt}\right) + \mu_k + \upsilon_{ikt}, \tag{D3}$$

where i indicates individuals, x_{ikt} is individual income, and μ_k are random effects at the participant level. In different regressions, we specify $f\left(x_{ikt}\right)$ as income rank within round t, categories of income level, log income, and income share within round t.

The estimates in Table D2 show that regardless of how income is specified it has a large and significant effect on the resources allocated to an individual and, hence, the share of health they get. For example, using income rank, the richest individual, on average ends up with a health share that is 3.69 percentage points smaller than the share of the poorest individual. Participants are, on average, pro-poor.

Table D2: Response of health share to income in Treatment B

	(1) Rank Coef./(S.E.)	(2) Level Coef./(S.E.)	(3) Log Coef./(S.E.)	(4) Prop. Coef./(S.E.)
Rank Income				
- Middle	-0.0163***			
	(0.004)			
- Richest	-0.0369***			
	(0.006)			
Income Level				
- £10,000		-0.0092**		
		(0.004)		
- £25,000		-0.0148***		
		(0.005)		
- £50,000		-0.0242***		
		(0.005)		
- £100,000		-0.0405***		
		(0.007)		
Log Income			-0.0125***	
			(0.002)	
Prop. Income				-0.0551***
				(0.009)
Constant	0.3511***	0.3511***	0.4588***	0.3517***
	(0.003)	(0.004)	(0.022)	(0.003)
Participants	337	337	337	337
Observations	10110	10110	10110	10110
R^2	0.0166	0.0139	0.0132	0.0156

Note: Table shows estimates of eq.(D3). Random effects models. Robust standard errors in parentheses. p-values: $^{*}\,p < 0.10$, $^{**}\,p < 0.05$, $^{***}\,p < 0.01$.

Appendix D.3   Response of Allocations to Causal Information

We estimate the following regression for the health shares of the poorest individual,

$$\tilde{h}_{ikt} = a + bI_{ikt} + \mu_k + \upsilon_{ikt}, \tag{D4}$$

where I_{ikt} is the exogenously varying causal information — the percentage of the variance in the multipliers (p_{ikt}) that is caused by variation in income (x_{ikt}) (Table 1) — i is identified by $x_{ikt} < x_{jkt}\,\forall j \neq i$, and μ_k are random effects at the participant level. We estimate

separate regressions for the cases in which the poorest individual is the a) least productive, $p_{ikt} < p_{jkt} \, \forall j \neq i$, and b) most productive, $p_{ikt} > p_{jkt} \, \forall j \neq i$.

Table D3 presents estimates of these regressions. In rounds where the poorest individual is also the least productive (lowest multiplier), information about the causal impact of income on productivity has no effect on the share of health allocated to the poorest individual. When that individual is the most productive, there is a significant, albeit small, effect. Increasing the percentage of the variation in the multipliers that is explained by income differences from 0% to 100% is estimated to increase the share of health given to the poorest individual by 1.65%.

Table D3: Effect of causal information on health share to poorest

|	(1) Low p_i Coef./(S.E.)	(2) High p_i Coef./(S.E.)
Causal Information, I	0.0000	0.0165***
	(0.004)	(0.005)
Constant	0.2866***	0.4330***
	(0.005)	(0.007)
Participants	337	337
Observations	1011	1011
R^2	0.0000	0.0022

Note: Table shows estimates of regression model eq.(D4) stratified by whether poorest individual has the lowest (left) or highest (right) multiplier in a round. Random effects model. Robust standard errors in parentheses. p-values: $^{*} p < 0.10$, $^{**} p < 0.05$, $^{***} p < 0.01$.

Figure D2 shows the distribution participant-specific estimates of regression model eq.(D4) (without the random effects) again stratified by whether the poorest individual is the least or most productive. Most participants do not respond to changes in the causal information — the null is not rejected (5% significance) for 88.1% and 90.8% of participants when the poorest individual is the least and most productive, respectively. When the poorest individual is most productive, the proportion of participants who give more health to the poorest when told that income causes a larger percentage of the difference in multipliers is

greater than the proportion who reduce the health share of the poorest when given such information. This difference results in the positive significant effect estimated with the pooled data shown in Table D3.

Figure D2: Distribution of effects of causal information on health share to poorest
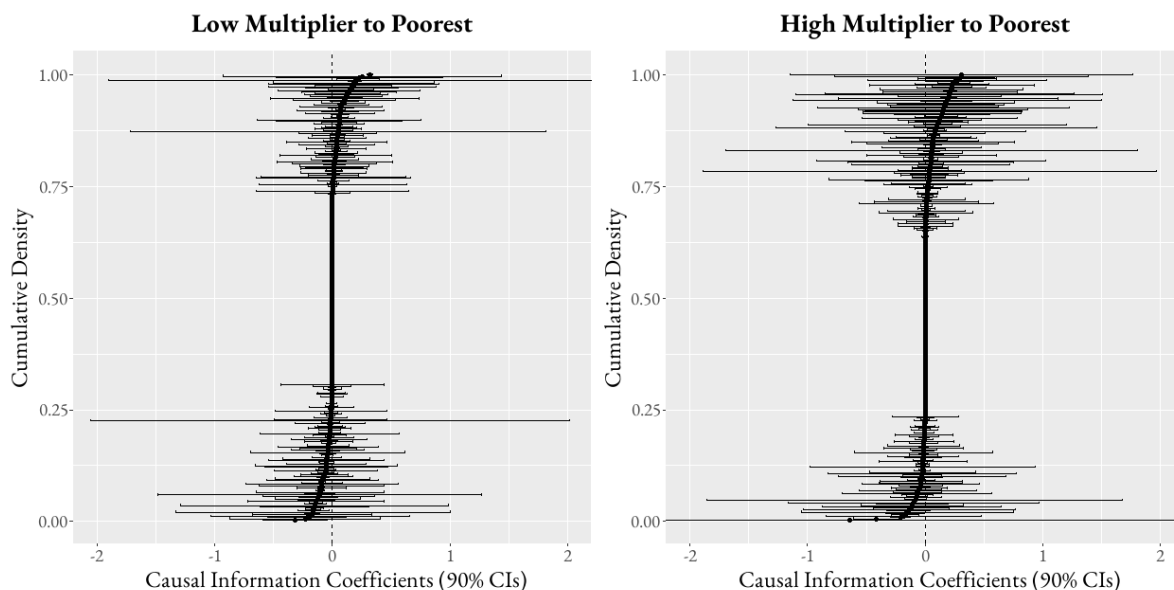
Note: Figure shows distributions of participant-specific estimates of slope coefficient from regression eq.(D4) (without random effects) stratified by whether poorest individual is least productive (left) or most productive (right). Interval lines show robust 95% confidence intervals.

Appendix D.4 Robustness to Exclusion of Participants

After the tutorial in the first session of the experiment, participants were asked five questions to check comprehension of the task. Those who answered fewer than three questions correctly were excluded from the analysis sample. Using all participants in the analysis sample plus those excluded, Table D4) shows estimates from OLS regressions of participant-specific estimates of (log) health inequality aversion (ε), precision (σ), and goodness-of-fit (GOF) measures on an indicator of exclusion from the analysis sample. The parameter estimates and GOF measures are from the Treatment A data.

70

Table D4: Differences between excluded participants and analysis sample

	(1) Ineq. Aver. Coef./(S.E.)	(2) Precision Coef./(S.E.)	(3) MPL Coef./(S.E.)	(4) MSE Coef./(S.E.)
Excluded	0.5281	-34.0155***	-0.0810***	0.0239**
	(0.822)	(11.293)	(0.026)	(0.011)
Constant	2.3367***	114.7929***	0.8917***	0.0172***
	(0.209)	(3.666)	(0.007)	(0.002)
Participants	381	381	381	381
R^2	0.0018	0.0253	0.0405	0.0247

Note: Each column gives estimates from OLS regression of participant-specific estimates of respective outcome on an indicator of sample exclusion. Outcomes are from Treatment A data. For inequality aversion, we use $log(\hat{\varepsilon})$ to deal with extreme values of $\hat{\varepsilon}$. MPL is what we call the Mean Proportional Likelihood. Define $PL_t = L_t/(L_t + L_t^U)$, where L_t is the likelihood in round t for the data and estimates and L_t^U is a likelihood for a uniform distribution draw. $MPL = 1/T \sum_t^T (PL_t)$. If $MPL = 0.5$, model fit to data is no better than fit uniform distribution draws. As $MPL \to 1$, data fit improves. $MSE = 1/T \sum_t^T (\tilde{y}_i - \tilde{y}_i^*)^2$ is the Mean Square Error. GOF is increasing with MPL and decreasing with MSE. Robust standard errors in parentheses. p-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Inequality aversion is not significantly different for the excluded participants. However, precision is significantly lower for these participants and both GOF measures show significantly worse fit to the estimates obtained for them. This suggests that those who are excluded from the analysis because of apparent poor comprehension of the task indeed give estimates that imply that they have less precise optimal allocations and they make more errors through choices that deviate from those allocations.

Appendix D.5   Distributions of Parameter Estimates

Figure D3 plots the empirical CDFs of participant-specific estimates of parameters obtained separately from Treatment A, B, and C data. The top-left panel reveals that Treatments A gives a larger proportion of lower estimates of ε, while Treatment C gives a larger proportion of higher estimates.

The top-right panel plots empirical CDFs of $\hat{\beta}$ from Treatment B and $\hat{\beta}_1$ from Treatment C. These parameters reflect aversion to income-related health inequality irrespective of its causal source. Estimates of β and β_1 may differ because they are obtained from different scenarios — multiplier and income combinations. Using Treatment B, 27.0% of participants have $0.95 \leq \beta \leq 1.05$, and so, approximately, display no discrimination in favour of either poorer or richer individuals. With Treatment C, the respective percentage of income neutral participants with $0.95 \leq \beta_1 \leq 1.05$ is lower, at 23.7%. The percentage displaying pro-poor weights is slightly larger for C (50.4%, $\beta_1 > 1.05$) than for B (49.3%, $\beta > 1.05$). The percentage displaying pro-rich weights is also larger for C (25.8%, $\beta_1 < 0.95$) than for B (23.7%, $\beta < 0.95$).

Figure D3: Distribution of preference parameter estimates by treatment
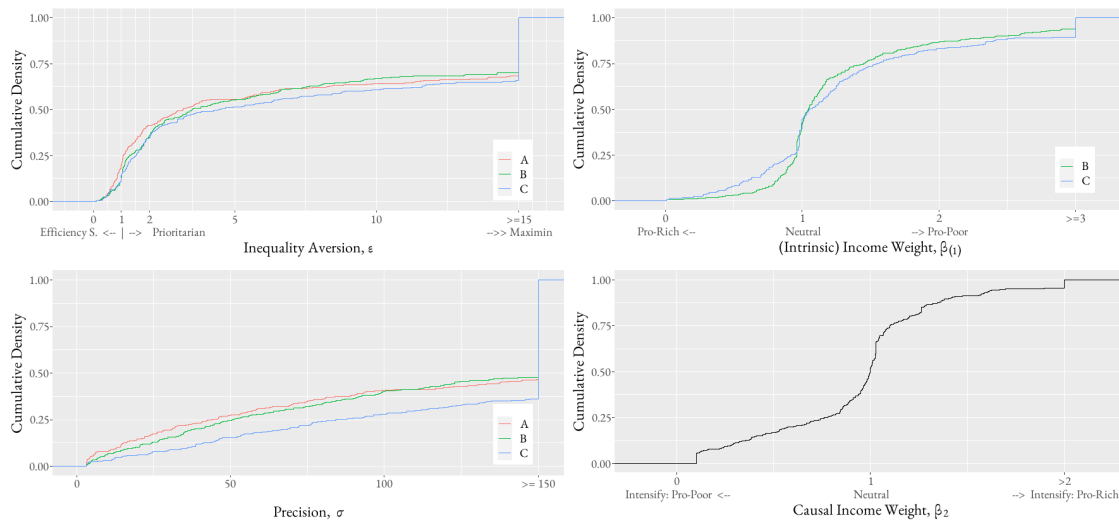
Notes: Empirical CDFs of participant-specific estimates of preference parameters obtained from choices made in each treatment.

The bottom-left panel shows CDFs of the precision parameter (σ) estimates. Treatment C appears to produce the most precise estimates (largest $\hat{\sigma}$). This is partly an artefact of the smaller number of rounds in this treatment. Up to about the 40th percentile, Treatment B is more precise than Treatment A. But the latter treatment gives a larger proportion

of very precise estimates. There are only 8.0%, 6.5%, and 3.0% of participants from A, B, and C, respectively, for whom $\hat{\sigma} < 10$. Comparing these estimates with the respective pooled estimates of 8.09, 8.71, and 8.99 confirms that allowing for heterogeneity across participants produces predicted allocations that are far more precisely centred around the optimal allocation.

Table D5 shows rank correlation coefficients between participant-level preference parameters, estimated within the three treatments: A, B and C. Between treatments, there are strong and significant correlations for each of ε and σ for all treatments, and between β and β_1, estimated in Treatment B and C. This shows consistency in the estimated preference parameters across different treatments. Within treatments, we do not see a significant correlation between ε and β in Treatment B, nor between ε and β_1 or β_1 and β_2 in Treatment C. We do see positive and significant correlations between ε and σ, this is in part explained by the low noise for *maximin* participants. We also see a positive correlation between ε in Treatment A and β in Treatment B, but not between ε and β in Treatment B.

Table D5: Within and Between Treatment Rank Correlation Coefficients for Participant-Specific Estimates

		ε A	ε B	ε C	β B	β_1 C	β_2 C	σ A	σ B	σ C
ε	A	1								
ε	B	0.76*	1							
ε	C	0.64*	0.83*	1						
β	B	0.15*	0.06	0.04	1					
β_1	C	0.11	-0.01	0.04	0.57*	1				
β_2	C	0.001	-0.02	-0.11*	-0.11*	-0.03	1			
σ	A	0.64*	0.51*	0.42*	0.14*	0.12*	0.01	1		
σ	B	0.48*	0.58*	0.51*	-0.05	-0.02	0.00	0.57*	1	
σ	C	0.38*	0.42*	0.49*	0.02	0.08	0.04	0.43*	0.56*	1

Note: Spearman's rank correlation coefficients. A, B, and C refer to treatments. For example, the second top cell in the first column shows the correlation between estimates of ε obtained from Treatment A and B. The third top cell in the second column shows the correlation between estimates of ε and β obtained from Treatment B. A star indicates that the correlation is significantly different from 0 at the 5% level.

Appendix E   Model Comparisons

Appendix E.1   Participant-Level Goodness of Fit

Using Treatment B data, Figure E1 shows cumulative distributions of participant-level Mean Proportional Likelihoods ($MPLs$). The left panel compares distributions obtained from the pooled and participant-specific estimates. The distribution for the pooled estimates is derived by using those same estimates to calculate a MPL for each participant. The distribution for the participant-specific estimates uses each participant's estimates to calculate their MPL. The distribution of $MPLs$ using the participant-specific estimates strictly dominates. A MPL of 0.5 indicates that a model fits the data no better than taking random draws from a uniform distribution. A $MPL < 0.5$ is even worse. $MPL \leq 0.5$ for 5.9% of participants using the pooled estimates and 0% of participants using the participant-specific estimates. There are 92.9% of participants for whom $MPL \geq 0.75$ with the participant-specific estimates, while 0% of participants reach such a high MPL with the pooled estimates.

The middle and right panels show cumulative distributions of $MPLs$ from participant-specific estimates obtained using alternative specifications of social welfare and weighting functions, respectively. The non-consequentialist welfare function clearly gives the worst data fit across the whole distribution. The Kolm welfare function is worse than Atkinson over the interval $0.5 < MPL \leq 0.9$. The Atkinson, Kolm, and non-consequentialist SWFs give the best fit (highest MPL) for 64.1%, 31.5%, and 4.5% of participants, respectively.

Constant weights clearly gives the worst fit, while there appears little difference in the fit given by income-rank and income-share weighting functions over the full MPL distribution. However, the income-rank specification gives the best fit for 46.9% of participants, while the income-share weighting function is the best fit for only 32.0% of participants. Constant weighting is the best fit for only 21.1% of participants.

Figure E1: Model comparisons using distributions of Mean Proportional Likelihoods
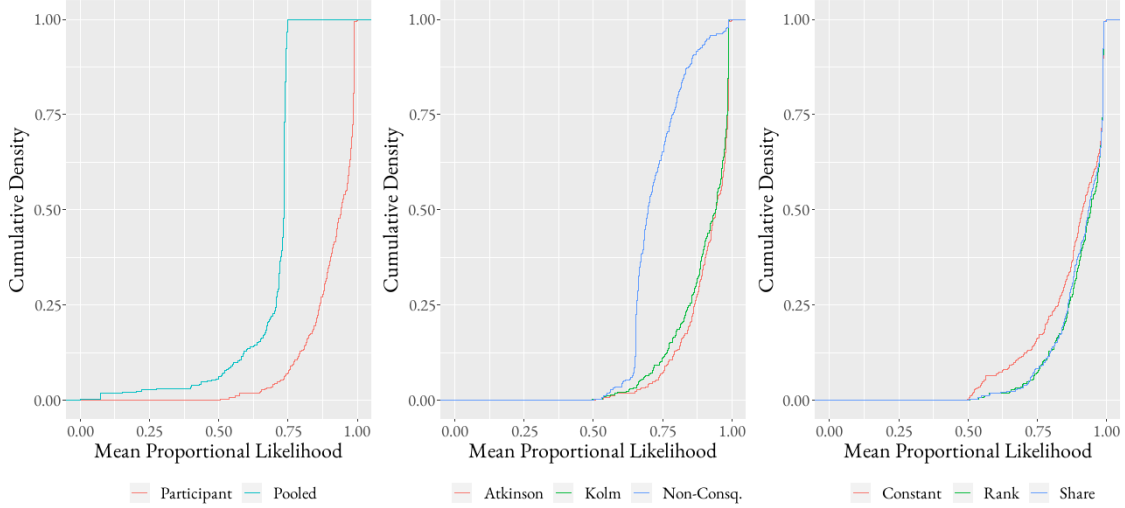

Figure shows empirical CDFs of Mean Proportional Likelihood (MPL) calculated with pooled vs participant-specific estimates (left), different specifications of SWF (middle), and different specifications of weighting function (right). $MPL = 1/T \sum_t^T (PL_t)$, where t is a round of the experiment, $PL_t = L_t/(L_t + L_t^U)$, L_t is the likelihood value for the data and estimates, and L_t^U is likelihood value for a uniform distribution draw. As $MPL \to 1$, the estimates have an increasingly high likelihood. If $MPL = 0.5$, a uniform random draw is equally likely, while as $MPL \to 0$ the uniform random draws become relatively more likely. Each CDF is drawn for 337 estimates of MPL, one for each participant.

Appendix E.2   Income-Related or Income-Caused: Treatment C

Using Treatment C data, Table E1 shows parameter estimates and goodness-of-fit measures for two alternative weighting functions. The first model is our main specification (see eq.(6)), which allows weights to depend both on the intrinsic- and causal-income parameters, β_1 and β_2, where $\beta = \beta_1 \beta_2^\lambda$. The second model restricts this function, by assuming $\beta_2 = 1$, and thus only estimates the income-related weight, β (see eq.(5)). The comparison of these two models allows for the identification of the importance the causal-income weight parameter.

Pooled estimates of the main model, reveal that β_2 is close to 1.[34] The estimates for ε are also similar to the restricted model, and $\beta \approx \beta_1$. This is reflected in the goodness-of-fit measures, with the MPL and MSE revealing no difference in fit, with the AIC showing

---

[34]Indeed, bootstrapped 95% confidence intervals of [0.892 - 1.048] show that it is not significantly different from 1.

Table E1: Income-Related or Income-Caused - Treatment C

		Parameters				Goodness-of-Fit		
		Inequality Aversion ε	Income-Related $\beta_{(1)}$	Income-Caused β_2	Precision σ	MPL	MSE	AIC
Pooled	Main	1.670	1.102	0.956	9.044	0.694	0.038	-4682.4
	Restricted	1.604	1.100		9.040	0.694	0.038	-4683.6
Participant	Main	4.366	1.067	0.999	150.0	0.941	0.010	-16240.3
	Restricted	3.363	1.103		157.1	0.931	0.011	-16590.4

Note: Pooled and participant-specific median estimates from Treatment C data. MPL is Mean Proportional Likelihood, MSE is Mean Square Error and AIC is Akaike Information Criterion as defined in Table 4.

that the restricted parameter is preferred. Participant-specific estimates are shown below the pooled estimates. The median β_2 is 0.999, again very close to 1. However, as shown in our main analysis, there is significant heterogeneity in β_2 estimates, with 10th and 90th percentile β_2 estimates of 0.294 and 1.416, respectively. For goodness-of-fit measures we do see a slightly better fit, with a higher MPL and lower MSE. However, as shown by the AIC the improvement in fit is not sufficient to compensate for the additional parameters.

In sum, this reaffirms our main findings that although participants are averse to health inequality and particularly income-related health inequality, they are not more averse to income-related health inequality when low income causes poor health.

Appendix E.3   Health-rank-dependent Weights

The model given by eq.(1), eq.(2), eq.(4) is subject to the criticism there is no distinction between inequality aversion and the concave valuation of health (Bleichrodt, Diecidue, and Quiggin, 2004). The parameter ε can be interpreted as reflecting either inequality aversion or diminishing marginal utility of health. The rank-dependent QALY model (Bleichrodt, Diecidue, and Quiggin, 2004; Bleichrodt, Doctor, and Stolk, 2005) makes this distinction.

In our experimental setup, participants choose optimal health levels, which determine health ranks. It is not possible to separately identify nonlinear aggregation of health (QALYs) and health-rank-dependent weights. A tractable alternative is to define the weights as a function of the fractional rank of the multipliers (p_i), such that the weights reflect the "potential" health rank of an individual. Using eq.(5), and replacing incomes by the productivity factors, with $p_i \geq p_{i-1}$, we specify the weights as:

$$\omega_i^P(\beta_p) = \int_{r(p_{i-1})}^{r(p_i)} \beta_p(1-q)^{\beta_p-1} dq \tag{E1}$$

Using this alternative weighting function we can estimate preference parameters, pooling data from Treatment A across all participants and rounds. The estimates and goodness-of-fit statistics are shown in Table E2. The estimate of β_p is smaller than 1, which would indicate that participants give lower weight to individuals with lower multipliers (productivity). Inclusion of β_p appears to slightly increase the estimate of ε. However, $\hat{\beta}_p$ is not significantly different from 1 (i.e. constant weights). Moreover, from the goodness-of-fit statistics there is little difference between the two models. Therefore, there is no strong support for a model with non-constant weights when individuals are not labelled by income.

Table E2: Health-rank-dependent weighting function - pooled estimates

|	Parameters			Goodness-of-Fit		
	Inequality Aversion ε	Productivity-Weight β_p	Precision σ	MPL	MSE	AIC
Equal	1.391 [1.323 - 1.470]		8.047 [7.354-8.910]	0.671	0.043	-7156.7
Productivity	1.545 [1.281 - 1.843]	0.927 [0.802 - 1.049]	8.049 [7.357-8.875]	0.671	0.043	-7156.8

Note: ε reflects the (representative) SDM's aversion to health inequality; see eq.(2). β_p determines rank multiplier weights in eq.(E1). σ is reflects precision of the SDM; see (eq.(13)). In brackets are 95% bootstrap confidence intervals obtained with the percentile method. MPL is Mean Proportional Likelihood, MSE is Mean Square Error and AIC is Akaike Information Criterion as defined in Table 4.

Appendix F   Details of Simulated Policy Evaluation

We simulate a population of 100,000 individuals characterised by income and health. Annual income, x_i, is drawn from a (rescaled) Beta distribution, where $x_i \sim Beta(1.1, 10.5) \times 360000$. Baseline health (QALYs), h_i, is a positive and stochastic function of log income, where $h_i = 45 + 2log(x_i) + \upsilon_i$, and $\upsilon_i \sim N(0, 3)$. We assume there are two policies, A and B, that have the same cost and do not change the distribution of income. We simulate the potential outcomes of each individual's health with Policy A and B, where $h_{iA} = h_i + 0.75log(x_i)(1 - r(x_i))$ for Policy A, and $h_{iB} = h_i + 0.75log(x_i)r(x_i)$ for Policy B. Intuitively, we can think of Policy A as allocating proportionately less resources to individuals according to their income rank, whilst Policy B allocates proportionately more. The productivity of these resources, in determining health, is a positive function of log income. The marginal and joint simulated distributions of income and health are shown in Figure F1.

Figure F1: Marginal and joint simulated distributions of income and health
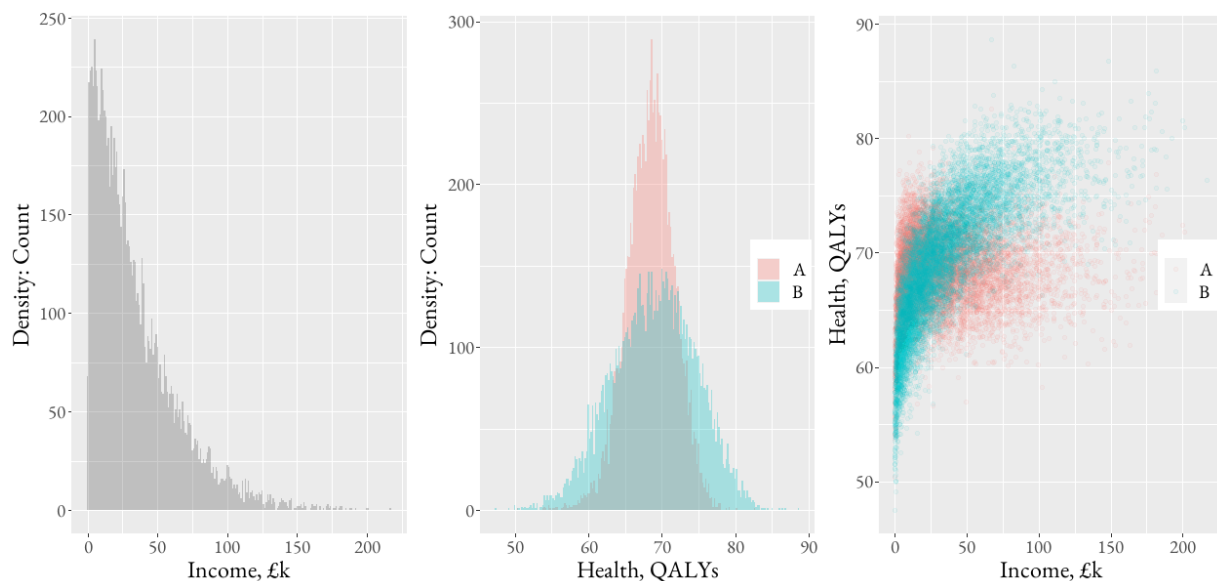
These simulated distributions are intended to demonstrate the potential to use our estimates to evaluate and rank policies. The minimum data requirement is a distribution of

health (QALYs) across individuals or groups for each policy. If data on incomes (or income ranks) of the individuals or groups are also available, then the evaluation can account for aversion to income-related health inequality as well as aversion to pure health inequality.

A data file containing the distributions of estimated parameters and code in R to conduct the analysis is available in an online repository, on Mendeley Data: https://doi.org/10.17632/9vy6f6g5k3.1. This includes an RData file, with the preference parameters estimated in this paper, alongside an R Script file, which provides a function to allow researchers and policy makers to evaluate other policies or interventions. All that is required to use our parameters to conduct a distributionally-sensitive policy evaluation is a dataset containing a treatment indicator, alongside health (and income) levels of individuals (or groups) with and without treatment. We provide example code to demonstrate how to use this function in R.

Appendix G   Causal Beliefs

After exposure to Treatment B and before exposure to Treatment C, we elicited beliefs about a causal effect of income on the multiplier (Appendix A.4). Before doing so, we gave a reminder that the multiplier determines health gained from resources. We told participants that, in any round of the experiment, the multipliers could differ between the three individuals by chance and, possibly, because of their characteristics.

We asked participants whether they would expect income to affect the multiplier. They used a slider to answer on a scale from -100 indicating strong belief that lower income increases the multiplier to 100 indicating strong belief that higher income increases the multiplier, with 0 indicating that income has no effect on the multiplier. We refer to this variable as the *General Causal Belief.* Two participants did not complete the belief elicitation exercise and so the sample size is 335 for analyses using these data.

The left panel of Figure G1 shows the cumulative density of this General Causal Belief. A median of 28 indicates weak belief that higher income causally increases the multiplier and so causes richer individuals to gain more health from a given allocation of resources. Almost two thirds (64.8%) of participants believe in this pro-rich generation of health. Less than a quarter (22.7%) of participants believe the causal relationship is pro-poor. Around an eighth (12.5%) express a belief that income has no impact on the multiplier and so the generation of health.

After eliciting the General Causal Belief, participants were shown two scenarios with multipliers and incomes of three individuals as in rounds 7 and 8 of Treatments B and C. In both scenarios, the multipliers are {0.33, 0.5, 1}. In one scenario, the multipliers increase monotonically with income across the three individuals (pro-rich). In the other, they decrease with income (pro-poor). In each case, we asked participants to use a slider, which was set to a random starting position between 0% and 100%, to indicate the percentage of the differences

Figure G1: Distributions of Causal Beliefs
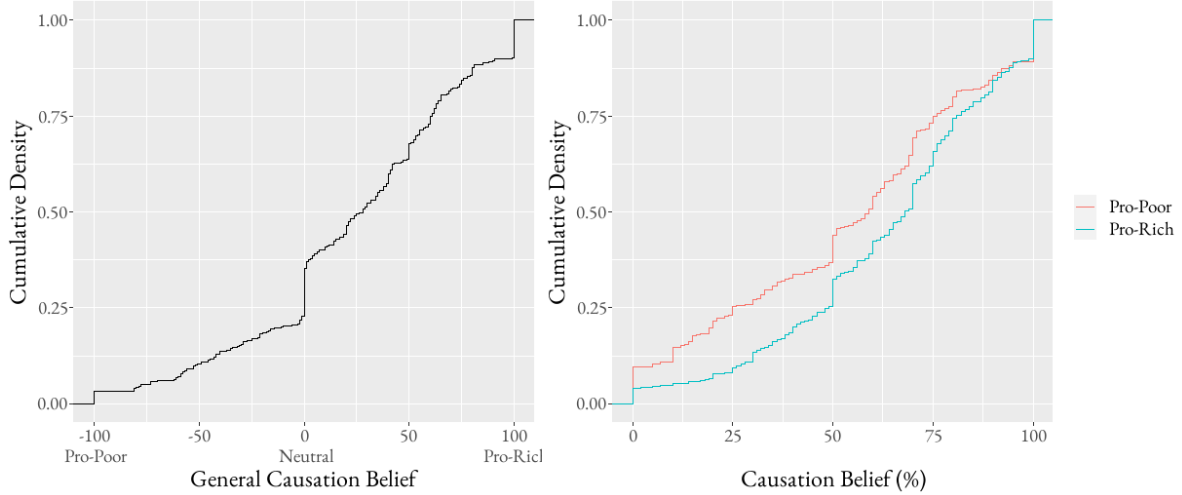
Note: Left panel shows the empirical cumulative density function (CDF) of *General Causal Belief* — extent to which strongly believe that income causally increases the multiplier (larger positive value) or that income causally reduces the multiplier (larger negative value). Right panel shows CDFs of *Pro-Rich Causal Belief* and *Pro-Poor Causal Belief*, which are beliefs about the percentage of multiplier differences caused by a positive and negative income effect, respectively. n=335 for both panels.

between the multipliers that was caused by the income differences (Appendix A.4). We refer to these responses as *Pro-Rich Causal Belief* and *Pro-Poor Causal Belief* for the scenarios in which the multipliers increase and decrease with income, respectively.

The right panel of Figure G1 shows cumulative densities of the Pro-Rich Causal Belief and the Pro-Poor Causal Belief. In the former case, participants tend to believe that a higher percentage of the variation in the multipliers is caused by income. The median belief is that income causes 68% of the differences in multipliers when they are positively associated with income. The respective median belief is 59% when the multipliers are negatively associated with income. Together with the distribution of the General Causal Belief, these results indicate stronger belief in a positive causal effect of income on the multiplier.

To check the validity of the belief elicitation instruments and participants' comprehension of them, we regress the General Causal Belief on the Pro-Rich Causal Belief and the Pro-Poor Causal Belief. We re-scale the latter two beliefs to 0-1. Table G1 shows that

the General Causal Belief is significantly positively associated with the Pro-Rich Causal Belief and significantly negatively associated with the Pro-Poor Causal Belief. This indicates consistency in responses between the questions. Participants who attribute a larger percentage of multiplier differences that are positively associated with income to a positive causal effect of income (larger Pro-Rich Causal Belief) report stronger belief that income causally increases the multiplier (positive and larger General Causal Belief). Participants who attribute a larger percentage of multiplier differences that are negatively associated with income to a negative causal effect of income (larger Pro-Poor Causal Belief) report stronger belief that income causally reduces the multiplier (negative and larger General Causal Belief). Although the partial associations and the R^2 statistic are small, the significance of the relationships between the elicited beliefs in the expected directions suggests that participants comprehended what they were being asked to do.

Table G1: Associations between Causal Beliefs

	General Causal Belief	
	Coef.	(SE)
Pro-Poor Causal Belief	-0.268***	(0.094)
Pro-Rich Causal Belief	0.402***	(0.107)
Constant	12.15*	(6.20)
Participants	335	
R^2	0.045	

Notes: Estimates from least squares regression of General Causal Belief on Pro-Rich Causal Belief and Pro-Poor Causal Belief. See notes to Figure G1 and text above for variable definitions. Pro-Rich Causal Belief and Pro-Poor Causal Belief re-scaled to 0-1. Data from Treatment B rounds with multipliers of {0.33, 0.5, 1}.

We now assess whether beliefs about the causal effect of income on the multipliers influence allocations. Figure G2 shows cumulative densities of the health share to the poorest individual ($\tilde{h}_i$) in rounds of Treatment B stratified by intervals of beliefs. In the left panel, we use rounds in which the multipliers are positively associated with income and stratification is by intervals of Pro-Rich Causal Belief. In the right panel, we use rounds in which

the multipliers are negatively associated with income and stratification is by intervals of Pro-Poor Causal Belief.

Figure G2: Distribution of Health Share to Poorest by Causal Beliefs
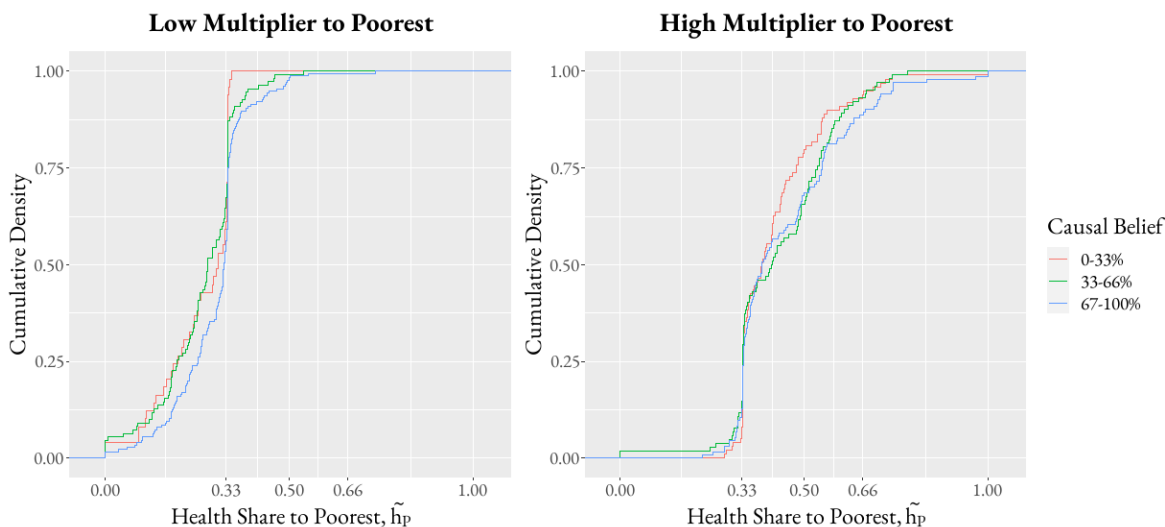
Note: Empirical cumulative density functions of health share to poorest individual in rounds of Treatment B by intervals of beliefs about percentage of multiplier differences caused by income. In all rounds, multipliers are 1, 0.5, and 0.33. Left (right) panel uses rounds with positive (negative) association between multipliers and income. n=335 in each panel.

Generally, participants who believe that a larger percentage of the differences in the multipliers is caused by income allocate resources to ensure that the poorest individual gets a larger share of health. This is true irrespective of whether the distribution of the multipliers favours the rich or the poor, although differences occur at the bottom of the distribution in the first case and toward the top in the second.

Regressing the health share to the poorest on the causal belief variables reveals that the share has the strongest and most significant association with beliefs when the multipliers increase with income (pro-rich). In column (1) of Table G2 we use rounds of Treatment B in which the multipliers are $\{0.33, 0.5, 1\}$ and increase monotonically with income. In this case, we estimate that a 100 point increase in the percentage of the difference in the multipliers believed to be caused by income is significantly associated with a 5.5 percentage

point increase in the health share to the poorest. In column (2), we use rounds in which the multipliers decrease with income (pro-poor). The estimated association between the health share to the poorest and the causal belief is still positive, but it is smaller and not significant. In column (3), we pool all 10 rounds and regress the health share to the poorest on the Pro-Rich and Pro-Poor Causal Beliefs. This reveals that the health share to the poorest is positively and significantly associated with the Pro-Rich Causal Belief, but not significantly associated with the Pro-Poor Causal Belief. Although these beliefs are not exogenously manipulated, these results are suggestive that a stronger belief in a pro-rich causal relationship between income and health leads participants to prioritise the poor. Further research could try to disentangle the longer run connection between beliefs and social preference formation.

Table G2: Association between health share to poorest and causal beliefs

	(1) Low p_i to Poor Coef./(S.E.)	(2) High p_i to Poor Coef./(S.E.)	(3) All Rounds Coef./(S.E.)
Pro-Rich Causal Belief	0.0546**		0.0375***
	(0.023)		(0.014)
Pro-Poor Causal Belief		0.0219	0.0064
		(0.024)	(0.011)
Constant	0.2460***	0.4280***	0.3241***
	(0.015)	(0.014)	(0.007)
Participants	335	335	335
Observations	335	335	3350
R^2	0.0194	0.0024	0.0076

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Notes: Each column gives estimates from OLS regression of health share to the poorest on Pro-Rich and/or Pro-Rich Causal Beliefs. Columns (1) and (2) use subsets of rounds with multipliers monotonically increasing (Pro-Rich) and decreasing (Pro-Poor), respectively, with income. Column (3) uses all 10 rounds. Robust standard errors (SE) in parentheses. p-values: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.