



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/245374/>

Version: Accepted Version

Article:

Trigg, L.A., Moolla, A., Angus, C. et al. (2026) A systematic review of elicitation methods for distributional preferences in healthcare regarding the concentration and dispersion of health benefits. *Value in Health*, 29 (6). pp. 961-968. ISSN: 1098-3015

<https://doi.org/10.1016/j.jval.2026.03.012>

© 2026 The Authors. Except as otherwise noted, this author-accepted version of a journal article published in *Value in Health* is made available via the University of Sheffield Research Publications and Copyright Policy under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Title: A Systematic Review of Elicitation Methods for Distributional Preferences in healthcare regarding the concentration and dispersion of health benefits

Authors: Laura A. Trigg MSc¹, Aisha Moolla MPH MBChB¹, Colin Angus MSc¹, Aki Tsuchiya PhD^{1,2}.

1 School of Medicine and Population Health, University of Sheffield, United Kingdom

2 School of Economics, University of Sheffield, United Kingdom

Contact details for corresponding author

Laura A. Trigg Msc

University of Sheffield

Regent Court

30 Regent Street

Sheffield

S1 4DA

Ltrigg1@sheffield.ac.uk

Précis:

This review explores and evaluates methods for eliciting preferences on concentrating or dispersing health benefits, highlighting implications for equity in health policy decision-making.

Authorship Confirmation:

All authors certify that they meet the ICMJE criteria for authorship.

Funding/Support:

This research was funded in whole, or in part, by the Wellcome Trust [Grant number 218462/Z/19/Z]. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Role of the Funder/Sponsor:

The funder had no role in the design and conduct of the study; collection, management, analysis and interpretation of the data; preparation, review or approval of the manuscript; and decision to submit the manuscript for publication.

A Systematic Literature Review of Elicitation Methods for Distributional Preferences in Healthcare Regarding the Concentration and Dispersion of Health Benefits

Abstract

Background

In conventional QALY maximisation, each unit of health is valued equally. It overlooks distributional preferences such as whether health improvements should be *concentrated* on a smaller number of individuals or *dispersed* across a larger population. It is plausible that these preferences follow an S-shaped pattern, where marginal social value of health gains initially increases before diminishing after a threshold. This review aims to identify and evaluate the methods used in empirical studies to elicit S-shaped preferences.

Methods

Searches were conducted in PubMed, PsycINFO, and EconLit in August 2024 and October 2025, with one round of citation chasing for four key papers. Studies were included if they elicited preferences on concentration and dispersion of health gains among otherwise identical individuals. Data synthesis was conducted narratively.

Results

Thirteen studies published between 1994-2025 were included, collectively involving 6,799 participants. All studies used a Person Trade-Off (PTO) approach, but two also used Time Trade-off. Five key methodological themes emerged: whether health gains should be measured in life years or health related quality of life; the threshold between concentration and dispersion; perspective; framing effects; and specification of the social welfare functions.

Conclusions

This review highlights methodological challenges associated with eliciting preferences for concentration and dispersion. Though there is a consensus on the use of PTO, other heterogeneity such as pool size or perspective, and sparse evidence on the threshold effect warrants further research.

Highlights

- 1.) This review addresses the evidence gap in evaluating methods used to elicit S-shaped distributional preferences with regarding health gains by systematically identifying and assessing the existing literature.
- 2.) Person Trade-Off (PTO) tasks are the most prevalent method, but studies vary considerably in implementation, including in framing and presentation of health states, meaning results are often incomparable. Preferences switching between concentration and dispersion was explored, though the point of this shift remains imprecisely quantified and warrants further empirical research.
- 3.) This review provides insight into strengths and weaknesses of previous empirical studies to provide guidance on the design of future preference elicitation work.

Introduction

Cost-Effectiveness Analysis (CEA) assumes every Quality Adjusted Life Year (QALY) is of equal social value regardless of how it is distributed¹; total welfare is an additive sum of QALYs. However, evidence suggests that when considering public preferences for the distribution of health gains, the use of an unweighted sum of QALYs does not accurately reflect social welfare²⁻⁵. Societal preferences often reflect inequality aversion, favouring a more equal distribution of health gains across individuals. This implies a concave health related social welfare function, where marginal value of additional health decreases as individual health increases. However, when per-person health-gain is small, efficiency considerations may take precedence over equality; with preferences to concentrate resources among fewer individuals (convex function), ensuring each receives a meaningful gain. In England, NICE have introduced modifiers to account for severity⁶, but there is no consideration into this type of distributional preference which could have implications for high-cost innovative technologies (e.g. TA588) where individual QALY gain is high, and preferences fall within the concave section of the objective function with diminishing marginal value. This review introduces the term ‘S-shaped preferences’ to describe this pattern of initially increasing and then diminishing marginal social value of health gain, as seen in figure 1. Prior studies have described and empirically estimated concentration and dispersion relationships, but terminology has not been formalised.

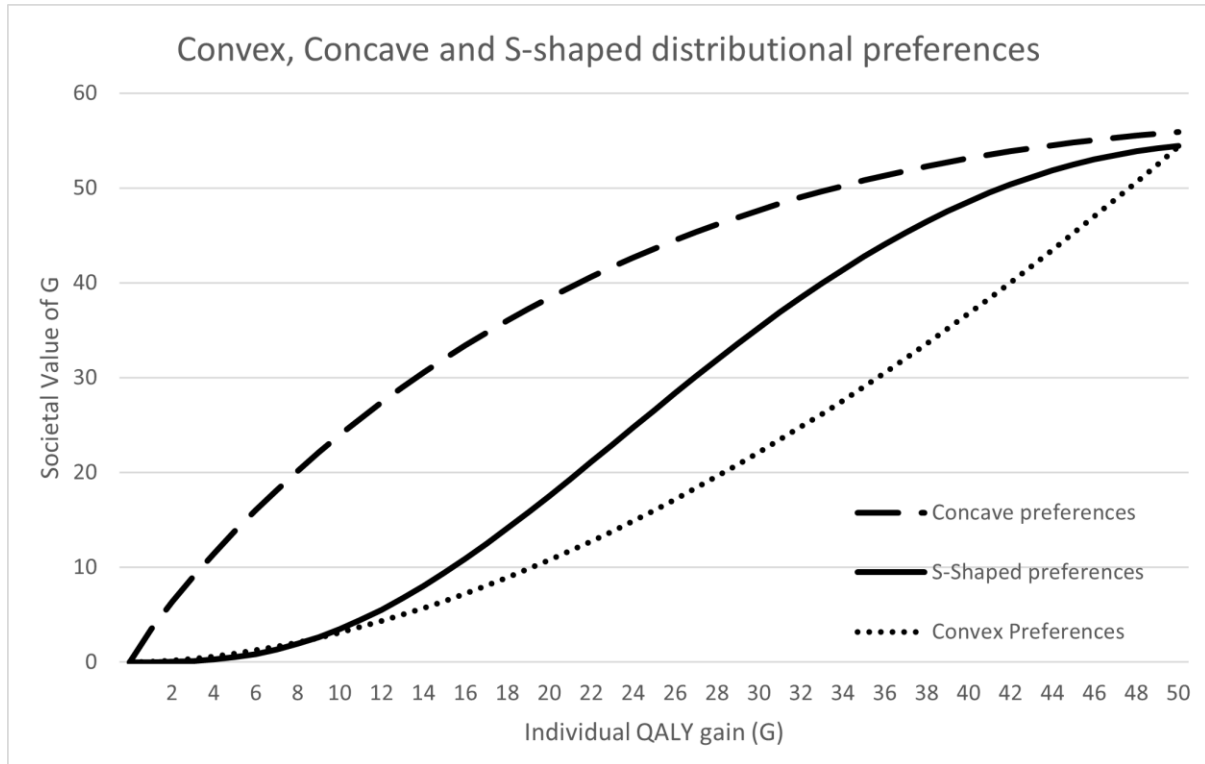


Figure 1: Convex, Concave and S-shaped distributional preferences

The underlying trade-off is represented by: (G,p) or (g,P) , where G or g is health-gain per head and P or p is number of people who benefit; with $G > g$, $P > p$ and $G \times p = g \times P$. When testing for S-shaped preferences, the null hypothesis is, ceteris paribus, distribution-neutral QALY maximisation holds; individuals are indifferent between (G,p) and (g,P) .

Whilst extensive literature on inequality aversion exists⁷, the concept must be distinguished from concentration and dispersion. Research on inequality aversion elicits preferences on health *outcomes* (*resulting level of health after an intervention*), whereas S-shaped preferences elicit preferences on health *gains* (*health improvement resulting from intervention*), distributed among *identical* individuals. Thus, research on distributional preferences based of health outcomes across different parties⁸, falls outside scope. Furthermore, preferences for concentration and dispersion relate to normative social preferences that concern interpersonal priority setting, as opposed to individual preferences over one's own health^{9,10}.

Whilst Discrete Choice Experiments (DCEs) present pairwise comparisons where attributes such as HRQoL, duration, and beneficiaries vary simultaneously¹¹ and participants are asked to choose between pairs, some stated preference methods elicit preferences by identifying an indifference point between alternatives. The Person-Trade-Off (PTO)¹² varies the number of beneficiaries, Gain-Trade-Off (GTO)¹³ the health-gain, and Time-Trade-Off (TTO) life-years. Methods differ in numeraire: the element changed to identify the point of indifference, as well as in cognitive complexity and theoretical grounding.

In a 2005 review, McKie and Richardson's found that societal preferences generally favour dispersion¹⁴ of health gain, though later studies^{15,16} reported an threshold where preferences switch from convex to concave as individual gains increase. Methodological challenges include how per-person health gains are measured, selection of numeraire range that captures the potential threshold, and accounting for framing effects.

The objective of this systematic review is to identify and critically evaluate methods used to elicit S-shaped distributional preferences for health gains. This review will focus on how existing methods have been designed, used and adapted to capture these preferences.

Research questions

1. What methods have been used to elicit s-shaped distributional preferences?
2. What are strengths and limitations of existing elicitation methods, and how can they inform design of more effective surveys?

Methods

This systematic review was conducted and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement¹⁷. A PRISMA checklist is available in Appendix A. This study was not pre-registered.

Search Strategy

Searches were carried out in PubMed, PsycINFO and EconLit in August 2024 with an update search in October 2025. The search strategy combined keywords related to "health," "distribution," "preferences," and "elicitation methods" (see Appendix B, Table 1 for full search strategy). One round of forward/backward citation chasing was performed on four key papers^{15,18–20}; selected because they represent seminal empirical evidence on S-shaped preferences. These studies were used to develop keywords for the keyword search and to assess sensitivity of the draft search, tested in MEDLINE. If the draft search did not return the four key papers, the search strategy was revised and re-run.

Screening

Studies were selected using predefined inclusion/exclusion criteria from the protocol (Appendix B, Table 2). After removing duplicates, one reviewer (LT) screened all records, with a second reviewer (AM) independently checking 20% at each stage; disagreements were resolved through discussion. Studies including trade-offs across heterogeneous individuals were excluded, but they did not need to explicitly explore both concentration and dispersion of health care to be included. One amendment was made to the protocol; conference papers were included due to high relevance of one paper that would have otherwise been excluded²¹.

Data Extraction

Data were extracted by a single reviewer (LT) using predefined fields. Extraction information included: author, year, title, study type, elicitation method, aims, year and country of fieldwork, mode of administration, population, sample size, response rate, priority setting context, perspective, pilot survey, survey details and results, statistical analysis, social welfare function,

theoretical arguments, and conclusions. Two additional fields, response rate and pilot survey, were added during pilot data extraction using two included papers.

Quality Assessment

Quality assessment was conducted by a single reviewer (LT) using the Joanna Briggs Institute (JBI) Critical Appraisal Tool for analytical cross-sectional studies²². One item ('Was the exposure measured in a valid and reliable way?') was removed as it was not deemed appropriate in this setting. Confounders such as duration or age effects were considered, and a dominant response check $[(g,p) < (G,P)]$ tested rationality.

Data Synthesis

Due to methodological heterogeneity across included studies quantitative meta-analysis was not appropriate. Therefore, narrative synthesis was conducted to compare how different study designs addressed common challenges in eliciting distributional preferences. Extracted data were grouped thematically based on methodological areas. Key characteristics and findings were tabulated, and the results were structured around the main themes from this synthesis.

Results

Searches and Screening

As shown in Figure 2, the keyword search identified 1,352 unique studies, with an additional 236 found through citation chasing. After full-text screening of 53 papers, 13 studies were included in the review.

PRISMA 2020 flow diagram for new systematic reviews which included searches of databases, registers and other sources

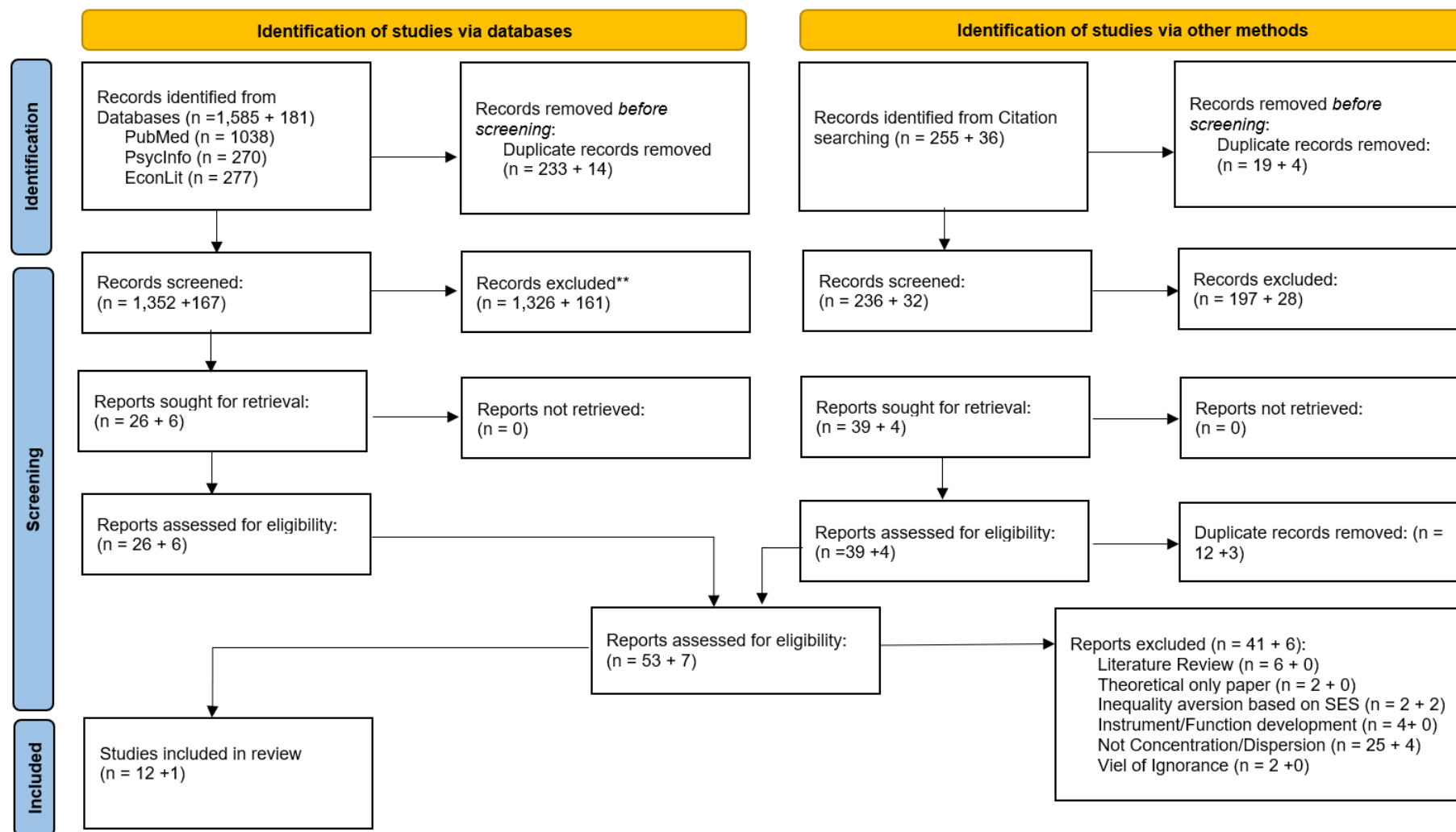


Figure 2: PRISMA Flow diagram

Table 1 summarises included studies. There were eleven studies that used a PTO and two that explored both PTO and TTO. All generally had low response rates where reported (28-71%). Three studies reported the year of fieldwork^{20,23,24}. Due to the age of many of the included studies, mode of administration was primarily in person. Six studies delivered the survey remotely (postal or online)^{18,20,24-26}. Most studies (9/13) did not report a pilot survey or conduct evaluation on the acceptability of the questions asked in the elicitation exercise.

Author (s) / Year	Elicitation Method	Population	Country	Sample Size (response rate)	Pilot
Choudhry <i>et al</i> , 1997 ²⁰	PTO	Ministry of Health officials	Canada	80 (61%)	No
Gyrd-Hansen and Kristiansen, 2008 ³²	PTO	General Population	Denmark	2923 (42.4%)	No
Fang <i>et al</i> ²³	PTO	General Population	China	500 (100%)	Yes
Nord, 1996 ³³	PTO	General Population	Australia	769 (28%)	No
Olsen, 1994 ¹⁹	PTO, TTO	Students	Norway	134 (NA)	No
Olsen, 2000 ²⁷	PTO	General Population	Norway	716 (34.3%)	Yes
Pinto-Prades and Lopez-Nicolás, 1998 ¹²	PTO	Students	Spain	83 (NA)	No
Richardson and Nord, 1997 ²⁵	PTO, TTO	General Population	Australia	156 (28%)	No

Richardson <i>et al</i> , 2012 ¹⁸	PTO	General Population	Australia	544 (NA)	Yes
Rodriguez-Miguez and Pinto-Prades, 2002 ¹¹	PTO	Students	Spain	61 (NR)	No
Tsuchiya and Hukin, 2005 ²¹	PTO	General Population	UK	63 (71%)	No
Ubel <i>et al</i> , 2001 ²⁶	PTO	General Population	USA	683 (NA)	No
Voorhoeve <i>et al</i> , 2018 ²⁸	PTO	Students/General Population	UK	82 (NA)	Yes

Table 1: Summary of Included Papers

Quality Assessment

Results of the quality assessment are shown in Table 2. The included papers represent mixed quality, reflecting variation in sampling, reporting and piloting procedures. Although no standardised criteria exist for identifying S-shaped preferences, most papers applied systematic and quantifiable elicitation methods, with several^{15,21,23} adapting existing designs. One study eliciting S-shaped preferences²³ was found to have a biased design (see Appendix C for details). Strategies to reduce ‘noise’ in the data included incorporating questions that displayed a dominant option and excluding respondents who failed these checks. All studies employed appropriate statistical analyses. Though the included studies were of mixed quality, the synthesis of this review focuses strengths and weaknesses of the methods used rather than the quantitative outcomes of the studies themselves. As such, the quality assessment highlights methodological heterogeneity rather than fundamental flaws, and results should be interpreted primarily in terms of their design rigour.

Author (s) / Year	Were the criteria for inclusion in the sample clearly described?	Is the sample representative of the general population?	Were the study subjects and settings described in detail?	Were objective, standard criteria used for measurement?	Were confounding factors identified?	Were strategies to deal with confounding factors stated?	Was appropriate statistical analysis used?
Choudhry <i>et al</i> , 1997 ²⁰	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gyrd-Hansen and Kristiansen, 2008 ³²	Yes	No	Yes	Yes	Yes	No	Yes
Fang et al, 2025 ²³	Yes	Yes	Yes	No	No	No	No
Nord, 1996 ³³	Yes	Unclear	Yes	Yes	No	No	Yes
Olsen, 1994 ¹⁹	No	No	Yes	Yes	No	No	Yes
Olsen, 2000 ²⁷	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Pinto-Prades and Lopez-Nicolás, 1998 ¹²	Yes	No	Yes	Yes	No	No	Yes
Richardson and Nord, 1997 ²⁵	Yes	Yes	Yes	Yes	Yes	No	Yes
Richardson <i>et al</i> , 2012 ¹⁸	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Rodriguez-Miguez and Pinto-Prades, 2002 ¹¹	Yes	No	Yes	Yes	Yes	Yes	Yes
Tsuchiya and Hukin, 2005 ²¹	Yes	No	No	Yes	Yes	Yes	Yes
Ubel <i>et al</i> , 2001 ²⁶	Yes	Unclear	Yes	Yes	No	Unclear	Yes
Voorhoeve <i>et al</i> , 2018 ²⁸	Yes	No	Yes	Yes	Yes	Yes	Yes

Table 2: Quality Assessment

Themes Identified within the literature

We identified five key methodological themes, which represent the papers' central methodological focus.

Capturing change in health in terms of Duration or HRQoL

Health-gain can be framed by duration, QoL, or a combination of both. Duration-based framing risks confounding distributional preferences with time or age biases, while HRQoL-based avoids this but increases cognitive burden. For example, Rodriguez-Miguez and Pinto-Prades¹⁵ asked participants to choose between a constant option (5 years for 20 people) and a varying one (10 years for a changing number), whereas Voorhoeve et al²⁴ used HRQoL scales to compare programmes offering different quality improvements.

Three studies, all using a PTO^{18,19,27} compared these approaches. They argued that time-based measures may conflate true preferences with time or risk attitudes, since positive time preference promotes dispersion while risk aversion favours concentration. Threshold effects can also explain preferences for concentration below the threshold. When G or g is varied solely in terms of years in full health, this S-shape may reflect conflicting biases rather than true preferences. For example, age-related confounding may occur as respondents may infer life expectancies intended in the survey. Richardson et al¹⁸ argue that PTOs involving trade-offs between beneficiary numbers and duration of benefit are confounded, whereas scenarios involving quality-of-life improvements avoid this issue by excluding time and age-related biases.

Olsen³⁰ discussed four ways to present health-gain G or g , in the (G,p) , (g,P) trade-off:

- [1] *Health-gain in terms of Life Years: $(G, p) = (10, 5)$.* 10 years of additional life years, given to 5 people each, with a fixed HRQoL.

- [2] *Health-gain in terms of QALYs (HRQoL * Life Years):* $(G, p) = (10, 5)$. 10 years of additional life years, in full health or an equivalent combination, given to 5 people each.
- [3] *Health-gain in terms of HRQoL:* $(G, p) = (0.8, 5)$. 0.8 increase in utility given to 5 people, no change to expected life years.
- [4] *Health-gain in terms of Number of years at a certain HRQoL value:* $(G, p) = (10, 5)$. 10 additional life years given to 5 people each, no change in HRQoL, patients all have a specified HRQoL.

Olsen highlighted that each method has limitations. Life-years [1] can be confounded by time and age preferences. QALYs [2] may be too complex for the public and still suffer from duration-confounding effects. Direct HRQoL changes [3] can be difficult to interpret, as respondents may not weigh utility intervals linearly; respondents may interpret a health state improvement from 0.2 to 0.4 as not exactly twice as good as 0.2 to 0.3^{28,29}. Olsen commented that the fourth approach, using years with a defined improvement, was the best of the alternatives as respondents are more likely to understand the patients in the question to have identical physical values when the gain is given in years. While this reduces utility misinterpretation and some confounding, challenges in presenting health gains and cognitive burden remain.

Overall, these studies show the way health-gain is represented may shape the preferences elicited. When expressed in life-years, results may be vulnerable to confounding from time and age effects, which may obscure genuine S-shaped preferences, though this was not empirically tested for in the included studies. Representing gains through HRQoL mitigates some of these issues but adds cognitive complexity for respondents. Thus, the choice of duration or HRQoL based framing may involve a methodological trade-off between conceptual validity and task comprehensibility.

Eliciting preferences for a threshold

Five studies^{15,16,21,23,27} explored the existence of a threshold level below which participants prefer to concentrate and above which to disperse health benefits. In general people prefer to disperse healthcare benefits across a wider population; but this may not be always true.

Tsuchiya and Hukin²¹ illustrate this with reference to an extreme scenario: extending the life of over half a million people for one minute each is unlikely to have the equivalent value to extending the life of one person by one year (525,600 minutes). This suggests that dispersing a fixed amount of total health benefits beyond a certain number of people is likely to diminish the value of the total benefit.

Pinto-Prades and Lopez-Nicolás¹⁶ tested the existence of a plateau effect in social health valuations a PTO. Respondents were reluctant to trade life-saving benefits for mild improvements, suggesting diminishing social value at higher health levels. Although no specific threshold was identified, results indicate a plateau beyond which additional gains hold little value. Importantly, they note this effect may reflect not just social preferences but the cognitive difficulty of making trade-offs between qualitatively distinct outcomes. Olsen²⁷ observed that respondents provided clearer and more ‘well-behaved’ answers when health-gain was small ($G = 1$ life-year), but their responses became more variable as the gain increased to 2 or 5 life-years. This pattern indicates a possible cognitive limit.

Rodriguez-Miguez and Pinto-Prades¹⁵ were the first to identify this threshold, at 9.1 years; above which preferences for dispersion were more dominant, and below which, concentration. Their study was conducted in Spain with an interviewer-led paper survey with a small, homogenous student sample ($n = 61$), likely reducing cognitive burden but limiting generalisability. Tsuchiya and Hukin²¹ replicated the study in the UK using card-based tasks,

shorter durations, and a broader community sample ($n = 63$), finding a lower threshold of 2.6 years. The design and sample differences, together with cultural or linguistic variation may have contributed to the lower threshold estimate. Fang et al²³ also replicated the Rodriguez-Migues and Pinto-Prades study, but found a reverse S-shape with an threshold at 4.6 years (see below for possible framing effects).

The importance of framing effects

Framing effects refer to systematic changes in respondents' choices that occur when identical trade-offs are presented in different contextual or numerical formats. Three studies focused on framing effects of preference elicitation survey^{21,24,26}. Tsuchiya and Hukin²¹ and Ubel²⁶ investigated whether the size of the population 'pool' in the background influenced responses, when the values of p and P were held constant. Evidence on these framing effects is mixed. While Tsuchiya and Hukin found little sensitivity to reference population size, Ubel et al found that framing significantly influenced preferences: when the same number of beneficiaries represented a larger share of the hypothetical population, respondents exhibited stronger preferences for dispersion. Ubel cautioned that such framing effects challenge the reliability of elicited preferences, as respondents may react differently when a treatment appears to benefit 'almost all' versus a smaller proportion of the population. This aligns with broader evidence that people are often more sensitive to relative rather than absolute changes when making allocation decisions^{30,31}. This suggests that observed preferences for dispersion under narrow framing may reflect sensitivity to relative impact, rather than purely distributive fairness.

Voorhoeve et al²⁴ examined framing effects through the lens of the similarity heuristic, where respondents disregard similar attributes across options and base decisions solely on differing ones. For example, a participant described individuals with an HRQoL of 0.95 as "almost fine," effectively treating them as if their HRQoL were 1.

Fang et al²³ found a reverse S-shape preference, which, at face value, is counter to earlier results. The study aimed to identify S-shape preferences using five PTO tasks covering five levels of health-gain. However, each PTO had different numbers of sub-tasks, with different maximum and minimum implied values of P*. Simulating random responses through the choices, we found that dispersion is more likely for 1, 2, and 5 year health gains and concentration for 20 and 50 year health gains. In other words, the framing of choices was biased in favour of finding reverse S-shape preferences, which is what the study found (see Appendix C for full details).

Different perspectives used to elicit preferences

Four studies explored the impact of the perspective used in the choice tasks^{16,25,32,33} i.e. if the respondent was asked to select the option that *they* would prefer for themselves, or if the respondent was asked to imagine themselves in the role of a decision maker on behalf of society. Richardson and Nord²⁵ compared three perspectives; a social perspective, where respondents act as decision-makers allocating resources between others; a socially inclusive personal (ex-ante) perspective, where respondents consider that they themselves may belong to one of the affected groups, and a purely personal perspective, involving trade-offs with ones own health³⁴. A standard PTO was used to assess the personal perspective, and a modified version of the PTO was used to evaluate the socially inclusive personal perspective (PTO-self), whereas a TTO was used to elicit the purely personal preferences. Health state values eliciting in the inclusive personal perspective were higher than those obtained using a standard PTO, where there was no personal stake in the decision. Similarly, Nord et al³³ found that the weight given to the duration of health benefits was sensitive to the perspective used: subjects placed greater value on the duration of benefits in a decision-maker perspective than when the treatment might benefit themselves.

Gyrð-Hansen et al³² also demonstrated that perspective meaningfully shaped preferences. When respondents evaluated programmes that varied in both the size of health gains (1 month to 3 years) for each person and the success rate of the treatment ($1 - 1/180$), the results showed differing results based on the perspective of the question. The results mirrored the findings of Richardson and Nord²⁵; from a societal perspective, respondents placed greater weights on the number of treatment recipients (preferences for dispersion). In contrast, when decisions were framed from a personal perspective, participants focused more on maximising their own expected gain. However, as this study did not examine threshold effects, preferences were framed as concentration *or* dispersion, rather than S-shaped preferences. Taken together, these studies suggest that adopting a decision-maker perspective may encourage preferences for greater dispersion.

Pinto-Prades and Lopez-Nicolás¹⁶ explored the relationship between individual and societal perspectives in the context of S-shaped preferences. Responding to O’Leary et al.³⁵, they agreed that treating minor conditions should not outweigh saving a life but noted that within-person TTO or Visual Analogue Scale methods capture only individual utilities, not societal preferences. They argued the relationship between individual and social values is more complex, and conventional TTO data cannot reveal the shape of the social welfare function or how incremental HRQoL gains compare to life-saving benefits.

Overall, respondents framed as decision-makers tend to display stronger willingness to make trade-offs, whereas self-regarding framing often elicit risk-averse preferences for dispersion. This highlights the importance of clearly specifying the chosen perspective in survey design.

Adaptation of objective functions

One study¹⁵ adapted a Health Related Social Welfare Function (HRSWF) to reflect S-shaped preferences. Under standard QALY maximisation, the social welfare function is additive, meaning that total health related social welfare is obtained by summing individual QALY gains across people. Each additional QALY contributes the same amount of social value, regardless of who receives it. Relaxing this assumption allows a non-linear relationship between social value and the number of QALYs gained. To allow for comparison to other additive HRSWFs, distributional preferences can be incorporated within the additive framework by allowing the marginal social value of each QALY to vary according to the recipient's health state or the size of the gain³⁶. Rodríguez-Míguez and Pinto-Prades adopted this latter approach, modifying the additive HRSWF to allow increasing or diminishing marginal social value. In the HRSWF below, t_i indicates the number of QALYs an individual i receives from a treatment, $U(t_i)$ is the social value of t_i , and α are scaling and shape parameters. The exponential term ($\alpha_1 e^{-\alpha_2 t_i}$) interacts with $t_i^{\alpha_3}$ to capture varying distributional preferences, allowing it to be either positive (dispersion) or negative (concentration) within the HRSWF, depending on the level of t_i .

$$U(t_i) = \alpha_1 e^{-\alpha_2 t_i} t_i^{\alpha_3}$$

A graphical representation of $U(t_i)$ is seen in Figure 1. Both Tsuchiya and Hukin²¹ and Fang et al²³ later applied this same functional form in their replication studies. However, no studies yet have proposed or tested questioned alternative forms that may provide a better fit to these preferences.

Discussion

This review set out to evaluate how empirical methods capture S-Shaped distributional preferences, where the marginal social value of health gains rise initially and diminish beyond a certain point. By synthesising methodological approaches across studies, the review highlights that design choices, particularly perspective and framing, shape the preferences elicited. There were several key themes within the literature, including how to present health-gain, the presence and value of a threshold level of health, framing effects, perspective, and adaptation of objective functions.

With respect to the first research question, ‘What methods have been used to elicit S-shaped distributional preferences on the concentration and dispersion of health gains?’, the included studies highlight significant methodological challenges and varied outcomes. All included studies used a PTO. Whilst theoretically, a GTO could capture distributional preferences, it remains largely unexplored in this context. Two of the included studies used an inter-personal TTO alongside the PTO; showing distributional preferences may be elicited using this method, but it is used less often. Though not included in this review as the authors elicited social preferences on patients in differing health states, Damschroder et al found that in a comparison between PTO and inter-personal TTO, TTO surveys were more likely to result in preferences for QALY maximisation³⁷.

Differences in how health gains were expressed often shaped the pattern of responses: when benefits were framed in terms of life-years, results could be confounded by age or time-preference effects, whereas HRQoL-based framing reduces those biases but increases cognitive burden and response variability. Studies that explored the reference-pool size found heterogenous results, suggesting further research is needed on its impact. A similar number of studies asked participants to take either a decision-maker or personal perspective. Preferences for dispersion were stronger in the decision-maker role, though the effect on the threshold

remains unclear^{25,32,26}. While the threshold was discussed in five studies, only two papers^{10,21} successfully attempted to quantify this, finding largely different results.

Our second research question asked: *What are the strengths and limitations of existing elicitation methods, and how can they inform the design of more effective surveys?* Overall, the reviewed studies show that current approaches are conceptually well-founded but methodologically sensitive. The PTO was most used and directly represents the normative trade-offs between concentration and dispersion of health gains. Several studies included dominance or comprehension checks to improve internal validity, and adaptations that estimated thresholds and parameterised social welfare functions represent progress in quantifying S-shaped preferences.

However, these methods also exhibit weaknesses that limit reliability and interpretability. PTO and inter-personal TTO tasks are cognitively demanding, and responses are highly sensitive to how the problem is framed. Small changes in the description of health-gain (e.g. life-years versus HRQoL), the reference pool size, or the decision-making perspective can substantially change preferences, suggesting that some findings reflect contextual framing rather than stable distributive values. Few studies reported pilot testing or comprehension evaluation, despite their importance in ensuring clarity for complex trade-off tasks³⁸. Low response rates, small samples, and the near-exclusive use of PTO with limited exploration of alternatives such as GTOs or DCEs limit the confidence in which results are to be interpreted.

Limitations

This review has several limitations. First, although it followed systematic review protocols, grey literature, policy documents and theses were not included. As a result, some relevant evidence may have been missed.

The included literature has important limitations. Most included studies are relatively old and predate the widespread use of online surveys, now the dominant mode of administration³⁹.

While some evidence suggests mode of administration does not affect results, it can influence how much and what type of information is presented, limiting the applicability of some findings to online settings. Several studies relied on niche samples, such as students, which may limit external validity and generalisability to broader populations. Finally, a threshold level was underexplored, limiting cross-population comparability and understanding of a more precise shape of distributional preferences. Although this lack of clarity is itself a notable finding, the concept remains poorly understood and warrants further investigation.

Conclusion

The included studies collectively provide a valuable evidence base. Their diversity of contexts and populations begins to reveal methodological drivers behind S-shaped preferences regarding improvements in health, and the consistent use of PTO enables tentative survey comparisons despite heterogeneity in framing and administration. Future research should explore uncertainties by including additional questions to tasks that explicitly changes one methodological element and testing for consistency of response. Greater methodological transparency and replication studies would help distinguish genuine societal preferences from methodological artefacts, thereby improving robustness and policy relevance.

References

1. Dolan P, Shaw R, Tsuchiya A, Williams A. QALY maximisation and people's preferences: A methodological review of the literature. *Health Econ.* 2005;14(2):197-208. doi:10.1002/hec.924
2. Richardson J, Schlender M. Health technology assessment (HTA) and economic evaluation: efficiency or fairness first. *J Mark Access Health Policy.* 2019;7(1):1557981. doi:10.1080/20016689.2018.1557981
3. Dolan P, Tsuchiya A. It is the lifetime that matters: public preferences over maximising health and reducing inequalities in health. *J Med Ethics.* 2012;38(9):571-573. doi:10.1136/medethics-2011-100228
4. Schwappach DLB. Resource allocation, social values and the QALY: a review of the debate and empirical evidence. *Health Expect Int J Public Particip Health Care Health Policy.* 2002;5(3):210-222. doi:10.1046/j.1369-6513.2002.00182.x
5. van de Wetering L, van Exel J, Bobinac A, Brouwer WBF. Valuing QALYs in Relation to Equity Considerations Using a Discrete Choice Experiment. *Pharmacoeconomics.* 2015;33(12):1289-1300. doi:10.1007/s40273-015-0311-x
6. Overview | NICE health technology evaluations: the manual | Guidance | NICE. January 31, 2022. Accessed December 15, 2025. <https://www.nice.org.uk/process/pmg36>
7. Cadham CJ, Prosser LA. Eliciting Trade-Offs Between Equity and Efficiency: A Methodological Scoping Review. *Value Health J Int Soc Pharmacoeconomics Outcomes Res.* 2023;26(6):943-952. doi:<https://dx.doi.org/10.1016/j.jval.2023.02.006>
8. Lancsar E, Wildman J, Donaldson C, Ryan M, Baker R. Deriving distributional weights for QALYs through discrete choice experiments. *J Health Econ.* 2011;30(2):466-478. doi:10.1016/j.jhealeco.2011.01.003
9. Kvamme MK, Gyrd-Hansen D, Olsen JA, Kristiansen IS. Increasing marginal utility of small increases in life-expectancy?: Results from a population survey. *J Health Econ.* 2010;29(4):541-548. doi:10.1016/j.jhealeco.2010.03.010
10. Lakdawalla DN, Phelps CE. The Generalized Risk-Adjusted Cost-Effectiveness (GRACE) Model for Measuring the Value of Gains in Health: An Exact Formulation. *J Benefit-Cost Anal.* 2023;14(1):44-67. doi:10.1017/bca.2023.6
11. Clark MD, Determann D, Petrou S, Moro D, de Bekker-Grob EW. Discrete Choice Experiments in Health Economics: A Review of the Literature. *Pharmacoeconomics.* 2014;32(9):883-902. doi:10.1007/s40273-014-0170-x
12. Nord E. The person-trade-off approach to valuing health care programs. *Med Decis Mak Int J Soc Med Decis Mak.* 1995;15(3):201-208. doi:10.1177/0272989x9501500302

13. Maestad O, Norheim OF. Eliciting people's preferences for the distribution of health: A procedure for a more precise estimation of distributional weights. *J Health Econ*. 2009;28(3):570-577. doi:<https://dx.doi.org/10.1016/j.jhealeco.2009.02.008>
14. McKie J, Shrimpton B, Richardson J, Hurworth R. Treatment costs and priority setting in health care: A qualitative study. *Aust N Z Health Policy*. 2009;6:11. doi:<https://dx.doi.org/10.1186/1743-8462-6-11>
15. Rodriguez-Miguez E, Pinto-Prades JL. Measuring the social importance of concentration or dispersion of individual health benefits. *Health Econ*. 2002;11(1):43-53. doi:<https://dx.doi.org/10.1002/hec.643>
16. Pinto-Prades JL, Lopez-Nicolás A. More evidence of the plateau effect: A social perspective. *Med Decis Making*. 1998;18(3):287-294. doi:10.1177/0272989X9801800306
17. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. doi:10.1136/bmj.n71
18. Richardson J, Sinha K, Iezzi A, Maxwell A. Maximising health versus sharing: Measuring preferences for the allocation of the health budget. *Soc Sci Med* 1982. 2012;75(8):1351-1361. doi:10.1016/j.socscimed.2012.05.036
19. Olsen JA. Persons vs years: two ways of eliciting implicit weights. *Health Econ*. 1994;3(1):39-46. doi:<https://dx.doi.org/10.1002/hec.4730030106>
20. Choudhry N, Slaughter P, Sykora K, Naylor CD. Distributional dilemmas in health policy: Large benefits for a few or smaller benefits for many? *J Health Serv Res Policy*. 1997;2(4):212-216. doi:10.1177/135581969700200405
21. Tsuchiya A, Hukin A. Dispersion vs. concentration of health benefits: preliminary report.
22. JBI Critical Appraisal Tools | JBI. Accessed December 15, 2025. <https://jbi.global/critical-appraisal-tools>
23. Fang N, Su C, Wu J. Eliciting Distributive Preferences in Health Care Resource Allocation: A Person Trade-Off Study. *Healthcare*. 2025;13(11):11. doi:10.3390/healthcare13111309
24. Voorhoeve A, Stefánsson A, Wallace B. Similarity and the trustworthiness of distributive judgements. *Econ Philos*. 2018;35(3):537-561. doi:10.1017/s0266267118000457
25. Richardson J, Nord EJ. The importance of perspective in the measurement of quality-adjusted life years. *Med Decis Mak Int J Soc Med Decis Mak*. 1997;17(1):33-41. doi:10.1177/0272989x9701700104
26. Ubel PA, Baron J, Asch DA. Preference for equity as a framing effect. *Med Decis Mak Int J Soc Med Decis Mak*. 2001;21(3):180-189. doi:<https://dx.doi.org/10.1177/0272989X0102100303>
27. Olsen JA. A note on eliciting distributive preferences for health. *J Health Econ*. 2000;19(4):541-550. doi:10.1016/S0167-6296(00)00035-7

28. Loomes G, McKenzie L. The use of QALYs in health care decision making. *Soc Sci Med*. 1989;28(4):299-308. doi:10.1016/0277-9536(89)90030-0
29. Roberts J, Dolan P. To what extent do people prefer health states with higher values? A note on evidence from the EQ-5D valuation set. *Health Econ*. 2004;13(7):733-737. doi:10.1002/hec.875
30. Brady D, Curran M, Carpiano RM. A test of the predictive validity of relative versus absolute income for self-reported health and well-being in the United States. *Demogr Res*. 2023;48:775-808. doi:10.4054/demres.2023.48.26
31. Li H, Xie X. Absolute outcome or relative standing? The difference between making decisions for the self and other. *Curr Psychol*. 2023;42(30):26065-26078. doi:10.1007/s12144-022-03676-w
32. Gyrd-Hansen D, Kristiansen IS. Preferences for “life-saving” programmes: Small for all or gambling for the prize? *Health Econ*. 2008;17(6):709-720. doi:https://dx.doi.org/10.1002/hec.1288
33. Nord. The significance of age and duration of effect in social evaluation of health care.
34. Dolan P, Olsen JA, Menzel P, Richardson J. An inquiry into the different perspectives that can be used when eliciting preferences in health. *Health Econ*. 2003;12(7):545-551. doi:10.1002/hec.760
35. O’Leary JF, Fairclough DL, Jankowski MK, Weeks JC. Comparison of Time-tradeoff Utilities and Rating Scale Values of Cancer Patients and Their Relatives: Evidence for a Possible Plateau Relationship. *Med Decis Making*. 1995;15(2):132-137. doi:10.1177/0272989X9501500205
36. Wagstaff A. QALYs and the equity-efficiency trade-off. *J Health Econ*. 1991;10(1):21-41. doi:10.1016/0167-6296(91)90015-F
37. Damschroder LJ, Roberts TR, Goldstein CC, Miklosovic ME, Ubel PA. Trading people versus trading time: What is the difference? *Popul Health Metr*. 2005;3(1):10. doi:10.1186/1478-7954-3-10
38. Hassan ZA, Schattner P, Mazza D. Doing A Pilot Study: Why Is It Essential? *Malays Fam Physician Off J Acad Fam Physicians Malays*. 2006;1(2-3):70-73.
39. Marshall DA, Veldwijk J, Janssen EM, Reed SD. Stated-Preference Survey Design and Testing in Health Applications. *Patient - Patient-Centered Outcomes Res*. Published online January 31, 2024. doi:10.1007/s40271-023-00671-6

Appendix A:



PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	Introduction / Air
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Introduction
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Research Questions
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Protocol, Appendix B
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Methods, Search Strategy
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Protocol, Appendix B
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	Methods, Screening
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Methods, Data Extraction
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Methods, Data Extraction
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Methods, Data Extraction
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	Methods, Quality Assessment
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	NA
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	Methods, Data Synthesis
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	Methods, Data Synthesis
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Methods, data synthesis
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	Methods, data synthesis
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	NA
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	NA
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	NA
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	NA
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Results, searching and screening
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	Results, PRISMA diagram
Study characteristics	17	Cite each included study and present its characteristics.	Results, Table 3
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	Results, Table 4
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	NA
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Results - Themes identified
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	NA
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	Results - Themes Identified
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	NA
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	NA
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	NA
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Discussion
	23b	Discuss any limitations of the evidence included in the review.	Discussion, Limitations
	23c	Discuss any limitations of the review processes used.	Discussion, Limitations
	23d	Discuss implications of the results for practice, policy, and future research.	Discussion, Conclusion

OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	Methods
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Methods, Screening, Appendix
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	Methods, Screening
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	Financial Declaration
Competing interests	26	Declare any competing interests of review authors.	COI declaration
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	Appendix B - Protocol

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71. This work is licensed under CC BY 4.0. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Appendix B:

Protocol: A systematic review of preference elicitation methods for the concentration and dispersion of healthcare.

Background:

The objective of this review is to explore elicitation methods used when eliciting social distributional preferences for the concentration or dispersion of public healthcare resources. The purpose is to inform subsequent studies to ensure the preference elicitation survey is based upon a strong evidence base. In the context of this literature review, distributional preferences refer to preferences regarding the allocation of resources within a population. Preferences are limited to other-regarding preferences, whereby individual preferences are influenced by the size of health gain received by others.

The review will also explore how these preferences translate/relate to social welfare functions, and if preferences differ in different settings, for example across countries. This review will include both experimental and theoretical studies in order to most comprehensively understand the various methods used to elicit preferences, and any nuances of how members of the public perceive the trade-offs between concentrated or dispersed healthcare, such as a threshold effect.

Research questions:

1. What preference elicitation methods have been used to elicit preferences on the concentration and dispersion of healthcare?
2. Evaluate the elicitation methods used across the literature, focusing on what strengths and weaknesses can be used in the development of a future elicitation survey.

Search Strategy:

Keyword Search

The following databases will be searched for relevant studies: PubMed, PsycINFO, EconLit.

The search strings will be developed using a combination of keywords and Boolean operators, based on the four key concepts of this literature review: concentration and dispersion, elicitation, preference, and health. The search term for each of the concepts is detailed below. For the search on PubMed, Medical Subject Heading (MeSH) keywords were included in the search.

Concept	Explanation	Search String
Concentration/Dispersion	How widely are healthcare resources distributed in a set population? A smaller gain to many, or a larger gain to few?	concentrat* or dispers* or maximis* or distribut* or share
Elicitation	Process of gathering information on preferences, for example, a person-trade-off study, benefit-trade off, willingness-to-pay. This may also include other methods of elicitation such as interviews, or postal surveys.	measuring or value or valuing or eliciting or elicit
Preference	Underlying values of and priorities that underpin the decision making in the elicitation	preference* or "inequality aversion" or "inequality-aversion" or "social value" or "inequality averse" or "aversion to inequalities" or weights or "social importance" or "distributional concerns"
Health		health or "quality-adjusted life years" or "QALYs" or "lifeyears" or "lifeyears"

Table 1: Search Strategy

As this review is aiming to examine both the elicitation methods used and the application of these methods, the search strings may be combined as below:

- 1.) Concentration/Dispersion search string
- 2.) Elicitation search string
- 3.) Preference search string
- 4.) Health search string
- 5.) 1 AND 2 AND 3 AND 4

Citation Chasing

The following four key studies have been selected as key papers for this review, in line with the below inclusion/exclusion criteria. These studies will be used for forward and backwards citation chasing.

1. Maximising health versus sharing: Measuring preferences for the allocation of the health budget. Richardson et al, 2012
2. Distributional Dilemmas in health policy: Large Benefits for a few or smaller benefits for many? Choundhry et al, 1997.
3. Measuring the social importance of concentration or dispersion of individual health benefits. Rodriguez-Miguez and Pinto-Prades, 2002
4. Persons vs years: two ways of eliciting implicit weights. Oslen, 1994

There will be one round of citation chasing, and the papers have been selected due to their high relevance with the research questions, and the papers have been published across a range of dates, and are therefore more likely to capture a wider range of relevant papers.

Screening:

Screening of the articles from the keyword search and citation chasing will be done following the PRISMA guidelines. This will include title and abstract screening, followed by full text screening. Inclusion and Exclusion decisions will be made in accordance with the Inclusion/Exclusion criteria as below. Screening will be done by one reviewer at the title and abstract and full text screening stage, with checks from a second review (AM) on 20% of the articles. Disagreements on inclusion will be resolved by a discussion between the reviewers.

Term	Inclusion Criteria	Exclusion Criteria
Population	Studies that examines the distributional preferences of individuals or groups with no specialist knowledge	Preferences from a population with specialist knowledge (patient preferences)

	General public including Student populations (including medical students) Preferences from decision makers, including clinicians, NHS (or other country equivalent) managers	Studies that focus on self-regarding preferences.
Intervention/Method	Elicitation Methods Preferences elicited in a hypothetical context that required a trade off. Preferences relating to trade offs. Treatment/healthcare is viewed as a public good	Elicitation of inequality preferences that do not focus on concentration/dispersion
Comparison	–	-
Outcome	Results of preference elicitation Type of preferences identified Social Welfare Function/Objective Function	
Study Design	Experimental	Protocol Instrument or function development Observational studies Literature Reviews

		Theoretical study
Other	Peer Review Journals Published in English	PhD thesis Not available in English Conference Papers

Table 2: Inclusion / Exclusion criteria

Data Extraction:

Once screening is complete, the data will be extracted according to the following fields:

1. Author
2. Year of publication
3. Year of fieldwork
4. Elicitation Method
5. Mode of administration
6. Sample Size
7. Population
8. Country
9. Priority setting context (health system, particular intervention or health condition, policies)
10. Perspective
11. Results
12. Reported evaluation of the method
13. Social Welfare Function

The data extraction will be trialled with the five pre-specified key papers, and adjusted to accommodate any additional relevant information.

Quality Assessment:

Quality assessment will be conducted using the Joanna Briggs Institute (JBI) Critical Appraisal Tools

Data Synthesis:

The extracted data will be synthesised using a narrative synthesis approach. This will involve summarising the findings from the included studies to address the research questions. The studies will

be categorised and described on the basis of: the elicitation method, the context of elicitation, and the application of the method or results. This categorization will be used to inform investigation of common links between select methods of preference elicitation and the application of the results.

Open Access:

All material used for this review, including template data collection forms and data extracted will be made available upon request to authors.

Appendix C

Simulation of Fang et al

Aim

To assess the effect of framing in Fang et al²³, we created a simulated data set where the answers to the five PTO variants used in Fang et al were randomly generated and contrasted with the empirical data.

Methods

In Fang et al, each of the five PTO variants has a different duration ranging from 1 year to 50 years. Each variant has a different number of titrating forced choice tasks, but within each variant, the number of choice tasks available is constant. In Excel, 500 participants were simulated for each PTO variant to replicate the sample size of the empirical study, by assigning the participants randomly across the titration paths within a given PTO variant. This means that value of P* for each of the participants was randomly generated within the constraints applied in each PTO variant.

Categorisation rules from Fang et al (Concentration, Diffusion, Maximising, Extreme Egalitarian, Inequality Seeking) were applied to the randomly simulated data. Additionally, a regression analysis was run using the HRSWF developed by Rodriguez-Migues and Pinto-Prades¹¹, used by Fang et al ($U(t_i) = \alpha_1 e^{-\alpha_2 t_i} t_i^{\alpha_3}$). The confidence intervals of the regression coefficients were compared to test the null hypothesis that the coefficients based on empirical data are no different from the corresponding coefficients based on randomly simulated data.

Results

The table below summarises the results:

Simulated (N = 500)	PTO variant				
	1 Year	2 Years	5 Years	20 Years	50 Years
Concentration	9.0%	17.2%	41.8%	91.8%	92.4%
Diffusion	88.6%	76.6%	42.2%	5.4%	4.4%
Maximising	1.2%	2.2%	4.6%	1.2%	1.2%
Extreme Egalitarian	0.2%	2.6%	5.0%	1.0%	0.8%
Inequality Seeking	1.0%	1.4%	6.4%	0.6%	1.2%
Total	100%	100%	100%	100%	100%

Table 3: Distribution of Preference types based on Simulated Data Set

For ease, we have included the results from Fang et al below:

Fang et al (N=500)	PTO variant				
	1 Year	2 Years	5 Years	20 Years	50 Years
Concentration	7.2%	24.0%	37.0%	78.0%	74.6%
Diffusion	89.0%	63.2%	33.6%	4.0%	17.4%
Maximising	3.4%	9.6%	21.2%	14.4%	4.0%
Extreme Egalitarian	0.2%	1.2%	6.8%	2.0%	1.6%
Inequality Seeking	0.2%	2.0%	1.4%	1.6%	2.4%
Total	100%	100%	100%	100%	100%

Table 4: Distribution of Preference types based on Fang et al²³

As seen from the two tables above, the patterns of response in Fang et al were similar to the randomly generated results. Notable differences are lower proportion of participants who are concentrators (in the higher evaluated years), and a higher proportion of maximisers.

Bias across Duration Variants

In a non-biased study design, the proportion of randomly simulated participants falling into each preference category should be consistent across PTO variants, so any observed variation in empirical data can be attributed solely to the duration of health gain. When we examine the randomly generated results, we can see that there are more responses for Diffusion in the shorter duration variants and more responses for Concentration under the longer duration variants. This highlights that the design itself is biased in favour of a reverse S-Shape – dispersing small gains and concentrating larger ones.

Including simulation of Rodriguez-Migues and Pinto Prades, and Tsuchiya and Hukin

Rodriguez-Migues and Pinto Prades¹¹ (RM-PP) and Tsuchiya and Hukin²¹ (T and H) were the other included studies that empirically estimated a threshold for S-Shaped preferences, with Tsuchiya and Hukin using the earlier study's design. We repeated the above exercise with their design. As the results were presented differently in these papers as compared to in Fang et al²³, we have presented the results in line with original studies to allow for ease of comparison.

In this case, we simulated 100 participants, and the P* value was randomly selected from the reference set given in RM-PP¹⁰ (always 10 years, but P = 1,2,5,8, 10, 12, 15,18, 20).

Health Gain	Number of Patients	Mean P* (Simulated -RM-PP and T-H design)	Mean P* (RM-PP ¹⁰)	Mean P* (T-H ²⁹)	Mean P* (simulated - Fang et al Design)	Mean P* (Fang et al ²³)

0.5	200	10.54	-	9.07	-	-
0.75	150	10.50	-	11.32	-	-
1	100	10.18	6.74	15.49	50.63	38.0
2	50	9.73	7.49	14.83	24.8	17.4
5	20	10.55	8.93	12.85	10.05	9.4
10	10	10	10	10	10	10
20	5	9.57	9.23	8.32	52.59	32.2
50	2	10.34	7.40	5.52	50.50	37.7

Table 5: Average P* Values for different levels of health gain

From the table above, when using the study design from Rodriguez-Migues and Pinto-Prades¹¹, the randomly generated results show mean P* values that are close to 10 (QALY maximisation) across all PTO variants and there are no biases for Concentration or Dispersion. The empirical results from these two papers show a clear pattern of average P* that initially increases with the health gain and then decreases, which is different from that of the simulated data. On the other hand, for Fang et al, the pattern of the P* values across the collected and simulated data are qualitatively the same, initially decreasing and then increasing, highlighting that the empirical results reported in the paper are likely to have been biased by the study design.

Comparison of results of the regression from Fang et al²³, Rodriguez-Migues and Pinto-Prades¹¹, and Tsuchiya and Hukin²¹, against the randomly generated results.

	Simulated based on Fang et al study design	Fang et al	Simulated based on RM-PP study design	Rodriguez-Miguez and Pinto-Prades	Tsuchiya and Hukin
Alpha1 (95% CI)	3.89 (3.66 - 4.12)	2.085 (1.9580 - 2.11)	77.35 (71.26 – 83.96)	0.446	0.941 (0.913 – 1.07)
Alpha2	-0.00419 (-0.000143 - 0.008526)	-0.035 (-0.030 - -0.029)	0.0048 (-0.0037 – 0.0133)	0.026	0.029 (0.032 – 0.022)
Alpha3	0.0783 (-0.00419 - 0.07831)	0.646 (0.587 - 0.705)	-0.9559 (-1.0456 - -0.8663)	1.435	1.173 (1.08 – 1.195)

The parameters based on empirical data from Fang et al lie outside the 95% confidence interval of the parameters based on simulated data using the same design, rejecting the null hypothesis that the empirical data are no different from randomly simulated data.

The null hypothesis is rejected for the corresponding results for Rodriguez-Miguez and Pinto-Prades and Tsuchiya and Hukin.

