Journal of Health Economics 94 (2024) 102856

Contents lists available at ScienceDirect

Journal of Health Economics

journal homepage: www.elsevier.com/locate/jhe

Aversion to health inequality — Pure, income-related and income-caused

Matthew Robson [a,b,d,*], Owen O'Donnell [a,b,c,d], Tom Van Ourti [a,b,c,d,1]

[a] Erasmus School of Economics, Erasmus University Rotterdam, The Netherlands
[b] Erasmus Centre for Health Economics Rotterdam, Erasmus University Rotterdam, The Netherlands
[c] Erasmus School of Health Policy and Management, Erasmus University Rotterdam, The Netherlands
[d] Tinbergen Institute, The Netherlands

A R T I C L E I N F O

Dataset link: https://data.mendeley.com/datasets/9vy6f6g5k3

JEL classification:
C90
D30
D63
I14

Keywords:
Inequality aversion
Social preferences
Health
Income
Experiment

A B S T R A C T

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 information on income and its causal effect. We gather data (26,286 observations) from a sample of UK adults (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.

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 (Wagstaff, 1991; Bleichrodt et al., 2004; Dolan et al., 2005; Richardson et al., 2012). Concern about systematic differences in health between individuals distinguished by income motivates prioritisation of health gains to the poor (Wagstaff et al., 1991; Bleichrodt and van Doorslaer, 2006). 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) (Dolan and Tsuchiya, 2009; Robson et al., 2017; Hurley et al., 2020; McNamara et al., 2021). This risks confounding one type of aversion with the other and so biasing distributionally sensitive evaluation of health

* Corresponding author at: Erasmus School of Economics, Erasmus University Rotterdam, The Netherlands.
E-mail addresses: robson@ese.eur.nl (M. Robson), odonnell@ese.eur.nl (O. O'Donnell), vanourti@ese.eur.nl (T. Van Ourti).
URL: https://www.mrobson92.com (M. Robson).

[1] This study was funded by the Erasmus Initiative Smarter Choices for Better Health, The Netherlands and the Wellcome Trust, UK (205427/Z/16/Z). Ethics approval given by the Internal Review Boards of the Erasmus Research Institute of Management and Erasmus School of Economics. We would like to thank Matt Adler, Richard Cookson, Tim Doran, Guido Erreygers, Paul Makdissi, Erik Schokkaert, Aki Tsuchiya, two anonymous reviewers, and participants at various seminars and conferences for helpful comments. A data repository including experimental data and code, statistical code, preference parameter estimates and policy evaluation code is available at https://data.mendeley.com/datasets/9vy6f6g5k3.

https://doi.org/10.1016/j.jhealeco.2024.102856
Received 12 April 2023; Received in revised form 22 December 2023; Accepted 2 January 2024
Available online 23 January 2024
0167-6296/

programmes that use the elicited preferences (Cookson et al., 2021). 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 (Arrow, 1971), 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 inequality in the univariate distribution of health — pure health inequality (Wagstaff and van Doorslaer, 2000). 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 prioritisation of health by income and aversion to pure health inequality that jointly determine aversion to income-related 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 (Atkinson and Bourguignon, 1982). This may motivate prioritisation of poorer individuals in the distribution of health resources in order to compensate material disadvantage with improved health.

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 left in worse health due to insufficient income to 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 sample that is representative of the UK adult population with respect to sex, and, to an extent, age and ethnicity. 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 et al., 2007; Hurley et al., 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 willingness 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 et al., 1991; Wagstaff, 2002; O'Donnell et al., 2008). Consequently, the model (Makdissi and Yazbeck, 2016) also encompasses Wagstaff's (2002) achievement index that penalises mean health for pro-rich health inequality without allowing for aversion to pure health inequality. 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 et al., 2016; Hurley et al., 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 et al., 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 et al., 2020; McNamara et al., 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, 2011; Robson et al., 2017; Cookson et al., 2018; Hardardottir et al., 2021; McNamara et al., 2021). This risks the mistaken inference of preferences from choices that are simply random errors. These studies often drop a substantial proportion of responses that appear irrational. Some studies allow for noise, but only when pooling data across participants (Edlin et al., 2012; Hurley et al., 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 were intended to be representative of the UK adult population in terms of sex, ethnicity, and age.

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 within a set of values chosen to facilitate ease of comprehension by participants, while providing sufficient variation to produce equity-efficiency trade-offs.

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 budget was randomly selected from a range of values that was set to ensure that the product of resources and the multipliers

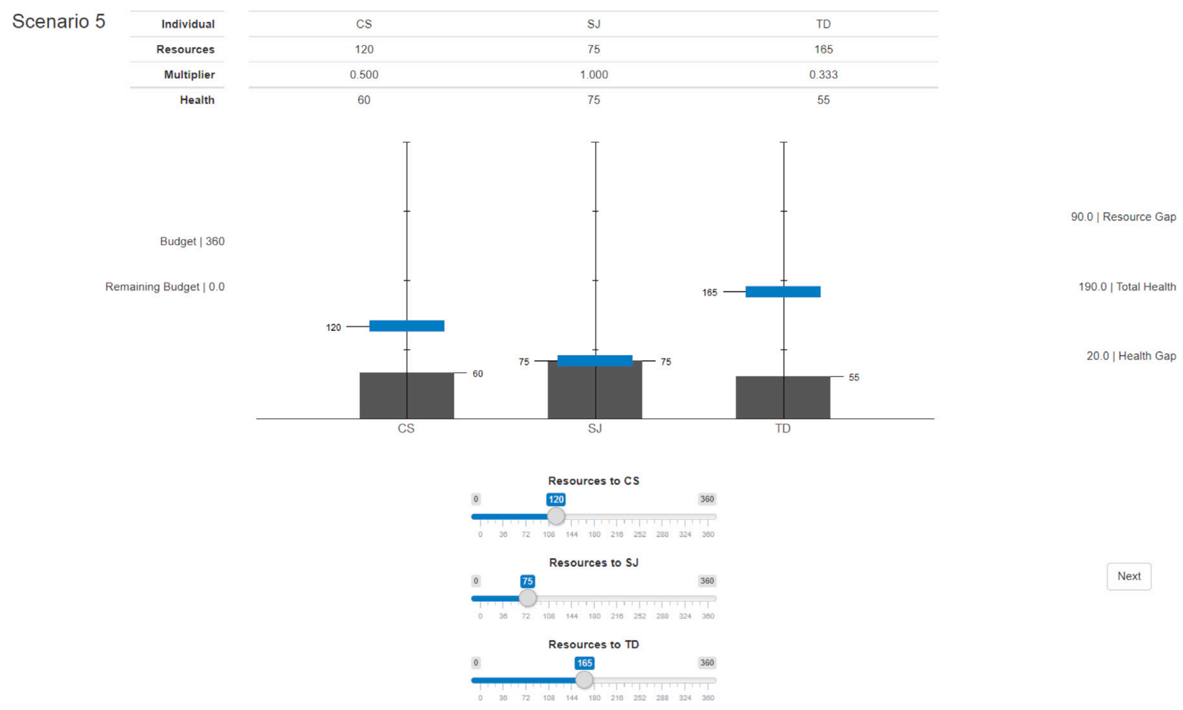

Fig. 1. Experiment Interface.

would produce plausible QALYs. For example, if the mean budget were evenly divided between three individuals all with the mean multiplier, then each would enjoy around 59 QALYs.

The task was completed using an online screen interface designed in R Shiny that is shown in Fig. 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. Timers were set to ensure that a participant took at least 30 s to complete each round.

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 labelled by randomly drawn initials (e.g. *CS*, *SJ*, and *TD*). We did not instruct the participants to consider the individuals to be identical, except for any difference in multipliers, because the purpose of this treatment is to estimate aversion to inequality in the marginal distribution of health, not in any conditional distribution. The multipliers (p_i) changed across 10 rounds (see Appendix A.3 Table A1). Between round differences in p_i force trade-offs between health maximisation and equalisation that are used to identify aversion to health inequality. Participants may vary in beliefs about possible sources of these differences, and these beliefs may affect their inequality aversion. 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. All participants completed the same 10 rounds of multiplier combinations but the order of these rounds was randomised across participants. 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. This information is used to identify prioritisation of health by income, which could depend on a participant's beliefs and ethical judgements about possible sources of income differences and their consequences for health. Each participant again allocated resources in 10 rounds of multiplier combinations with the order randomised across participants. 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 $\{£5000, £10,000, £25,000, £50,000, £100,000\}$ and used them to label the three individuals, e.g. *HD: £10,000*.[2] 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 of each of the three individuals but also the extent to which income differences cause differences in multipliers. We presented each participant with two scenarios that differ in the direction of association between incomes and the same set of multipliers, $\{0.33, 0.5, 1\}$.[3] In one scenario, the 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 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.[4] 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 min in the first session and 27.6 min in the second.[5] See Appendix A.1 for a graphical overview of the experiment.

On the Prolific platform, eligible subscribers can sign up to participate in a study until the target sample size is achieved. Hence, there is no response rate. Out of 426 subscribers who started the first session, 24 (5.6%) did not complete the session and so were dropped from the sample. Of the remaining 402 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 representative with respect to sex, but is younger than the UK adult population and has a lower proportion of participants who are white (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.

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),

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

[2] 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 facilitate ease of comprehension by participants, to give sufficient variation for identification and to cover the 10th to 99th percentiles of the UK personal income distribution.

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

[4] 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.

[5] 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.

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.

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 et al., 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, $U(h_i) = (h_i^{1-\varepsilon} - 1)/(1-\varepsilon)$ for $\varepsilon \geq 0$ and $\varepsilon \neq 1$, and $U(h_i) = ln(h_i)$ for $\varepsilon = 1$ (Atkinson, 1970). Welfare can be measured by the equally distributed equivalent (EDE) level of health,

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

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).[6] 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.[7] 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.[8]

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, $h_i = p_i \times y_i$.[9] Maximisation of social welfare (h_{EDE}) subject to a binding budget constraint, $\sum_{i=1}^{N} y_i = m$, gives the optimal allocations,

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

[6] We assume that health does not directly determine the weights. This distinguishes Eq. (1) from the (nonlinear) rank-dependent QALY model (Bleichrodt et al., 2004, 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.

[7] 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.

[8] 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.

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

3.2. Equity weights

Specification of the weights distinguishes aversion to pure health inequality from aversion to income-related health inequality and the latter from aversion to income-caused health inequality. Treatments A, B, and C, allow specification of constant weights, ω_i^A, income-dependent weights, ω_i^B, and causality-dependent weights, ω_i^C, respectively.

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, $x_{i-1} \le x_i \ \forall_i$. 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) indices to measure income-related health inequality (Wagstaff et al., 1991; O'Donnell et al., 2008). We specify the weights in a way that is standard for these indices (Donaldson and Weymark, 1980, 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.[10] 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 et al., 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 et al., 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.[11]

The model of social preferences given by Eqs. (2) and (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) for aversion to income-related health inequality through the extended concentration index but with no aversion to pure health inequality ($\varepsilon = 0$).[12]

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.[13] The sign of λ indicates the perceived direction of the causal effect that income may be believed to have on health. The magnitude of this parameter is the perceived proportion of health differences that are believed to be 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 health is not causally affected by income. 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.

---

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

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

[12] Wagstaff's (2002) achievement index is $A(\beta) = (1 - C(\beta))\,\bar{h}$, where $C(\beta)$ is the extended concentration index and $\bar{h}$ is mean health.

[13] 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.[14]

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 ($\boldsymbol{I}$) 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 $\lambda = \boldsymbol{I}\left(\mathbb{1}\left(Cov\left(p_i, x_i\right) > 0\right) - \mathbb{1}\left(Cov\left(p_i, x_i\right) < 0\right)\right)/100$. 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.

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,[15]

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

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

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

3.3.2. Absolute invariance

Eq. (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 $U\left(h_i\right) = -exp\left(\theta h_i\right), \theta < 0$ in Eq. (1). The respective EDE health is

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

The allocation of resources that maximises this indicator of welfare is

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

The non-consequentialist SDM who is concerned about absolute resource inequality would make an optimal allocation that is given by Eq. (10) with multipliers $\left(p_i, p_j\right)$ 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 et al., 1991; O'Donnell et al., 2008), it does not allow for the possibility that 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 normalise them to sum to 1[16]:

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

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

[15] 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.

[16] These weights can be derived from Eq. (5) by replacing income ranks $\left(r(x_i)\right)$ with income shares $\left(\tilde{x}_i\right)$, 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 et al., 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.

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.

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).[17] The α_i parameters determine the shape of the distribution. Here, they capture the relative weight given to each individual.

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,[18]

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

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 ($\sum_{k=1}^K LL_k$), and estimating one set of parameters that capture the preferences and weights of a representative SDM.

5. Results

5.1. Non-parametric analyses

5.1.1. Equity-efficiency trade-off

Using data from Treatment A, Fig. 2 plots cumulative density functions (CDFs) of resource shares ($\tilde{y}_i$) and health shares ($\tilde{h}_i$) conditional on relative multipliers ($\tilde{p}_i = p_i/\sum p_i$).[19] 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.

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

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

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

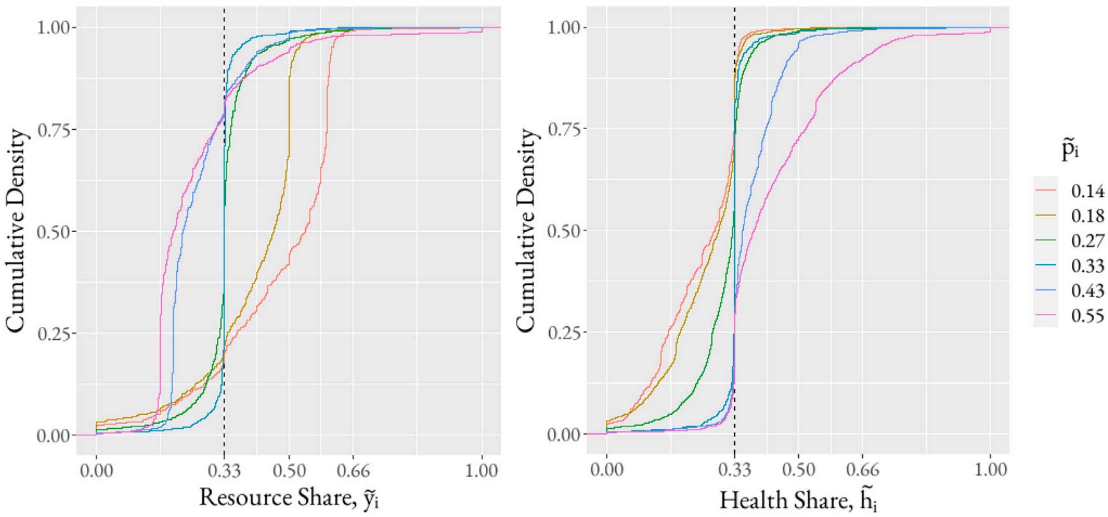

Fig. 2. Distributions of resource and health shares by relative multipliers.
Note: Empirical cumulative density functions of resource shares ($\tilde{y}_i = y_i / \sum y_i$) and health shares ($\tilde{h}_i = h_i / \sum h_i$) to individuals distinguished by relative multipliers ($\tilde{p}_i = p_i / \sum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

The left panel shows that individuals who are more productive than average $\left(\tilde{p}_i > 1/3\right)$ are given less than an equal (1/3) share of resources in more than three quarters (78.5%) of the allocations. Those who are least productive $\left(\tilde{p}_i \leq 0.18\right)$ are given more than an equal resource share in about three quarters of allocations. These results reveal that most participants are inequality averse — they sacrifice efficiency in the production of health for less inequality in its distribution by allocating less resources to individuals who would gain most health from them. Some participants, at least some of the time, do not display this inequality aversion. Individuals who are more productive than average are given more than an equal share of resources in about 17% of the allocations. But very few participants maximise the production of health by not allocating any resources to individuals who are less productive than average.

The right panel shows that inequality aversion is, in most cases, insufficient to eliminate inequality in the distribution of health. Mostly, the less productive individuals get a less than an equal (1/3) share of health, and the more productive get a greater share. However, in more than one third (34.6%) of allocations, the inequality aversion is sufficient to leave above average productivity individuals with an equal or smaller share of health. And almost half of the allocations result in an equal or greater share of health to individuals who are slightly less productive than average $\left(\tilde{p}_i = 0.27\right)$.

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 in social preferences (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.[20]

5.1.2. Prioritisation by income

Using data from all choices in Treatment B, the left panel of Fig. 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).[21]

The right panel of Fig. 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.

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

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

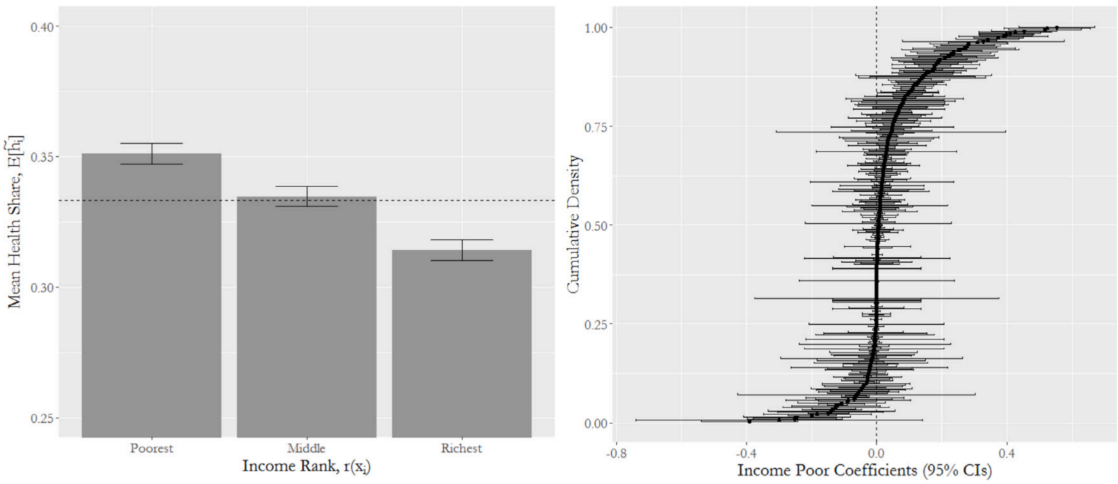

Fig. 3. Health shares by income - pooled (left) and heterogeneous (right).
Note: Left panel shows mean of health share ($\tilde{h}_i$) by within round income rank of individual. In this panel, data are pooled and averaged over all participants (n = 337) and rounds (10) in Treatment B (n = 10,110). Right 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.

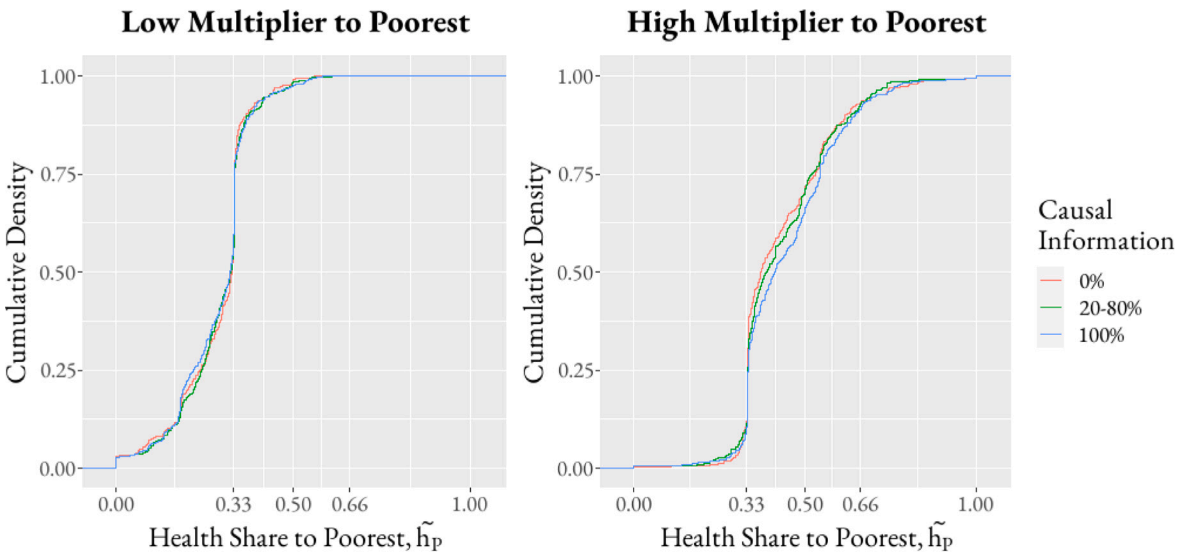

Fig. 4. Distribution of health share to poorest by causal information.
Note: CDFs of health share to poorest individual within each round ($\tilde{h}_p$, $p = i$ with $x_i < x_j \forall j \neq i$) stratified by exogenously varying values of causal information on percentage of multiplier differences caused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

5.1.3. Sensitivity to causality

Using the data from Treatment C, Fig. 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.

Within 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).

Table 2
Pooled estimates of social welfare function parameters.

Treatment	Health inequality aversion ε	Income weight β (β_1 in C)	Causal income weight β_2	Precision σ
A	1.391 [1.323–1.470]			8.047 [7.354–8.910]
B	1.563 [1.474–1.663]	1.105 [1.075–1.136]		8.677 [7.938–9.599]
C	1.670 [1.477–1.793]	1.102 [1.060–1.147]	0.956 [0.892–1.048]	9.044 [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.

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.[22] 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 population of three a weight that is 21.5% larger than the weight given to the richest individual.[23] 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 Eqs. (2) and (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.[24]

We use random variation in beliefs generated by the causal information (I) provided in Treatment C to set values of λ in Eq. (6) and so identify β_2 separately from β_1. The point estimate of β_2 obtained from the Treatment C data is very close to 1 and is not significantly different from this value. This implies that, on average, the weight given to poorer individuals 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.[25]

5.3. Heterogeneous estimates of preference parameters

Table 3 shows percentiles of the distribution of participant-specific estimates of each preference parameter.[26] 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.[27] 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 Fig. 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

[22] The welfare loss generated by inequality is the difference between mean health and EDE health (Eq. (2)). 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.

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

[24] 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).

[25] We confirm robustness to an alternative identification strategy that uses elicited beliefs to set values of λ. 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.

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

[27] For example, using data from Treatment A and the 3-person scenario from fn. 22, 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).

Fig. 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.

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 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: $\hat{\varepsilon} \geq 15$.[28]

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 Fig. 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: $\hat{\beta} < 0.95$.

For a majority of the sample, estimates of ε and β are greater than 0 and 1, respectively, which implies rejection of restrictions on the general model (Eqs. (2) and (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 Fig. 6, which plots participant-specific estimates (from Treatment B) of the optimal health share to the poorest individual ($\tilde{h}_p^*$) against $\hat{\beta}$. Symbols distinguish between participants with Efficiency Seeking, Cobb–Douglas, Prioritarian, and Maximin preferences

[28] 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).

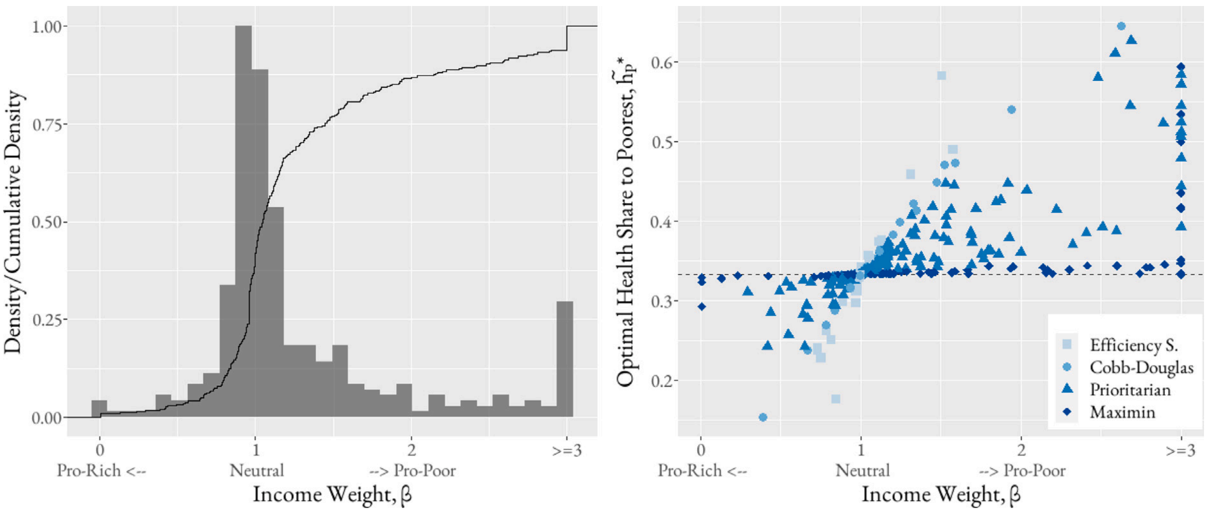

Fig. 6. Distribution of income weight parameter estimates, $\hat{\beta}$.

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 $\hat{\beta} > 3$ are 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 Efficiency Seeking, Cobb–Douglas, Prioritarian, and Maximin preferences defined by $\hat{\varepsilon} \leq 0.9$, $0.9 < \hat{\varepsilon} \leq 1.1$, $1.1 < \hat{\varepsilon} < 15$, and $\hat{\varepsilon} \geq 15$, respectively.

based on their $\hat{\varepsilon}$. We consider a case with constant multipliers to ensure that there is no equity-efficiency trade-off. Hence, with income-neutral weights ($\hat{\beta} \approx 1$), the poorest individual optimally gets an equal share of health ($\tilde{h}_p^* \approx 1/3$). As income weights become more pro-poor – that is, $\hat{\beta}$ increases above 1 – the optimal share to the poorest rises above $1/3$, but 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 so their optimal allocations to the poor rise steeply with more pro-poor income weights, given that there is no efficiency cost.

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).

Participants with larger estimates of ε obtained from one treatment tend to have larger estimates of the same parameter obtained from another treatment (Appendix D.5 Table D5). This strong and significant between-treatment rank correlation, which is also observed for estimates of σ, lends face validity to the analysis. There is also a strong, positive rank correlation between estimates of β from Treatment B and β_1 from Treatment C (Table D5). Interestingly, estimates of ε from Treatment A are positively and significantly rank correlated with estimates of β from Treatment B, while within Treatment B the rank 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 rank 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.

With multiple rounds within each treatment, there is potential for order effects. The analyses reported in Appendix D.6 produce no evidence of such effects, which may be due to the steps taken in the design to reduce the scope for them.

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.

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

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.

SWF is strongly preferred to the non-consequentialist SWF and is slightly preferred to the Kolm SWF (see 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 ($\theta = 0$).[29]

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 ($\beta > 1$).[30]

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 adult sample can be used in policy evaluation. They can be applied to an estimated policy-specific distribution of health over individuals or groups, which may be ordered by income, to simulate the distribution of support for the respective policy.[31]

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.[32] We evaluate two policies that have the same cost and do not change the distribution of income.

[29] 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.

[30] 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].

[31] Appendix F explains how to access and use our estimates, at https://data.mendeley.com/datasets/9vy6f6g5k3. 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.

[32] 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 $h_i = 45 + 2log(x_i) + \upsilon_i$, where $\upsilon_i \sim N(0, 3)$.

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 (median)	68.21	68.17	A	51.3%
With Income Weights				
Pooled	68.31	68.16	A	
Heterogeneous (median)	68.03	67.14	A	70.9%

Note: See footnotes 32 and 33 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 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.

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.[33]

The top panel of Table 5 shows, for each policy, the mean and standard deviation of health and its correlation with income.[34] 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 sample, Policy A gives the larger EDE health and so this policy would be chosen under simple majority voting.

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 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 sample with demographics similar to those of the UK adult 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 et al., 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).

[33] 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.

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

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 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 et al., 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 et al., 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 et al., 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.

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 et al., 2021; Hurley et al., 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 et al., 2012). We are not the first to estimate that a sizeable proportion of a population would prioritise the health of richer people. Hardardottir et al. (2021) find that slightly more than one quarter of a representative Swedish sample displays a pro-rich bias, while Hurley et al. (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 et al., 2013), although the evidence is mixed (De Nardi et al., 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, prioritisation 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 explanation would be 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). It is also possible that muted responses to the causal information (see Appendix D.3) are due to the difficulty many participants may have had in processing it. This is a cognitively demanding task. Refining it in future research would be worthwhile in order to evaluate the often-posed but seldom-tested hypothesis that aversion to health inequality depends on its causes.

Our approach rests on an assumption that preferences elicited from allocations to three hypothetical individuals are reasonably good approximations to preferences over the distribution of health in a whole population. There can be legitimate concerns about the validity of this assumption. For example, someone may observe that the richest of three individuals is in the best health and attribute it to chance, but they must accept that an association between health and income in a population is systematic. This could cause downward bias in our estimates of aversion to income-related health inequality. On the other hand, by asking participants to choose between three individuals, we avoid the *psychic numbing* (Slovic, 2007) that tends to cloud perceptions of injustice that is suffered by multitudes. This psychological bias could suppress aversion to inequality if that inequality were presented as health deficits experienced by many poor people. The number of individuals that experiment participants can be asked to compare is constrained by cognitive capacity. Subject to this constraint, it would be interesting in future research to assess the direction and magnitude of any bias in the estimation of health inequality aversion that is related to the size of an experiment's hypothetical population.

Most studies have elicited aversion to health inequality by asking participants to choose between groups (Dolan and Tsuchiya, 2011; Robson et al., 2017; Pinho and Botelho, 2018; Hurley et al., 2020; McNamara et al., 2021). This avoids extrapolation from choices made between a few individuals. We deviate in order to distinguish aversion to pure health inequality over individuals who can differ in multiple characteristics from aversion to health inequality between individuals who are known to differ by income. It would be difficult to do this with a group framing because reference to a group begs the question of what characterises it. Distinguishing groups by some characteristic, such as disease (McNamara et al., 2021), would make the aversion specific to inequality by that characteristic. Asking participants to contemplate groups that are identical in all respects other than health (Pinho and Botelho, 2018; Hurley et al., 2020) may leave them puzzled about how such health differences could arise. Logically, they would have to attribute the differences to chance arising from social risk. But aversion to this type of health inequality is not the parameter of interest. Decision makers do not allocate resources in order to change the distribution of health over otherwise identical groups, or individuals. They respond to health differences that can have multiple correlates and causes. Our design, including comparison between individuals rather than groups, allows identification of the extent to which inequality aversion is contingent on knowledge of those correlates and causes.

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.

CRediT authorship contribution statement

Matthew Robson: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Owen O'Donnell:** Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing. **Tom Van Ourti:** Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

A data repository including experimental data and code, statistical code, preference parameter estimates, and policy evaluation code is available at: https://data.mendeley.com/datasets/9vy6f6g5k3.

Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jhealeco.2024.102856.

References

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

Abellan-Perpiñan, J., Pinto-Prades, J., 1999. Health state after treatment: A reason for discrimination? Health Econ. 8 (8), 701–707. http://dx.doi.org/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. Q. J. Econ. 85 (3), 409–415. http://dx.doi.org/10.2307/1885930.

Atkinson, A.B., 1970. On the measurement of inequality. J. Econom. Theory 2 (3), 244–263. http://dx.doi.org/10.1016/0022-0531(70)90039-6.

Atkinson, A.B., Bourguignon, F., 1982. The comparison of multi-dimensioned distributions of economic status. Rev. Econom. Stud. 49 (2), 183–201. http://dx.doi.org/10.2307/2297269.

Bleichrodt, H., 1997. Health utility indices and equity considerations. J. Health Econ. 16 (1), 65–91. http://dx.doi.org/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. J. Health Econ. 23 (1), 157–171. http://dx.doi.org/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. J. Health Econ. 24 (4), 655–678. http://dx.doi.org/10.1016/j.jhealeco.2004.10.001.

Bleichrodt, H., van Doorslaer, E., 2006. A welfare economics foundation for health inequality measurement. J. Health Econ. 25 (5), 945–957. http://dx.doi.org/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. Amer. Econ. Rev. 97 (3), 818–827. http://dx.doi.org/10.1257/aer.97.3.818.

Conte, A., Moffatt, P.G., 2014. The econometric modelling of social preferences. Theory and Decision 76, 119–145. http://dx.doi.org/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 Econ. 27 (11), 1754–1771. http://dx.doi.org/10.1002/hec.3799.

Cookson, R., Griffin, S., Norheim, O.F., Culyer, A.J., 2021. Distributional Cost-Effectiveness Analysis. Oxford University Press, Oxford.

Cropper, M., Krupnick, A., Raich, W., 2016. Preferences for equality in environmental outcomes. Working Paper 25447, National Bureau of Economic Research, http://dx.doi.org/10.3386/w22644.

De Nardi, M., French, E., Jones, J.B., 2010. Why do the elderly save? The role of medical expenses. J. Polit. Econ. 118 (1), 39–75. http://dx.doi.org/10.1086/651674.

Dirichlet, P.G.L., 1839. Sur une nouvelle méthode pour la détermination des intégrales multiples. J. Math. Pures Appl. (9) 4, 164–168.

Dolan, P., 1998. The measurement of individual utility and social welfare. J. Health Econ. 17 (1), 39–52. http://dx.doi.org/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 Econ. 14 (2), 197–208. http://dx.doi.org/10.1002/hec.924.

Dolan, P., Tsuchiya, A., 2009. The social welfare function and individual responsibility: Some theoretical issues and empirical evidence. J. Health Econ. 28 (1), 210–220. http://dx.doi.org/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. Appl. Econ. 43 (18), 2241–2250. http://dx.doi.org/10.1080/00036840903166244.

Donaldson, D., Weymark, J.A., 1980. A single-parameter generalization of the Gini indices of inequality. J. Econom. Theory 22 (1), 67–86. http://dx.doi.org/10.1016/0022-0531(80)90065-4.

Donaldson, D., Weymark, J.A., 1983. Ethically flexible Gini indices for income distributions in the continuum. J. Econom. Theory 29 (2), 353–358. http://dx.doi.org/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 Econ. 21 (12), 1416–1426. http://dx.doi.org/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. J. Health Econ. 31 (1), 257–270. http://dx.doi.org/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. Int. J. Environ. Res. Public Health 14 (7), http://dx.doi.org/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. J. Eur. Econom. Assoc. 11 (1), 221–258. http://dx.doi.org/10.1111/j.1542-4774.2012.01101.x.
Fisman, R., Kariv, S., Markovits, D., 2007. Individual preferences for giving. Amer. Econ. Rev. 97 (5), 1858–1876. http://dx.doi.org/10.1257/aer.97.5.1858.
Fleming, M., 1952. A cardinal concept of welfare. Q. J. Econ. 66 (3), 366–384. http://dx.doi.org/10.2307/1885309.
Gyrd-Hansen, D., 2004. Investigating the social value of health changes. J. Health Econ. 23 (6), 1101–1116. http://dx.doi.org/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 Econ. 30 (10), 2531–2546. http://dx.doi.org/10.1002/hec.4395.
Harless, D.W., Camerer, C.F., 1994. The predictive utility of generalized expected utility theories. Econometrica 1251–1289. http://dx.doi.org/10.2307/2951749.
Harsanyi, J.C., 1955. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. J. Polit. Econ. 63 (4), 309–321. http://dx.doi.org/10.1086/257678.
Hurley, J., Mentzakis, E., Walli-Attaei, M., 2020. Inequality aversion in income, health, and income-related health. J. Health Econ. 102276. http://dx.doi.org/10.1016/j.jhealeco.2019.102276.
Kolm, S.-C., 1976. Unequal inequalities. I. J. Econom. Theory 12 (3), 416–442. http://dx.doi.org/10.1016/0022-0531(76)90037-5.
Makdissi, P., Yazbeck, M., 2016. Avoiding blindness to health status in health achievement and health inequality measurement. Soc. Sci. Med. 171 (1), 39–47. http://dx.doi.org/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. Eur. J. Health Econ. http://dx.doi.org/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. Soc. Sci. Med. 269, 113573. http://dx.doi.org/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. World Bank.
Parfit, D., 2000. In: Clayton, M., Williams, A. (Eds.), The Ideal of Equality. Palgrave MacMillian, pp. 81–125, Ch. Equality or Priority?.
Pinho, M., Botelho, A., 2018. Inference procedures to quantify the efficiency–equality trade-off in health from stated preferences: A case study in Portugal. Appl. Health Econ. Health Policy 16, 503–513. http://dx.doi.org/10.1007/s40258-018-0394-6.
Pollak, R.A., 1971. Additive utility functions and linear engel curves. Rev. Econom. Stud. 38 (4), 401–414. http://dx.doi.org/10.2307/2296686.
Ratcliffe, J., 2000. Public preferences for the allocation of donor liver grafts for transplantation. Health Econ. 9 (2), 137–148. http://dx.doi.org/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. Soc. Sci. Med. 75 (8), 1351–1361. http://dx.doi.org/10.1016/j.socscimed.2012.05.036.
Robson, M., 2021. Inequality aversion, self-interest and social connectedness. J. Econ. Behav. Organ. 183, 744–772. http://dx.doi.org/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 Econ. 26 (10), 1328–1334. http://dx.doi.org/10.1002/hec.3430.
Roemer, J., 2002. Equality of opportunity: A progress report. Soc. Choice Welf. 19, 455–471. http://dx.doi.org/10.1007/s003550100123.
Schokkaert, E., Devooght, K., 2003. Responsibility-sensitive fair compensation in different cultures. Soc. Choice Welf. 21, 207–242. http://dx.doi.org/10.1007/s00355-003-0257-3.
Slovic, P., 2007. "If I look at the mass I will never act": Psychic numbing and genocide. Judgm. Decis. Mak. 2 (2), 79–95. http://dx.doi.org/10.1017/S1930297500000061.
Ubel, P.A., Loewenstein, G., 1996. Distributing scarce livers: The moral reasoning of the general public. Soc. Sci. Med. 42 (7), 1049–1055. http://dx.doi.org/10.1016/0277-9536(95)00216-2.
Vickrey, W., 1960. Utility, strategy, and social decision rules. Q. J. Econ. 74 (4), 507–535. http://dx.doi.org/10.2307/1884349.
Wagstaff, A., 1991. QALYs and the equity-efficiency trade-off. J. Health Econ, 10 (1), 21–41. http://dx.doi.org/10.1016/0167-6296(91)90015-F.
Wagstaff, A., 2002. Inequality aversion, health inequalities and health achievement. J. Health Econ. 21 (4), 627–641. http://dx.doi.org/10.1016/S0167-6296(02)00006-1.
Wagstaff, A., Paci, P., Van Doorslaer, E., 1991. On the measurement of inequalities in health. Soc. Sci. Med. 33 (5), 545–557. http://dx.doi.org/10.1016/0277-9536(91)90212-U.
Wagstaff, A., van Doorslaer, E., 2000. Chapter 34 Equity in health care finance and delivery. In: Handbook of Health Economics, vol. 1, Elsevier, pp. 1803–1862. http://dx.doi.org/10.1016/S1574-0064(00)80047-5.
Yitzhaki, S., 1983. On an extension of the Gini inequality index. Internat. Econom. Rev. 24 (3), 617–628. http://dx.doi.org/10.2307/2526814.