

Validity, Reliability, and Framing Effects in Equity-Efficiency Trade-Off Studies: A Systematic Review

Running title: Validity and reliability in trade-off studies

Victoria Chechulina¹, Andrew Chu¹, Quang Hung Lam¹, Camryn Kabir-Bahk BMSc¹, Shehzad Ali MBBS MPH MSc PhD^{1,2}

1. Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada
2. WHO Collaborating Centre for Knowledge Translation and Health Technology Assessment in Health Equity, Ottawa, Ontario, Canada

Corresponding author:

Shehzad Ali, MBBS, MPH, MSc, PhD

Department of Epidemiology and Biostatistics, Western University

3rd floor, 1465 Richmond Street, London, Ontario, Canada, N6G 2M1

shehzad.ali@uwo.ca

Phone: 519-661-2111 X 80573

Abstract

Objectives

Health system resources are limited, and decision makers often need to make trade-offs between equity and efficiency. Such trade-offs are guided by public values that are elicited using choice experiments. There is no standard approach to elicit equity-efficiency trade-offs. Previous studies have found significant variability in public values, raising concerns about the robustness of trade-off experiments. The objective of this systematic review was to determine whether and how equity-efficiency trade-off studies consider validity, reliability and framing effects.

Methods

We searched Medline, EMBASE, and Web of Science in January 2024 for health-related equity-efficiency trade-off studies. Two reviewers independently screened titles and abstracts, followed by full-text review, data extraction, and quality assessment.

Results

115 equity-efficiency trade-off studies were identified, of which 29 studies (25.2%) investigated validity, reliability, and/or framing effects. 14 (12.2%) studies assessed validity, 9 (7.8%) studies assessed reliability, and 10 (8.7%) assessed framing or cognitive effects. Validity was most frequently assessed by comparing results to hypothesized expectations, while reliability was commonly assessed by providing a repeated test or questionnaire. Framing and cognitive effects were assessed by varying question order or changing the wording or framing of the

scenario. 19 of 23 studies reported high or acceptable validity or reliability, and 5 of 10 studies found no significant framing or cognitive effects.

Conclusion

This systematic review highlights the need to consider robustness of elicited values in choice experiments. It will also guide future choice experiments that aim to estimate inequity aversion.

Introduction

Health equity is a major public policy concern; however, limited progress has been made in recent decades.^{1,2} Health inequities can be mitigated through fairer allocation of health system resources.^{3,4} The standard economic evaluation framework used for resource allocation decisions is based on the efficiency principle which asserts that decisions should maximize the overall health, irrespective of its distribution.⁵ In contrast, the equity principle focuses on fairness in the population distribution of health.⁶ Lately, alternative economic evaluation approaches have been proposed, such as the distributional cost-effectiveness analysis (DCEA), that explicitly incorporate equity concerns in decisions.^{7,8} These approaches use evidence on public values concerning the trade-offs between health maximization (efficiency) and health equity to inform the optimal allocation of resources. Incorporating public values allows decision-makers to explicitly and transparently consider equity in decision-making.

To elicit public preferences on equity-efficiency trade-offs, preference elicitation studies are frequently used in the health economics literature.⁹ These studies typically present individuals with hypothetical scenarios where they are asked to choose between different levels of health gains and their distribution across social groups, in turn revealing the level of inequity aversion of the respondent.¹⁰ Previous systematic reviews of equity-efficiency trade-off studies have summarized public values on inequity aversion.¹¹⁻¹³

There are no standard questionnaire instruments to elicit equity-efficiency trade-offs. Studies in the literature have used different questionnaire designs and have estimated different levels of inequity aversion.¹¹⁻¹³ For example, in a systematic review of equity-efficiency trade-off studies

based in the UK, McNamara et al (2020) identified eight studies reporting evidence of inequality aversion, two finding no aversion, and five reporting mixed evidence.¹¹ The estimated Atkinson inequality aversion parameter ranged between 1 (i.e. no aversion) and 28.9, and the weight placed on a marginal gain to group with low lifetime health ranged between 1 and 166.2. While this large difference in inequity aversion in studies conducted in the same country can partly be explained by between-study heterogeneity, this degree of variation poses a challenge in determining which estimate can be reliably used for policy decisions.

One consideration that has received limited attention in the literature is the validity, reliability, and framing effects in equity-efficiency trade-off studies. While validity and reliability are frequently studied in relation to questionnaire instruments, such as quality-of-life tools; however, they are less frequently reported for questionnaire-based choice experiments.¹⁷ While the trend is changing with a number of discrete choice experiments explicitly investigating reliability, validity and framing effects,^{18–20} their use in equity-efficiency trade-off experiments has not been studied.

Janssen et al (2017) have developed a conceptual framework to study validity and reliability of choice experiments, focusing on the measurement and choice-related domains.²¹

To investigate validity, it is important to consider the normative, value-based nature of equity-efficiency trade-off experiments which is distinct from conventional discrete choice experiments where the expected direction of attribute preference can often be reasonably assumed (e.g. more health benefit being better than fewer). While not all aspects of validity are pertinent to equity-efficiency trade-off experiments, several domains may be relevant,

including face validity, convergent validity, external validity of measurement, monotonicity, and non-compensatory choices. Face validity of responses can be assessed by examining deviations from hypothesized response behaviors,²² such as the level of pro-rich or extreme egalitarian values among respondents,¹⁴ the willingness-to-pay for larger-sized quality of life gains, or treatment preferences for patients that are younger and have no comorbidities. Convergent validity of the elicited values can be assessed by comparing within or between-subject responses to other choice scenarios/instruments measuring the same underlying construct. For instance, an investigator may compare individual responses against their own social values as measured by social attitude questions within the same survey. Alternatively, the elicited values can be compared with previously published public attitude surveys. Other forms of validity, such as external validity, are more difficult to assess due to limited opportunity to compare elicited values with the predicted or observed choices in the real-world.²¹ Monotonicity relates to consistency in the direction of responses across choice sets. In case of equity-efficiency trade-off studies, this may relate to consistency in relation to a respondent's preference for welfare increasing choice options.¹⁴ Finally, non-compensatory choices imply that individuals might only consider one attribute to make choices in the experiment. In case of equity-efficiency trade-off studies, this relates to giving exclusive consideration to the equity attribute such that no level of efficiency loss would compensate for inequity. This may present as an 'extreme egalitarian' position when a respondent treats equity as a 'sacred value'.¹⁴

Reliability in this context refers to the ability of a measure to produce consistent results over time and space.²³ While not all domains of reliability are pertinent to equity-efficiency trade-off experiments, test-retest reliability, version consistency and level recoding can be relevant. Test-

retest reliability is typically assessed by administering the questionnaire twice, at two different timepoints, assuming stability of preferences.²¹ Version consistency assesses if preferences remain stable across different versions of the questionnaire administered to separate groups of participants.²¹ Finally, level recoding refers to the extent to which participants process the numeric values presented in attributes. To assess this, researchers can administer two versions of the survey with different level values of the health benefits being traded. However, if the units or the range of health benefits are changed across versions, such as changing healthy years to healthy months, the level of trade-offs can be expected to be different.¹⁴

In addition to validity and reliability, other areas considered in choice experiments are framing and cognitive effects. Framing effects occurs when preferences vary based on how choice scenarios are presented to participants.¹⁴ The framing may vary in terms of the presentation of scenarios, such as gains versus losses, relative versus absolute changes and small versus large units (such as days versus weeks). For instance, Howard et al (2009) found that framing the accuracy of a cancer detection test in terms of cases found versus cases missed can significantly influence the trade-offs.²⁴ Similar framing effects were reported by Veldwijk et al (2016) in relation to the outcome of genetic testing expressed either as survival (positive framing) or mortality (negative framing).²⁵ In the context of equity-efficiency trade-off literature, Ali et al (2017) found that the level of inequity aversion varied based on abstract (i.e. not specific to a health condition) versus concrete (i.e. condition-specific) scenario descriptions.¹⁴ This could be attributed to the idea of agentic responsibility when a specific condition is presented compared to wider structural determinants of health when considering broad health system inequities.

Framing effects may violate the principle of invariance, which states that decisions should not vary as a function of equivalent descriptions of the same situation.²⁶

Preferences may also vary across scenarios based on the cognitive burden of choice experiments. For instance, probabilistic outcomes may be harder to process cognitively in a choice experiment than fixed frequencies.²⁷ As such, the cognitive load of the choice experiment may impact the elicited values.²⁸ Researchers should determine whether the results of their preference elicitation studies have been skewed by framing effects or other cognitive biases. Results that have been influenced by cognitive biases may not be reliable enough to inform health policy making, as they may not generalize from the study scenario to real-world settings.¹⁴

Choice experiments may also employ other approaches to assess the robustness of responses. This may include sensitivity analysis, whereby researchers may exclude participants who did not understand the survey, provided inconsistent responses, or who showed no real preference on the choice task.^{29–31} Such analyses lend credibility to the estimates.

While several systematic reviews of equity-efficiency trade-off studies have been published,^{9,11–13,32} to our knowledge, there has been no systematic review that has examined the validity, reliability or framing effects of equity-efficiency trade-off studies. This review aims to fill this important gap in the literature. If the results of equity-efficiency trade-off studies are to be used in health resource allocation decisions, policymakers need to know if the elicited inequity aversion parameters have been robustly estimated. This study will also provide important

methodological insights to improve the design and conduct of future equity-efficiency trade-off studies.

Methods

This systematic review was designed and reported according to the Preferred Reporting Items for Systematic reviews and Meta-analyses (PRISMA) 2020 guidelines.³³

Eligibility Criteria

Study inclusion criteria were determined prior to study screening. We included studies that were: (1) peer-reviewed; (2) elicited value judgements and preferences of individuals with respect to health system allocation decisions or asked individuals to make a health-related equity-efficiency trade-off; (3) included equity or equality as a dimension of choice; and (4) were in English. All stated preference studies were eligible regardless of the elicitation technique. Studies that were theoretical in nature, did not consider health, did not include equity or equality as a dimension of choice, or did not include trade-offs were excluded. There were no restrictions placed on the year of publication, study sample, or study setting. Peer-reviewed original research was included, and conference abstracts were excluded.

Information Sources

Three databases were searched on December 4th, 2023: Ovid MEDLINE (1946 – 04/12/2023), Ovid EMBASE (1947 – 04/12/2023), and Web of Science (1900 – 04/12/2023). Searches were updated on January 12th, 2024. Reference lists of systematic reviews of equity-efficiency trade-off studies were manually searched to identify potentially eligible studies.^{11–13}

Search Strategy

The search strategy was designed to capture health-related equity-efficiency trade-off studies by combining three concepts: *health AND preference elicitation study AND equity or equality*. Terms for each of these three concepts were identified by consulting previous systematic reviews of equity-efficiency trade-off studies.¹¹⁻¹³ No limits were placed on date or language. The sensitivity of the search strategy was validated by ensuring that studies included in previously published systematic reviews were captured by the search strategy. The full search strategies can be found in Appendix A.

Study Selection

Records identified from the searches were uploaded to Covidence for screening. Two reviewers (VC, AC, QHL, CKB) independently screened all titles and abstracts, followed by full-text review. Conflicts were resolved by a third reviewer (AC, QHL, CKB). Two reviewers (VC, AC, QHL) independently extracted data from each included study. Discrepancies were discussed with a third reviewer (VC, AC, QHL).

Data Items and Outcomes

We assessed the proportion of equity-efficiency trade-off studies that evaluated validity, reliability, and/or framing effects. Reviewers then identified how these were assessed, and what the findings were. If a study provided quantitative values for validity, reliability, or framing effects, these values were extracted. If a study discussed them narratively, this information was synthesized.

Data were collected for the following study characteristics in studies that assessed validity, reliability, or framing effects: elicitation method; sample size; geographic location; type of

sample (e.g., general public, policymakers, healthcare professionals, etc.); and survey administration method (e.g., face-to-face interview, online survey, etc.).

Study Quality Assessment

Study quality was assessed using the PREFS checklist, which is suitable for assessing data quality of preference elicitation studies.³⁴ The checklist assesses five study domains: purpose; respondent sampling; explanation of methods; findings; and significance testing. Each item is scored 0 (unacceptable) or 1 (acceptable) and each study receives a score from 0 to 5.³⁵ Quality assessment was conducted for studies that assessed validity, reliability, and/or framing effects, independently by two reviewers (VC, AC, QHL), and discrepancies were resolved with a third reviewer (VC, AC, QHL). A narrative synthesis was conducted to describe the reporting of validity, reliability, and framing effects in the equity-efficiency trade-off literature.

Results

The database searches returned 5279 results (1273 MEDLINE, 1424 EMBASE, 2582 Web of Science) and hand searching of reference lists returned three additional citations. After deduplication, 3240 titles and abstracts were screened. The full texts of 200 articles were reviewed, and 115 studies were included. The most common reason for exclusion was the study not being a preference elicitation study ($n = 39$). The study identification, screening, and inclusion process is summarized in a PRISMA flow diagram (Figure 1).³³

Study design characteristics of the studies that assessed validity, reliability, or framing effects are summarized in Table 1. Of the 115 included studies, 29 (25.2%) assessed validity, reliability, or framing effects.^{14,26,30,36–61} The other 86 included studies did not assess or discuss them.^{15,16,29,31,62–143}

14 studies assessed validity (12.2%), 9 studies assessed reliability (7.8%), and 10 assessed framing or cognitive effects (8.7%).

Table 2 presents the characteristics of validity, reliability, and framing effects as assessed in equity-efficiency trade-off studies.

Assessment of Validity

Fourteen studies assessed validity. Many studies did not explicitly specify the type of validity they assessed, although their assessment of validity could be aligned with definitions of face validity, convergent validity, or external validity. Several studies ascertained the face validity of their measurement instruments by comparing their results with *a priori* expectations or with

previously published literature or policy,^{50,53,55,60,61} by conducting a pilot study to ascertain participant understanding,⁴⁹ or by asking participants to verbally explain their reasoning.³⁷ In all five studies that had *a priori* expectations for allocation preferences, researchers reported that results were in the hypothesized direction, suggesting good face validity. In the study that asked participants' to explain their choices as a measure of face validity, investigators reported high validity because participants' choices on the questionnaire were highly consistent with the verbal statements they made to explain their reasoning.³⁷ Jehu-Appiah and colleagues assessed convergent validity by comparing the rank ordering from their main DCE exercise to the rank ordering produced from a simpler DCE exercise conducted on a subset of the sample. They found a strong correlation (Spearman rank order correlation = 0.79, $p = 0.004$) between the two DCE exercises.³⁹ In a different study, researchers measured validity by giving participants the option to 'opt out' when faced with a difficult allocation decision.⁴⁶ Only 5% of participants decided to opt out, suggesting that they were not avoiding difficult decisions and that their stated preferences reflected their true beliefs, indicating good validity.⁴⁶ External validity was discussed in two equity-efficiency trade-off studies. Comerford and colleagues assessed external validity by comparing individuals' preferences in an abstract scenario with their preferences in a real-world COVID-19 pandemic scenario. In the real-world scenario, participants were less likely to 'level down' and choose as if an additional year of life has negative utility if offered to the most privileged.³⁶ This discrepancy in preferences between real-world and abstract resource allocation scenarios suggests that abstract scenarios may not be externally valid in eliciting individuals' true preferences. Li and colleagues discussed external validity by referencing previous literature; they stated that their method, which measures

behaviour in the laboratory, correlates well with behaviour outside of the laboratory.⁴⁴ In the equity-efficiency studies that assessed validity, investigators did not explicitly reference monotonicity or compensatory preferences. Overall, most equity-efficiency studies reported acceptable validity of their measurement instruments.

Assessment of Reliability

Of the nine studies that assessed reliability, five assessed reliability generally,^{36,43,48,51,59} three assessed test-retest reliability,^{30,38,58} and one assessed version consistency.⁵⁴ In three of the studies assessing reliability, researchers constructed indices to determine participants' value orientations, such as a dispositional optimism index, with the goal of including participants' scores on the index as a covariate in the model of resource allocation preferences. To assess the reliability of these indices, researchers used Cronbach's alpha, with two studies reporting good internal consistency of their indices^{48,51} and one study reporting modest internal consistency.³⁶ To assess test-retest reliability, researchers provided repeated choice tasks³⁰ or a repeated question at the end of the questionnaire.⁵⁸ In another study, a subsample of individuals was asked to repeat the survey one week after the initial survey.³⁸ Generally, researchers reported high test-retest reliability. For example, in one study, 81.1% of respondents made the same choices when provided with a repeated question,⁵⁸ and in another study, there was strong or excellent agreement between the values provided at the initial test and at retest.³⁸ To assess version consistency, van Exel and colleagues administered their survey both online and in-person two weeks apart and compared the results from both modes of administration.⁵⁴ They stated that the comparability of results from both modes of

administration was good.⁵⁴ Level recoding was not explicitly mentioned in any of the equity-efficiency trade-off studies that discussed reliability.

Assessment of Framing and Cognitive Effects

Ten studies assessed framing or cognitive effects. Half of these studies varied their question or attribute framing,^{14,26,38,41,45} and others assessed order effects by presenting questions in different orders.^{47,56} Two studies used a pilot study or focus groups to rule out framing and order effects prior to the main investigation.^{40,43} Five studies did not observe any significant framing or order effects.^{14,38,43,47,56} Conversely, three studies found that respondents' choices were affected by question framing or order.^{26,41,45} In one study, investigators attempted to reduce pro-egalitarian cognitive bias by randomizing participants to receive an e-learning intervention designed to prompt participants to think about questions in a more complete manner prior to taking the questionnaire.⁴² Respondents that received the e-learning intervention still displayed high levels of inequality aversion, but there was a substantial new minority of non-egalitarian responses.⁴² In another study, Ali and colleagues tested four different cognitive biases based on question framing. They found that presenting small vs. unrealistically large inequality reductions in their hypothetical scenarios, and population- vs. individual-level descriptions did not affect health inequality aversion.¹⁴ The presentation of concrete scenarios instead of abstract scenarios, and an online administration format led to some reductions in health inequality aversion, but the median respondent was still inequality averse.¹⁴ In summary, while some studies reported that there were no major framing effects present in their study, others reported that framing effects significantly affected respondents'

choices, and others reported only a modest framing effect. As such, it is unclear to what extent framing effects influence the results of equity-efficiency trade-off studies.

Other Methods of Assessing Robustness

In addition to validity, reliability, and framing effects, we identified other methods that equity-efficiency trade-off studies used to assess the robustness of their results. Several studies reported sensitivity analyses as a method of checking robustness.^{16,29,61} Three studies assessed robustness by repeating their analyses after excluding a certain group of participants. One study excluded participants who showed no real preferences in the choice task.²⁹ Investigators in this study reported that their results were affected by excluding participants who showed no real preference.²⁹ Another study repeated analyses after excluding participants who may not have understood or engaged with the survey, found the survey difficult, took a long or short time to complete the survey, or selected the same response for all questions.³¹ This study found that excluding certain participants affected the magnitude of all calculated coefficients, but this effect was only statistically significant for a small number of models and the impact was not systematic.³¹ A third study excluded participants who reported a lower clarity or certainty score for the choice task, took a long or short time to complete the task, or stated a lower willingness-to-pay for larger quality of life gains.⁶¹ Other equity-efficiency trade-off studies assessed robustness by including individuals with inconsistent responses³⁰ and control participants.^{36,44} These studies found that results were robust to the inclusion of inconsistent responders and controls. One study assessed whether median inequality aversion parameters were sensitive to larger sample sizes and found that these parameters were indeed robust to

sample size.¹⁶ In general, most studies reported that their results were robust to the inclusion or exclusion of certain participants, variations in methods, and differences in sample size.

The results of study quality assessment using the PREFS checklist can be found in Table 3.

Discussion

This systematic review investigated whether and how equity-efficiency trade-off studies assess validity, reliability, or framing effects. Of the 115 included studies, only 29 (25.2%) did so. Less than 15% of the studies assessed validity and less than 10% of the studies assessed reliability or framing effects. In the studies that assessed them, investigators used a wide array of quantitative and qualitative assessment methods. Face validity was commonly assessed by comparing results to hypothesized expectations. Convergent validity was assessed in one study where researchers provided a simple version of the exercise to a subsample of participants and compared the results of the simple exercise to those of the main choice experiment. External validity was assessed in one study by comparing abstract and real-world scenarios, and in another study by referencing previously published literature. Level recoding and compensatory preferences were not explicitly referenced in any of the studies that discussed validity or reliability. Researchers assessed test-retest reliability and version consistency by offering a repeated choice task or by offering a different version of the questionnaire. Reliability was also assessed by determining the internal consistency of indices used to ascertain participants' value orientations. Generally, equity-efficiency trade-off studies reported good reliability and validity, with fewer studies reporting low validity and reliability. Multiple studies reported no framing and cognitive effects, while others reported that their results were significantly affected by framing and cognitive effects. Studies that used other methods for assessing the robustness of their results, such as sensitivity analyses, reported that their results were robust to variations in method or exclusion of certain respondents.

To our knowledge, this is the first systematic review to discuss the validity, reliability, and framing effects of equity-efficiency trade-off studies. Several review papers have assessed the validity of DCEs and other preference elicitation techniques. Merlo and colleagues conducted a validity assessment of DCEs of primary healthcare professionals, with specific assessment of internal and external validity.¹⁷ A meta-analysis by Quaife and colleagues assessed external validity of health-related DCEs to determine how well DCEs predict health choices in real life.¹⁴⁴ Other reviews have investigated validity and concordance of DCEs and best-worst scaling methods in health,³⁴ participant understanding in health-related DCEs,¹⁴⁵ and external validity of DCEs in health economics.¹⁴⁶ A systematic review by Ryan and colleagues described preference elicitation techniques in the health literature, including their validity, reliability, and generalizability.¹⁴⁷ Although these reviews discuss the validity and reliability of preference elicitation techniques, they review health-related DCEs more broadly, without a specific focus on equity as a dimension of choice.

It is unsurprising that few equity-efficiency trade-off studies report validity, reliability, and framing effects. A previous systematic review of self-report research utilization measures in healthcare found that only 33% of studies assessed reliability.¹⁴⁸ The authors of this review noted that there is significant underdevelopment in psychometric assessment.¹⁴⁸ Although equity-efficiency trade-off studies are not psychometric assessments, these conclusions from the assessment of psychometric literature suggest that validity and reliability may be even more underreported in preference elicitation choice experiments, where validity and reliability are not traditionally systematically assessed. The findings of our systematic review support this

conclusion. It appears that validity and reliability are underreported across health research disciplines.

Previous studies have reported mixed findings about the validity and reliability of preference elicitation techniques. A systematic review and meta-analysis of external validity of DCEs found that DCEs have moderate, but not exceptional, accuracy in predicting health choices.¹⁴⁴

Conversely, a preference elicitation study for health states that used visual analogue scale, time trade-off, and standard gamble methods found that these methods had poor test-retest reliability and construct validity.¹⁴⁹ Another study showed that DCEs are able to accurately predict health choices, mimicking real-world decisions, if scale and preference heterogeneity are considered.¹⁵⁰ Given these differing findings in the health economics literature, it is unsurprising that we found varying reports of validity and reliability in equity-efficiency trade-off studies, with some studies reporting high or moderate validity and reliability and other studies reporting poor validity and reliability. These conflicting findings may be due to differing preference elicitation methods, study samples, or other context-specific factors.

Only a small proportion of equity-efficiency trade-off studies assessed or discussed the validity, reliability and framing effects of their measurement instruments. This finding is concerning, given that the purpose of equity-efficiency trade-off studies is to inform high-impact health policy and health resource allocation decisions. Researchers in other fields have noted that reporting validity and reliability is necessary to evaluate the usefulness of measurement instruments and meaningfully apply results to clinical practice and policy.¹⁵¹ Establishing the validity and reliability of stated preference methods is important if the results of studies are to be used to inform health policy.¹⁴⁶ Assessment of validity and reliability lends credibility to the

method used.¹⁵² The results of these studies must be interpreted critically and cautiously if there is no such assessment.

Researchers should assess validity and reliability every time they are conducting a new equity-efficiency trade-off study, even if they are using a preference elicitation technique for which validity and reliability have previously been established. Validity and reliability are not fixed and depend on study context, study population, and other factors.²³ Although several studies have assessed the validity of common preference elicitation techniques such as discrete choice experiments,^{17,144} these assessments are not specific to the equity-efficiency trade-off. When conducting an equity-efficiency trade-off study, it is not sufficient to cite previous studies and suggest that all discrete choice experiments are valid. The validity and reliability of preference elicitation techniques will vary greatly depending on the context in which they are applied.

A limitation of this systematic review is that the assessment methods of included studies were heterogeneous. Studies assessed validity and reliability using different methods and reported a wide range of qualitative and quantitative measures. Moreover, studies varied widely in their preference elicitation techniques, with methods ranging from discrete choice experiments to person trade-off to integrated citizens juries and others. This heterogeneity of study methods precluded a meta-analysis or a direct comparison of validity and reliability across equity-efficiency trade-off studies. This review serves as a descriptive paper to outline the current state of validity and reliability assessment and reporting in the equity-efficiency literature.

Another limitation of this systematic review is that study reviewers were not blinded to study outcomes. As such, a selection bias is possible if reviewers were more inclined to include a

study if it mentioned validity or reliability in the title or abstract. Although such a selection bias is possible, it was likely mitigated due to duplicate study screening by two independent reviewers and resolution of conflicts by a third reviewer.

Currently, researchers use a wide array of preference elicitation techniques in their equity-efficiency trade-off studies, such as discrete choice experiments, person trade-off, best-worst scaling, and many others. Future studies should compare the validity and reliability of these various preference elicitation methods directly to ascertain whether certain preference elicitation methods are more valid and reliable than others in deriving preferences for equity and efficiency.

Another promising avenue for future research is determining a ‘threshold’ for validity and reliability in equity-efficiency trade-off studies, if such a threshold exists. Researchers and policy makers should consider how high validity and reliability of an equity-efficiency trade-off study need to be for this study to be considered acceptable for use in health policy decision making.

Researchers may be unaware of the need to assess the validity and reliability of their methods. As such, future work should focus on developing guidelines to support researchers in the transparent assessment and reporting of validity and reliability in their studies. Such guidelines may outline minimum standards for the reporting of validity and reliability in preference elicitation studies.

Conclusion

This systematic review described validity, reliability, and framing effects in equity-efficiency trade-off studies. Most equity-efficiency trade-off studies did not assess validity, reliability, or framing effects. The studies which assessed them presented a wide range of qualitative and quantitative metrics, with most studies reporting good validity and reliability and fewer reporting low validity and reliability. The results of equity-efficiency trade-off studies should be interpreted with caution if there is no assessment of important measures such as validity and reliability. Equity-efficiency trade-off studies should only be used to inform high-impact health policy decisions if authors demonstrate acceptable validity and reliability.

Figures and Tables

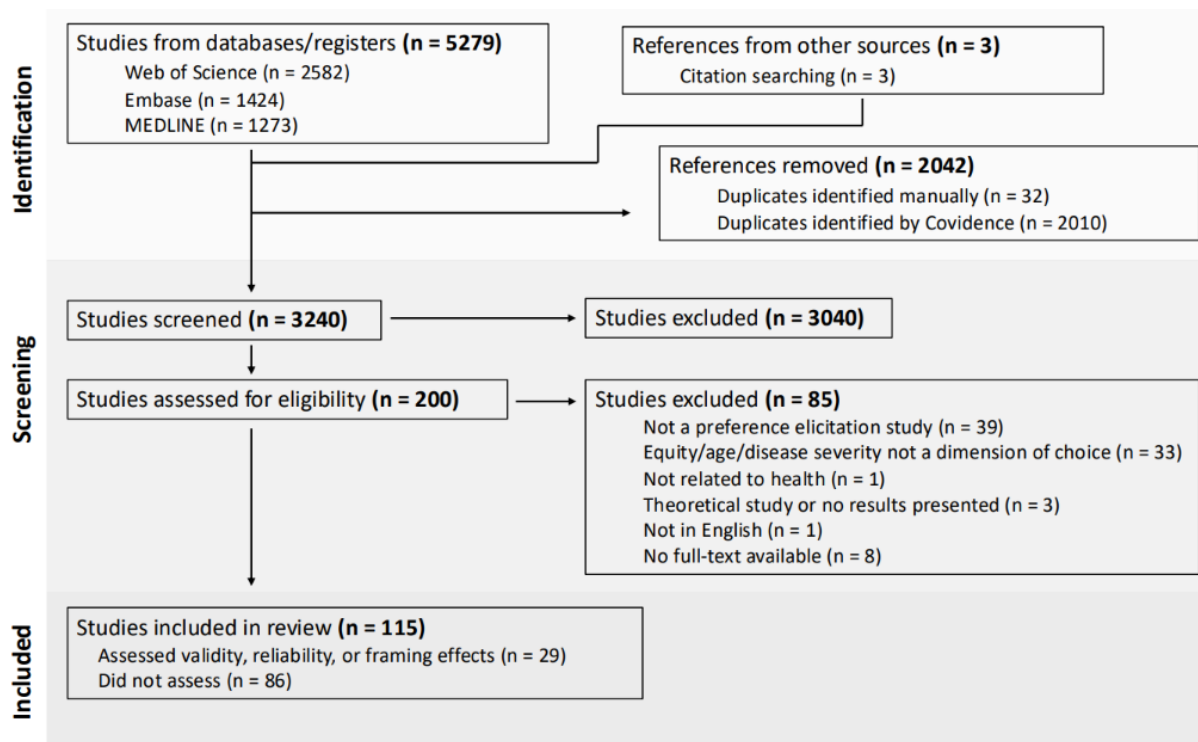


Figure 1. PRISMA flow diagram

Table 1. Study design characteristics of studies that assessed validity, reliability, or framing effects.

Study	Preference elicitation method	Sample size	Geographic location	Type of sample	Administration format	Property evaluated
Gulácsi 2012	Discrete choice experiment	200	Hungary	General practitioners	Face-to-face interviews and online questionnaires	Validity
Attema 2022	Person trade-off	500	The Netherlands	National sample	Online	Reliability
Stolk 2005	Priority ranking experiment	65	The Netherlands	Students, researchers, health policy makers	Not stated	Framing effects and reliability
Ratcliffe 2000	Conjoint analysis	303	Britain	Local sample of university employees	Questionnaire sent by mail	Validity
Werner 2009	Survey	624	North Israel	Local sample of community-dwelling adults	Face-to-face interview	Reliability
Luyten 2019	Discrete choice experiment	750	Belgium	National sample	Survey (not stated whether online or in-person)	Reliability
Damschroder 2004	Person trade-off	95	Philadelphia, USA	Local sample (prospective jurors)	Computer-administered survey and face-to-face interview	Reliability and validity
Jehu-Appiah 2008	Discrete choice experiment	63	Ghana	Policymakers	In-person survey	Validity
Reckers-Droog 2021	Willingness-to-pay	2023	The Netherlands	National sample	Online	Validity
Ubel 1996 (a)	Choice experiment	169	Pennsylvania, US	Local sample of jurors	Paper questionnaire	Validity
Whitty 2014	Discrete choice experiment and	930	Queensland, Australia	Local sample	Online	Reliability

	best-worst scaling					
Lancsar 2011	Discrete choice experiment	587	England	National sample of the general population	Face-to-face using Computer Assisted Personal Interview (CAPI)	Framing effects
Petrou 2013	Person trade-off	2500	United Kingdom	National sample	Online survey	Reliability and framing effects
Aidem 2017	Focus groups and semi-structured interviews	27	Norway	Hospital administrators, policymakers, practitioners, seniors, and university students	In-person	Validity
Whitty 2015	Discrete choice experiment	1994	Queensland and South Australia	Local sample	Online survey	Face validity
Cookson 2018	Questionnaire with pairwise choices	60	York, England	Local sample	Face-to-face	Framing and cognitive effects
Ahlert 2017	Questionnaire	166	Germany	Local sample of university students	In-person	Validity
Ali 2017	Choice experiment	135	York, England	Local sample	Online and face-to-face	Framing/cognitive effects
Li 2022	Questionnaire	1862	United States	Physicians, compared with the general public, an 'elite' subsample of Americans, and a nationwide sample of medical students	Online survey	Validity
Ubel 2001	Questionnaire	615	Philadelphia, USA	Local sample of randomly selected jurors	In-person survey	Framing effects
Ubel 1999	Choice experiment	479	Philadelphia, US	Local sample of jurors	In-person survey	Framing effects

Abásolo 2013	Choice experiment	1211	Spain	National sample	Face-to-face interview	Framing effects
Schoon 2022	Integrated citizens jury and discrete choice experiment	27	Taiwan	Local sample	Face-to-face	Validity
McKie 2019	Small-group discussion and interview	66	Victoria, Australia	Local sample	Semi-structured, face-to-face, small-group discussion	Framing effects
Comerford 2023	Experiment with pairwise choices	495	United States and United Kingdom	National sample	Online survey	Reliability and validity
Baltussen 2006	Discrete choice experiment	30	Ghana	Health policy makers	In-person survey	Validity
Ubel 1996 (b)	Choice experiment	169	Pennsylvania, US	Prospective jurors	Face-to-face survey	Framing effects
Green 2009	Choice experiment	261	Southampton, England	Local sample	Face-to-face interview	Validity
van Exel 2015	Q methodology	294	Ten countries: Hungary, Palestine, Poland, Denmark, Norway, France, Sweden, the UK, the Netherlands, Spain	General public, national samples	In-person and online	Reliability

Table 2. Characteristics of validity, reliability, and framing effects in equity-efficiency trade-off studies.

Study	Type of property	How was it assessed?	Findings
Validity			
Baltussen 2006	Face validity	Researchers examined the signs of the relations elicited in the choice task to determine whether these signs were in the expected direction and followed logical and expected allocation rules. Furthermore, the results of the choice task were compared to the priorities of the international health community as a measure of face validity.	All relations had the expected sign, and the results showed a large overlap with the priorities of the international health community. Researchers stated that these observations support the face validity of the results, although they did not operationalize face validity, provide a test of statistical significance, or describe how the threshold for 'high' validity was set.
Ahlert 2017	Face validity	As a test of validity, researchers asked participants to explain their reasoning and the principles behind their allocation preferences verbally. They compared these statements with participants' choices on the preference elicitation questionnaire and presented the results in a table. Researchers compared individual's choices in the questionnaire with verbal statements meant to elucidate their reasoning. No statistical significance testing was conducted.	Researchers reported that respondents followed their verbally reported rules in their questionnaire choices consistently.
Damschroder 2004	Validity	A computer-based administration mode for the preference elicitation task was compared to a face-to-face interview format to determine the validity of the computer-based administration format with the 'gold-standard' face-to-face format. Person trade-off equivalence values were calculated for each of the six scenarios in both the online and the face-to-face administration format.	The researchers did not find evidence of respondents giving different PTO equivalence values in the computerized and face-to-face administration formats. They concluded that the computerized protocol produced results of similar quality to the face-to-face protocol.

		P-values and confidence intervals were provided.	
Jehu-Appiah 2008	Convergent validity	As a test of validity, the rank ordering derived from the main DCE experiment was compared to the rank ordering produced from a simple DCE conducted on a smaller subset of the same sample. A correlation coefficient and p-value were provided.	The results showed a strong correlation with the simple exercise (Spearman rank order correlation = 0.79, p = 0.004).
Reckers-Droog 2021	Face validity	Investigators hypothesized that willingness-to-pay would be higher for larger-sized quality of life gains and for respondents with a higher household income. Results in support of this hypothesis would indicate acceptable theoretical validity. Researchers determined β values and reported confidence intervals for these results.	Researchers stated that the results confirm the validity of the elicited WTP. Higher household income of the respondent and higher quality-of-life gains corresponded to higher willingness to pay.
Ubel 1996 (a) - <i>Distributing scarce livers: the moral reasoning of the general public</i>	Validity	Researchers compared the results of the preference elicitation task with current organ allocation policies to determine if these policies are valid. Researchers did not set any numerical thresholds for validity or describe any statistical significance tests.	Researchers concluded that "the validity of basing transplant policies, at least in part, on public values is supported by the similarities between present allocation policies and the results of this study."
Aidem 2017	Validity	"To address validity, the author interviewed stakeholders from different levels of the health system about the same questions of interest, documented all research activities for critical evaluation of the methodology and shared findings with a sample of respondents to assess the authenticity of the author's interpretations." "In addition, member checks were conducted with a sample of respondents in order to test the validity of the	Findings were valid and in line with the author's interpretations. Validity was assessed qualitatively so no numerical values, significance tests, or thresholds were provided for validity.

		descriptions and interpretations of stakeholder views.”	
Whitty 2015	Face validity	<p>“Extensive pilot testing was undertaken to confirm the face validity of the instrument, prior to main data collection.” A convenience sample of adults (n = 20) completed the survey, with qualitative exploration of understanding of the instrument along with estimation of a preliminary choice model.</p>	The pilot survey suggested that the concepts presented in the choice task were easy to understand. Face validity was not quantified or measured, but researchers stated that face validity was confirmed during pilot testing.
Li 2022	External validity	Investigators established the external validity of their experimental method by referencing their previous work and other literature. They cite other works to suggest that behaviour inside the laboratory, as measured by their method, correlates well to behaviour outside the laboratory.	The investigators stated that their experimental method is externally valid based on other literature, although they do not provide any quantitative measures of or thresholds for external validity.
Gulácsi 2012	Face validity	Internal validity of the DCE was investigated by determining the extent to which respondents’ preferences aligned with <i>a priori</i> expectations. Investigators expected respondents to prefer younger patients, diseases with high mortality and a significant effect on quality of life, and to treat patients who can recover fully from the disease and have no comorbidities. Results in alignment with these expectations would indicate validity. Researchers reported coefficients, standard errors, and p-values for each of the parameters in the choice experiment.	Researchers stated that many of the coefficients were statistically significant and had the expected sign, suggesting good face validity. No numerical value or threshold was provided for validity, other than the qualitative determination that ‘many of’ the coefficients had the expected sign.

Schoon 2022	Face validity	Coefficients, confidence intervals, and p-values were calculated for each of the attributes in the preference task. Researchers compared the signs of the coefficients with <i>a priori</i> expectations to establish the face validity of the model.	The signs of coefficients in the model were consistent with <i>a priori</i> expectations, so the researchers concluded that the model was valid, although no numerical summary value or threshold of validity was provided.
Comerford 2023	External validity	Two surveys were given - one describing an abstract allocation scenario similar to previous equity-efficiency trade-off studies, and one describing real-world COVID-19 conditions. This was meant to examine whether preferences in abstract scenarios would match the details that people would make in real life. Researchers reported p-values from two different models to determine whether the real-life COVID condition has a statistically significant effect. Researchers provided coefficients, standard errors, and p-values for each of the attributes in both experimental conditions.	In the real-world COVID-19 scenario, participants are less likely to 'level down' - i.e. choosing as if an additional year of life has negative utility if offered to the most privileged. Both models showed that the effect of the COVID-19 condition was statistically significant ($p = 0.19$, $p = 0.13$). This suggests that in real-world scenarios, participants may prefer efficiency over equity, in contrast with previous equity-efficiency studies which commonly show a strong health inequality aversion in abstract scenarios.
Ratcliffe 2000	Face validity	Results of the preference elicitation task were compared to <i>a priori</i> expectations. It was expected that individuals would allocate more donor livers to a group if they had a longer expected survival, and less livers if they had alcoholism, were older, or were a re-transplant. Results in line with these expectations would suggest face validity. For each attribute in the preference task, researchers provided coefficients from the regression model, 95% confidence intervals, and p-values.	Researchers stated there is evidence of the model's validity because all coefficients (except the re-transplant attribute) had the expected sign. Researchers did not provide a summary measure or threshold of theoretical validity.
Green 2009	Validity	Respondents were given an option to 'opt out' and 'let others choose' when faced with a difficult	Very few (5%) respondents chose to opt out, suggesting that the most common preference of dividing resources equally may

		allocation decision. Opting out would indicate that participants were avoiding difficult decisions and that their preferences may not be valid. Researchers quantified the number of participants that decided to opt out.	be a true preference, rather than participants avoiding difficult decisions. Investigators determined that this indicated validity of the preference task, although no pooled measure of validity, significance test, or threshold was provided.
Reliability			
Comerford 2023	Inter-item reliability	Investigators assessed whether participants treated equity as a sacred value. They used a sacred value scale and evaluated inter-item reliability of this scale. They included the sacred value score as a covariate to evaluate its association with participant's health inequality aversion. Cronbach's alpha was used to assess inter-item reliability of the sacred values scale. The 'desirable value' for interitem reliability was determined by consulting previous literature (Nunnally and Bernstein, 1994).	The sacred values scaled had only modest inter-item reliability. Cronbach's alpha = 0.556, and a desirable value would be 0.7.
Stolk 2005	Internal consistency	An internal consistency check was used to determine the robustness of the preferences elicited in the paired comparison task. Significance tests for the coefficient of consistence (individual consistency) and coefficient of agreement (agreement among different respondents) were performed.	The observed proportions agreed well with the expected proportions. The absolute average discrepancy was 0.048. The coefficient of consistence showed a mean value of 0.947 ($p < 0.001$), indicating a high level of consistency. The coefficient of agreement was high (0.721, $p < 0.001$), suggesting that participants were consistent among each other. No set thresholds were used to ascertain 'high' consistency and agreement.
Werner 2009	Internal reliability	Researchers constructed three indexes to reflect participants' value orientations with the goal of determining the association between values and preferences. Cronbach's alpha was calculated for each of the three value indexes to assess reliability of each index. Researchers did not specify a	The first index had excellent internal reliability (alpha = 0.84), the second had very good internal reliability (alpha = 0.76), the third had good internal reliability (alpha = 0.74).

		threshold which represented 'excellent' or 'very good' reliability.	
Luyten 2019	Internal consistency	<p>Researchers used the Revised Life Orientation Test (LOT-R) to determine participants' dispositional optimism and then measure the association between dispositional optimism and preferences. An internal consistency check was performed on the LOT-R to assess the reliability of obtained dispositional optimism scores. Cronbach's alpha was calculated for the life orientation, optimism, and pessimism variables. The association between these dispositions and priority setting was then calculated.</p>	<p>The internal consistency of the three variables was acceptable [Cronbach's α = 0.75 (life orientation), 0.78 (pessimism) and 0.72 (optimism)].</p>
Damschroder 2004	Internal consistency	<p>Researchers assessed three measures of internal consistency in subjects' responses to the person trade-off task: ordinal consistency, reference group size consistency, and triad consistency.</p> <p>Ordinal consistency was determined by measuring whether subjects' equivalence ranking of the severity of the three health conditions presented to them followed a logical order.</p> <p>Reference group size consistency was measured by determining whether elicitation mode influenced the degree of inconsistency with respect to reference group size. 3 reference group size consistency ratios were computed for each subject by dividing the PTO equivalence value for the tradeoff when 100 people were in the reference group by the PTO equivalence value for the</p>	<p>Only 5% of responses violated criteria for ordinality, and the odds of having violations in ordinality did not differ between the computer and face-to-face conditions. Both groups displayed significant bias with respect to reference group size consistency. The two elicitation modes did not differ in triad inconsistency.</p>

		<p>tradeoff when 10 people were in the reference group.</p> <p>Triad consistency was determined to elicit the effects of triad measures on outcomes by subtracting the expected log of the PTO comparison from the actual log of the PTO comparison and dividing by the log of the worst possible deviation.</p>	
Whitty 2014	Test-retest reliability	<p>A repeated question was presented at the end of the Discrete Choice and the Best-Worst Scaling tasks. Researchers reported the percentage of respondents that gave consistent responses in each task but did not provide a measure of statistical significance. No threshold was provided for what constituted an acceptable level of consistency.</p>	<p>Consistent responses were given to the DCE repeat choice task by 75.7% of respondents. Consistent responses were given to the BWS repeat choice task by 64.5% for the most preferred attribute level, 49.4% for the least preferred attribute level, and 35.5% for both the most and least preferred. Researchers cited past work that stated that inconsistent responders are not necessarily irrational, and thus retained the inconsistent responders in the model.</p>
Petrou 2013	Test-retest reliability	<p>The first 500 individuals (one fifth of the sample) were asked to complete a re-test survey with identical questions, in the same order and using the same online format, approximately one week after the first survey. These individuals did not have access to their previous responses. The test-retest reliability was calculated using the intraclass correlation coefficient.</p>	<p>There was strong agreement (0.7-0.8) between values at test and retest for 3 of 8 person trade-off questions, and excellent agreement (>0.8) for 5 of 8 person trade-off questions. Researchers cited published literature to determine the thresholds for poor, fair, moderate, strong, and excellent agreement.</p>
Attema 2022	Test-retest reliability	<p>A repeated question was presented at the end of the questionnaire to see if participants would choose the same option as a test of reliability. No measure of statistical significance was calculated. No further detail was provided.</p>	<p>81.1% of the subjects made the same choice in the repeated questions. Researchers stated that this indicated a high level of reliability, although no justification was provided for what constitutes a 'high' level of reliability.</p>
van Exel 2015	Version consistency	<p>Materials were designed such that questions could be administered in</p>	<p>Researchers stated that the "comparability of results from both modes of</p>

		person or online. A repeated in-person and online interview pilot study was conducted two weeks apart to determine the consistency of the two methods of administration. No further detail was provided.	administration was good ($\rho=0.8$).” A coefficient was provided although the authors did not explain what this coefficient was or how it was derived. No justification was provided for what constituted a ‘good’ threshold of consistency.
<i>Framing Effects</i>			
Lancsar 2011	Question formats, graphical presentation of attributes	Several focus groups were conducted to determine the effect of question framing. The manners of question presentation & framing included: stating quality of life using EQ5D; ‘table-style’ approaches such as using a visual analogue scale to represent quality of life; explaining QALYs to respondents; and testing QALYs gained as an attribute.	Researchers did not provide a quantitative or qualitative measure of framing effects. However, they stated that they experimented with question framing in the focus groups to determine the best framing for each question.
Petrou 2013	Framing effect	“To assess whether estimated age-related weights for health gains were influenced by the framing of the trade-off questions, the reference number of individuals in each age group was set at 100 for one half of the study participants and at 1000 for the other half of study participants.” The researchers used Student’s t-test to determine whether there was a significant difference in the age-related weights when the number of individuals in the reference group varied.	The t-test revealed no significant differences in the age-related weights between the two reference groups.
Cookson 2018	"Pro-strict egalitarian" cognitive bias	Participants were randomized to receive an e-learning intervention before taking the questionnaire to allow the respondents to think about questions more completely. The first e-learning intervention presented a video animation debate to simulate tradeoffs. The second e-learning intervention was a spreadsheet-based interactive	Respondents that received the e-learning intervention still had a high level of health inequality aversion but there was a substantial new minority of non-egalitarian responses. These results suggest that learning interventions may affect health inequality aversion.

		questionnaire to provide feedback on the implications of each choice. Investigators reported model coefficients for each attributed, along with standard errors and p-values.	
Ali 2017	Cognitive biases based on question framing	Investigators tested four different cognitive biases based on question framing and presentation: (1) small vs. unrealistically large health inequality reductions (2) population-level vs. individual-level descriptions of health inequality reduction (3) concrete vs. abstract intervention scenarios (4) online vs. face-to-face mode of administration. Investigators determined the percent difference in the percentage of respondents that were strong egalitarian and non-egalitarian across the various conditions, with p-values. They also performed a Wilcoxon equality test.	Small vs. unrealistically large inequality reductions and population- vs. individual-level descriptions did not significantly affect health inequality aversion. Concrete scenarios and online format led to some reductions in health inequality aversion, but the median respondent was still health inequality averse.
Ubel 2001	Framing effects	Investigators varied attributes of the questions and scenarios to determine if this would affect participants' preferences in the equity-efficiency trade-off. They varied the number of people who could receive a test, the number of lives which would be saved, the proportion of Medicaid enrollees who could receive the test, whether scenarios involved Medicaid enrollees from the same or different states, and the order in which questions were presented. P-values were reported.	Individuals' preferences varied significantly depending on the ways in which scenarios were framed. Preference for equity was stronger when the choice covered the entire population than half the population. Preferences were significantly susceptible to order effects. Preferences for equity and efficiency varied when an identical scenario was framed in different ways.
Ubel 1999	Framing effects - question wording	Minor changes were made to question wording, by clarifying the scenario, prompting subjects to think about self-interest, and giving subjects the explicit vs. no explicit	Less subjects prioritized severely ill patients when they were given the explicit option to divide resources evenly. Researchers reported that minor changes in wording

		option to divide resources evenly. Researchers determined whether these minor changes in framing affected participants preferences. They used a Sign test and reported p-values for each framing change.	significantly affect people's preferences for prioritizing severely ill patients.
Abásolo 2013	Titration effects (based on the order in which the questions are presented, and the 'gap' between the options)	Some participants were presented with random question order, others with a 'titrated' question order, with gradually increasing differences between each question. To test for two framing issues, experimenters presented titrated questions to determine whether respondents could delay switching their choice to the point where they are indifferent to either choice. Investigators calculated p-values to determine whether titration effects significantly affected respondents' choices.	No significant framing effect was found ($p > 0.05$), so investigators reported that titration effects do not significantly affect participants' choices.
Stolk 2005	Framing effects based on presentation of health states	A pilot study was conducted among 20 students to determine whether the use of varying labels and descriptors in the scenarios introduced biases or framing effects in the observed rank order.	The pilot study confirmed that the presentation of labels and health state descriptors did not affect the observed rank order. No quantitative measures of framing effects or significance tests were provided.
McKie 2019	Order effects	Health states were presented in different orders. Participants were asked to partake in a critical self-examination process designed to prompt them to think beyond their initial reactions, with the goal of eliminating order effects. Investigators reported the proportions of participants choosing the health maximizing option, confidence intervals, and p-values.	Unlike previous studies, this study did not observe order effects. Order effects were eliminated by giving participants the opportunity to directly compare the pre-existing and onset cases, encouraging in-depth discussion of the issues, and ensuring participants were ethically informed.
Ubel 1996 (b) - <i>Public perceptions of the importance</i>	Framing effects	Investigators tested whether framing recipients of transplant organs as distinct groups vs. as individual recipients affected	Subjects were more inclined to use prognostic information in determining their

<i>of prognosis in allocating transplantable livers to children</i>		respondent's allocation preferences. P-values were reported for respondents' likelihood of allocating organs to various groups.	choices when patients were presented as individuals rather than groups.
Other Methods for Assessing Robustness			
Reckers-Droog 2019	Sensitivity analysis	Investigators performed sensitivity analyses where they excluded respondents who reported that the choice task was unclear or who showed no real preference on the choice task.	Investigators reported that the ratios obtained in the preference task were 'pulled downwards' by including participants who showed no real preference. Otherwise, the sensitivity analyses showed consistent results.
Whitty 2014	Sensitivity analysis	Researchers examined the performance of their models when including responders with inconsistent responses to ascertain the resilience and stability of the models to variations in respondent behavior.	Investigators stated that their "models were robust to the inclusion of inconsistent responders." Numerical data was not provided, but the authors stated that it could be obtained upon request.
Robson 2017	Sensitivity analysis	Sensitivity analyses were conducted to determine whether median inequality aversion parameters were sensitive to larger sample sizes.	Investigators stated that the obtained median inequality aversion parameters were robust to larger sample sizes. Numerical data was not provided directly in the manuscript, but the authors stated that the robustness analysis could be found in the appendix.
Rowen 2016	Sensitivity analysis	Robustness was determined by measuring the impact of excluding individuals who may not have understood or engaged with the survey, those who found the survey difficult, those who took a very long or short time to complete it, those who selected to treat the same patient group for all questions, and exclusion of the first and last question (susceptible to learning or fatigue effects). Investigators calculated coefficients for all attributes and reported significance with p-values.	Excluding certain respondents "affected the magnitude of all coefficients but only affected their significance for a small number of models containing burden of illness, where the impact was not systematic."

Ali 2017	Robustness of results to cognitive biases	Investigators tested four different cognitive biases based on question framing and presentation: (1) small vs. unrealistically large health inequality reductions (2) population-level vs. individual-level descriptions of health inequality reduction (3) concrete vs. abstract intervention scenarios (4) online vs. face-to-face mode of administration. Investigators reported results of a Wilcoxon equality test and p-values.	Investigators found that their findings of health inequality aversion were mostly robust to cognitive biases. Median respondents in all scenarios displayed substantial health inequality aversion.
Li 2022	Sensitivity analysis	Investigators tested whether results of the choice task were robust to the inclusion of controls for gender and census region.	Investigators stated that their results were robust to the inclusion of the controls, although they did not explicitly delineate any numerical values or significance tests.
Baker 2010	Robustness analysis of methods	Investigators used multiple data aggregation methods in the relativities study to determine if the results would be robust to variations in method.	Patterns in the data were robust to the choice of method. There was a general tendency to give preference to younger age groups and those with poorer health.
Comerford 2023	Sensitivity analysis	To test robustness of perceived health inequality aversion, investigators included control variables in the model and excluded respondents who believed the poor live longer than the rich. P-values were reported to determine significance.	Results were robust to the inclusion of controls (Extreme Egalitarian: $z = 2.80$, $p = 0.005$; Health Maximizer: $z = 3.30$, $p = 0.001$). Results were unchanged when dropping respondents who believed the poor live longer than the rich (p-values from both multivariate regressions <0.01).
Reckers-Droog 2021	Sensitivity analysis	Sensitivity analyses were conducted, excluding respondents who reported a lower clarity and certainty score for the choice task, who took an extremely long or short time to complete the task, and who stated a lower willingness-to-pay for larger quality of life gains. Welch's t-test was used to determine statistically significant associations.	"The sensitivity analyses indicated that respondents' stated WTP in the practice task had a marginal effect on the stated WTP in the subsequent tasks (models 1 to 7: β 0.01) and that [the] results were robust."

BWS: Best-worst scaling

DCE: Discrete choice experiment

PTO: Person trade-off

QALY: Quality-adjusted life year

WTP: Willingness-to-pay

Table 3. Study quality assessment of studies that assessed validity, reliability, or framing effects using the PREFS checklist.

Study	Purpose (0-1)	Respondents (0-1)	Explanation (0-1)	Findings (0-1)	Significance (0-1)	Total (0-5)
Werner 2009	1	0	1	1	1	4
Luyten 2019	1	0	1	1	1	4
Damschroder 2004	1	0	1	0	1	3
Jehu-Appiah 2008	1	0	1	1	1	4
Reckers-Droog 2021	1	0	1	0	1	3
Ubel 1996(a)	1	0	1	1	0	3
Whitty 2014	1	0	1	1	1	4
Lancsar 2011	1	0	1	0	1	3
Petrou 2013	1	0	1	1	1	4
Aidem 2017	1	0	1	1	0	3
Whitty 2015	1	0	1	1	1	4
Cookson 2018	1	0	1	1	1	4
Ahlert 2017	1	0	1	0	0	2
Ali 2017	1	0	1	1	1	4
Li 2022	0	1	1	0	1	3
Ubel 2001	1	0	1	0	1	3
Gulácsi 2012	1	0	1	1	1	4
Ubel 1999	1	0	1	1	1	4
Abásolo 2013	1	0	1	0	1	3
Schoon 2022	1	0	1	1	1	4
Attema 2022	1	0	1	0	1	3
Stolk 2005	1	0	1	1	1	4
McKie 2019	1	0	1	1	1	4

Comerford 2023	1	0	1	0	1	3
Baltussen 2006	1	0	1	1	1	4
Ubel 1996(b)	1	0	1	1	1	4
Ratcliffe 2000	1	0	1	0	1	3
Green 2009	1	0	1	1	1	4
van Exel 2015	1	0	1	0	1	3

References

1. Love-Koh J, Asaria M, Cookson R, Griffin S. The Social Distribution of Health: Estimating Quality-Adjusted Life Expectancy in England. *Value Health J Int Soc Pharmacoeconomics Outcomes Res.* 2015;18(5):655-662. doi:10.1016/j.jval.2015.03.1784
2. Woolf SH. Progress In Achieving Health Equity Requires Attention To Root Causes. *Health Aff (Millwood).* 2017;36(6):984-991. doi:10.1377/hlthaff.2017.0197
3. Cookson RA, Mirelman A, Asaria M, Dawkins B, Griffin S. Fairer decisions, better health for all: Health equity and cost-effectiveness analysis. *Cent Health Econ Univ York.* Published online September 2016:1-43.
4. Love-Koh J, Griffin S, Kataika E, Revill P, Sibandze S, Walker S. Methods to promote equity in health resource allocation in low- and middle-income countries: an overview. *Glob Health.* 2020;16(1):6. doi:10.1186/s12992-019-0537-z
5. Culyer T. Efficiency, equity and equality in health and health care. *Effic Equity Equal Health Health Care.* Published online December 2015:1-20.
6. Cookson R, Drummond M, Weatherly H. Explicit incorporation of equity considerations into economic evaluation of public health interventions. *Health Econ Policy Law.* 2009;4(2):231-245. doi:10.1017/S1744133109004903
7. Asaria M, Griffin S, Cookson R. Distributional Cost-Effectiveness Analysis: A Tutorial. *Med Decis Making.* 2016;36(1):8-19. doi:10.1177/0272989X15583266
8. Cookson R, Griffin S, Norheim OF, Culyer AJ. *Distributional Cost-Effectiveness Analysis: Quantifying Health Equity Impacts and Trade-Offs.* Oxford University Press; 2020.
9. Gu Y, Lancsar E, Ghijben P, Butler JR, Donaldson C. Attributes and weights in health care priority setting: A systematic review of what counts and to what extent. *Soc Sci Med.* 2015;146:41-52. doi:10.1016/j.socscimed.2015.10.005
10. Ali S, Ronaldson S. Ordinal preference elicitation methods in health economics and health services research: using discrete choice experiments and ranking methods. *Br Med Bull.* 2012;103(1):21-44. doi:10.1093/bmb/lds020
11. McNamara S, Holmes J, Stevely AK, Tsuchiya A. How averse are the UK general public to inequalities in health between socioeconomic groups? A systematic review. *Eur J Health Econ.* 2020;21(2):275-285. doi:10.1007/s10198-019-01126-2
12. Khor S, Elsis ZA, Carlson JJ. How Much Does the US Public Value Equity in Health? A Systematic Review. *Value Health.* 2023;26(3):418-426. doi:10.1016/j.jval.2022.08.009
13. Cadham CJ, Prosser LA. Eliciting Trade-Offs Between Equity and Efficiency: A Methodological Scoping Review. *Value Health.* 2023;26(6):943-952. doi:10.1016/j.jval.2023.02.006

14. Ali S, Tsuchiya A, Asaria M, Cookson R. How Robust Are Value Judgments of Health Inequality Aversion? Testing for Framing and Cognitive Effects. *Med Decis Mak Int J Soc Med Decis Mak*. 2017;37(6):635-646. doi:10.1177/0272989X17700842
15. Baker R, Bateman I, Donaldson C, et al. Weighting and valuing quality-adjusted life-years using stated preference methods: preliminary results from the Social Value of a QALY Project. *Health Technol Assess*. 2010;14(27):1-+. doi:10.3310/hta14270
16. Robson M, Asaria M, Cookson R, Tsuchiya A, Ali S. Eliciting the Level of Health Inequality Aversion in England. *Health Econ*. 2017;26(10):1328-1334. doi:10.1002/hec.3430
17. Merlo G, van Driel M, Hall L. Systematic review and validity assessment of methods used in discrete choice experiments of primary healthcare professionals. *Health Econ Rev*. 2020;10(1):39. doi:10.1186/s13561-020-00295-8
18. Cornelsen L, Quaife M, Lagarde M, Smith RD. Framing and signalling effects of taxes on sugary drinks: A discrete choice experiment among households in Great Britain. *Health Econ*. 2020;29(10):1132-1147. doi:10.1002/hec.4123
19. For better or worse? Investigating the validity of best–worst discrete choice experiments in health - Krucien - 2019 - Health Economics - Wiley Online Library. Accessed August 7, 2024. <https://onlinelibrary.wiley.com/doi/abs/10.1002/hec.3869>
20. Johnson FR, Yang JC, Reed SD. The Internal Validity of Discrete Choice Experiment Data: A Testing Tool for Quantitative Assessments. *Value Health*. 2019;22(2):157-160. doi:10.1016/j.jval.2018.07.876
21. Janssen EM, Marshall DA, Hauber AB, Bridges JFP. Improving the quality of discrete-choice experiments in health: how can we assess validity and reliability? *Expert Rev Pharmacoecon Outcomes Res*. Published online November 2, 2017. Accessed June 22, 2024. <https://www.tandfonline.com/doi/abs/10.1080/14737167.2017.1389648>
22. Ryan M, Gerard K. Using discrete choice experiments to value health care programmes: current practice and future research reflections. *Appl Health Econ Health Policy*. 2003;2(1):55-64.
23. Souza AC de, Alexandre NMC, Guirardello E de B. Psychometric properties in instruments evaluation of reliability and validity. *Epidemiol E Serv Saude Rev Sist Unico Saude Bras*. 2017;26(3):649-659. doi:10.5123/S1679-49742017000300022
24. Howard K, Salkeld G. Does Attribute Framing in Discrete Choice Experiments Influence Willingness to Pay? Results from a Discrete Choice Experiment in Screening for Colorectal Cancer. *Value Health*. 2009;12(2):354-363. doi:10.1111/j.1524-4733.2008.00417.x
25. Veldwijk J, Essers BAB, Lambooi MS, Dirksen CD, Smit HA, de Wit GA. Survival or Mortality: Does Risk Attribute Framing Influence Decision-Making Behavior in a Discrete Choice Experiment? *Value Health*. 2016;19(2):202-209. doi:10.1016/j.jval.2015.11.004
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.

27. Gigerenzer G, Hoffrage U. How to improve Bayesian reasoning without instruction: Frequency formats. *Psychol Rev.* 1995;102(4):684-704. doi:10.1037/0033-295X.102.4.684
28. Caussade S, Ortúzar J de D, Rizzi LI, Hensher DA. Assessing the influence of design dimensions on stated choice experiment estimates. *Transp Res Part B Methodol.* 2005;39(7):621-640. doi:10.1016/j.trb.2004.07.006
29. Reckers-Droog V, van Exel J, Brouwer W. Equity Weights for Priority Setting in Healthcare: Severity, Age, or Both?. *Value Health J Int Soc Pharmacoeconomics Outcomes Res.* 2019;22(12):1441-1449. doi:10.1016/j.jval.2019.07.012
30. Whitty JA, Ratcliffe J, Chen G, Scuffham PA. Australian Public Preferences for the Funding of New Health Technologies: A Comparison of Discrete Choice and Profile Case Best-Worst Scaling Methods. *Med Decis Mak Int J Soc Med Decis Mak.* 2014;34(5):638-654. doi:10.1177/0272989X14526640
31. Rowen D, Brazier J, Mukuria C, et al. Eliciting Societal Preferences for Weighting QALYs for Burden of Illness and End of Life. *Med Decis Mak Int J Soc Med Decis Mak.* 2016;36(2):210-222. doi:10.1177/0272989X15619389
32. 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
33. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Syst Rev.* 2021;10(1):89. doi:10.1186/s13643-021-01626-4
34. Whitty JA, Oliveira Gonçalves AS. A Systematic Review Comparing the Acceptability, Validity and Concordance of Discrete Choice Experiments and Best-Worst Scaling for Eliciting Preferences in Healthcare. *Patient - Patient-Centered Outcomes Res.* 2018;11(3):301-317. doi:10.1007/s40271-017-0288-y
35. Joy SM, Little E, Maruthur NM, Purnell TS, Bridges JFP. Patient Preferences for the Treatment of Type 2 Diabetes: A Scoping Review. *Pharmacoeconomics.* 2013;31(10):877-892. doi:10.1007/s40273-013-0089-7
36. Comerford D, Tufte-Hewett A, Bridger E. Public preferences to trade-off gains in total health for health equality: Discrepancies between an abstract scenario versus the real-world scenario presented by COVID-19. *Ration Soc.* Published online 2023. doi:10.1177/10434631231193599
37. Ahlert M, Schwettmann L. Allocating health care resources: a questionnaire experiment on the predictive success of rules. *Int J Equity Health.* 2017;16(1):112. doi:10.1186/s12939-017-0611-1
38. Petrou S, Kandala NB, Robinson A, Baker R. A person trade-off study to estimate age-related weights for health gains in economic evaluation. *Pharmacoeconomics.* 2013;31(10):893-907. doi:10.1007/s40273-013-0085-y
39. Jehu-Appiah C, Baltussen R, Acquah C, et al. Balancing equity and efficiency in health priorities in Ghana: the use of multicriteria decision analysis. *Value Health J Int Soc Pharmacoeconomics Outcomes Res.* 2008;11(7):1081-1087. doi:10.1111/j.1524-4733.2008.00392.x

40. 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
41. Ubel PA, Loewenstein G. Distributing scarce livers: the moral reasoning of the general public. *Soc Sci Med* 1982. 1996;42(7):1049-1055.
42. Cookson R, Ali S, Tsuchiya A, Asaria M. E-learning and health inequality aversion: A questionnaire experiment. *Health Econ*. 2018;27(11):1754-1771. doi:10.1002/hec.3799
43. Stolk EA, Pickee SJ, Ament AHJA, Busschbach JJV. Equity in health care prioritisation: an empirical inquiry into social value. *Health Policy Amst Neth*. 2005;74(3):343-355.
44. Li J, Casalino L, Fisman R, Kariv S, Markovits D. Experimental evidence of physician social preferences. *Proc Natl Acad Sci U S A*. 2022;119(28). doi:10.1073/pnas.2112726119
45. Ubel P. How stable are people's preferences for giving priority to severely ill patients? *Soc Sci Med*. 1999;49(7):895-903. doi:10.1016/S0277-9536(99)00174-4
46. Green C. Investigating public preferences on "severity of health" as a relevant condition for setting healthcare priorities. *Soc Sci Med*. 2009;68(12):2247-2255. doi:10.1016/j.socscimed.2009.03.020
47. Abásolo I, Tsuchiya A. Is more health always better for society? Exploring public preferences that violate monotonicity. *THEORY Decis*. 2013;74(4):539-563. doi:10.1007/s11238-011-9292-1
48. Werner P. Israeli lay persons' views on priority-setting criteria for Alzheimer's disease. *Health Expect Int J Public Particip Health Care Health Policy*. 2009;12(2):187-196. doi:10.1111/j.1369-7625.2008.00523.x
49. Whitty J, Ratcliffe J, Kendall E, et al. Prioritising patients for bariatric surgery: building public preferences from a discrete choice experiment into public policy. *BMJ OPEN*. 2015;5(10). doi:10.1136/bmjopen-2015-008919
50. Gulácsi L, Hajdú O. Prioritization Preferences among General Practitioners in Hungary. *EAST Eur Polit Soc*. 2012;26(1):20-32. doi:10.1177/0888325410390207
51. Luyten J, Kessels R, Desmet P, Goos P, Beutels P. Priority-Setting and Personality: Effects of Dispositional Optimism on Preferences for Allocating Healthcare Resources. *Soc JUSTICE Res*. 2019;32(2):186-207. doi:10.1007/s11211-019-00329-5
52. Ubel PA, Loewenstein G. Public perceptions of the importance of prognosis in allocating transplantable livers to children. *Med Decis Mak Int J Soc Med Decis Mak*. 1996;16(3):234-241.
53. Ratcliffe J. Public preferences for the allocation of donor liver grafts for transplantation. *Health Econ*. 2000;9(2):137-148.

54. van Exel J, Baker R, Mason H, Donaldson C, Brouwer W. Public views on principles for health care priority setting: findings of a European cross-country study using Q methodology. *Soc Sci Med* 1982. 2015;126(ut9, 8303205):128-137. doi:10.1016/j.socscimed.2014.12.023
55. Schoon R, Chi C, Liu T. Quantifying public preferences for healthcare priorities in Taiwan through an integrated citizens jury and discrete choice experiment. *Soc Sci Med*. 2022;315. doi:10.1016/j.socscimed.2022.115404
56. Mckie J, Richardson J. Social preferences for prioritising the treatment of disabled and chronically ill patients: beyond the order effect. *Health Econ POLICY LAW*. 2019;14(4):443-467. doi:10.1017/S1744133118000154
57. Aidem JM. Stakeholder views on criteria and processes for priority setting in Norway: a qualitative study. *Health Policy Amst Neth*. 2017;121(6):683-690. doi:10.1016/j.healthpol.2017.04.005
58. Attema A, Brouwer W, Pinto J. The Role of Perceived Utility of Full Health in Age Weighting. *VALUE Health*. 2022;25(9):1559-1565. doi:10.1016/j.jval.2022.04.1733
59. Damschroder LJ, Baron J, Hershey JC, Asch DA, Jepson C, Ubel PA. The validity of person tradeoff measurements: randomized trial of computer elicitation versus face-to-face interview. *Med Decis Mak Int J Soc Med Decis Mak*. 2004;24(2):170-180. doi:10.1177/0272989X04263160
60. Baltussen R, Stolk E, Chisholm D, Aikins M. Towards a multi-criteria approach for priority setting: an application to Ghana. *Health Econ*. 2006;15(7):689-696.
61. Reckers-Droog V, van Exel J, Brouwer W. Willingness to Pay for Health-Related Quality of Life Gains in Relation to Disease Severity and the Age of Patients. *Value Health J Int Soc Pharmacoeconomics Outcomes Res*. 2021;24(8):1182-1192. doi:10.1016/j.jval.2021.01.012
62. Dolan P, Cookson R. A qualitative study of the extent to which health gain matters when choosing between groups of patients. *Health POLICY*. 2000;51(1):19-30. doi:10.1016/S0168-8510(99)00079-2
63. Pinho M, Borges AP. A Scoring Index of Prioritization Factors Between Patients: An Iberian Comparison. *Health Care Manag*. 2019;38(3):267-275. doi:10.1097/HCM.0000000000000274
64. Diederich A, Winkelhage J, Wirsik N. Age as a Criterion for Setting Priorities in Health Care? A Survey of the German Public View. *PLOS ONE*. 2011;6(8). doi:10.1371/journal.pone.0023930
65. Richardson J, Mckie J, Iezzi A, Maxwell A. Age Weights for Health Services Derived from the Relative Social Willingness-to-Pay Instrument. *Med Decis Making*. 2017;37(3):239-251. doi:10.1177/0272989X16645576
66. Grepin KA, Pinkstaff CB, Hole AR, et al. Allocating external financing for health: a discrete choice experiment of stakeholder preferences. *Health Policy Plan*. 2018;33(suppl_1):i24-i30. doi:10.1093/heapol/czx017

67. Lindholm L, Rosen M, Emmelin M. An epidemiological approach towards measuring the trade-off between equity and efficiency in health policy. *Health Policy Amst Neth*. 1996;35(3):205-216.
68. Schwappach D. Are preferences for equality a matter of perspective? *Med Decis Making*. 2005;25(4):449-459. doi:10.1177/0272989X05276861
69. van de Wetering EJ, van Exel NJA, Rose JM, Hoefman RJ, Brouwer WBF. Are some QALYs more equal than others?. *Eur J Health Econ HEPAC Health Econ Prev Care*. 2016;17(2):117-127. doi:10.1007/s10198-014-0657-6
70. Li DG, Wong GX, Martin DT, et al. Attitudes on cost-effectiveness and equity: a cross-sectional study examining the viewpoints of medical professionals. *BMJ Open*. 2017;7(7):e017251. doi:10.1136/bmjopen-2017-017251
71. Quintal C. Aversion to geographic inequality and geographic variation in preferences in the context of healthcare. *Appl Health Econ Health Policy*. 2009;7(2):121-136. doi:10.2165/10899530-000000000-00000
72. Berhane Defaye F, Danis M, Wakim P, Berhane Y, Norheim OF, Miljeteig I. Bedside Rationing Under Resource Constraints-A National Survey of Ethiopian Physicians' Use of Criteria for Priority Setting. *AJOB Empir Bioeth*. 2019;10(2):125-135. doi:10.1080/23294515.2019.1583691
73. Pinho M, Borges A, Zahariev B. Bedside healthcare rationing dilemmas: A survey from Bulgaria and comparison with Portugal. *Soc THEORY Health*. 2017;15(3):285-301. doi:10.1057/s41285-017-0029-2
74. Diederich A, Swait J, Wirsik N. Citizen Participation in Patient Prioritization Policy Decisions: An Empirical and Experimental Study on Patients' Characteristics. *PLOS ONE*. 2012;7(5):e36824. doi:10.1371/journal.pone.0036824
75. Howard K, Jan S, Rose J, et al. Community Preferences for the Allocation & Donation of Organs - The PArADOx Study. *BMC PUBLIC Health*. 2011;11. doi:10.1186/1471-2458-11-386
76. Browning C, Thomas S. Community values and preferences in transplantation organ allocation decisions. *Soc Sci Med*. 2001;52(6):853-861. doi:10.1016/S0277-9536(00)00187-8
77. Chim L, Salkeld G, Kelly PJ, Lipworth W, Hughes DA, Stockler MR. Community views on factors affecting medicines resource allocation: cross-sectional survey of 3080 adults in Australia. *Aust Health Rev Publ Aust Hosp Assoc*. 2019;43(3):254-260. doi:10.1071/AH16209
78. Baji P, Garcia-Goni M, Gulacsi L, Mentzakis E, Paolucci F. Comparative analysis of decision maker preferences for equity/efficiency attributes in reimbursement decisions in three European countries. *Eur J Health Econ HEPAC Health Econ Prev Care*. 2016;17(7):791-799. doi:10.1007/s10198-015-0721-x
79. Goto R, Mori T. Comparison of Equity Preferences for Life Expectancy Gains: A Discrete Choice Experiment Among the Japanese and Korean General Public. *Value Health Reg Issues*. 2019;18(101592642):8-13. doi:10.1016/j.vhri.2018.05.004

80. Youngkong S, Baltussen R, Tantivess S, Koolman X, Teerawattananon Y. Criteria for priority setting of HIV/AIDS interventions in Thailand: a discrete choice experiment. *BMC Health Serv Res.* 2010;10. doi:10.1186/1472-6963-10-197
81. Mirelman A, Mentzakis E, Kinter E, et al. Decision-making criteria among national policymakers in five countries: a discrete choice experiment eliciting relative preferences for equity and efficiency. *Value Health J Int Soc Pharmacoeconomics Outcomes Res.* 2012;15(3):534-539. doi:10.1016/j.jval.2012.04.001
82. Ratcliffe J, Buxton M, Young T, Longworth L. Determining priority for liver transplantation: a comparison of cost per QALY and discrete choice experiment-generated public preferences. *Appl Health Econ Health Policy.* 2005;4(4):249-255.
83. Tsuchiya A, Dolan P. Do NHS clinicians and members of the public share the same views about reducing inequalities in health?. *Soc Sci Med 1982.* 2007;64(12):2499-2503.
84. Shmueli A. Do the equity-efficiency preferences of the Israeli Basket Committee match those of Israeli health policy makers?. *Isr J Health Policy Res.* 2017;6(101584158):20. doi:10.1186/s13584-017-0145-4
85. Kolasa K, Lewandowski T. Does it matter whose opinion we seek regarding the allocation of healthcare resources? - a case study. *BMC Health Serv Res.* 2015;15(101088677):564. doi:10.1186/s12913-015-1210-8
86. Denburg A, Ungar W, Chen S, Hurley J, Abelson J. Does moral reasoning influence public values for health care priority setting?: A population-based randomized stated preference survey. *Health POLICY.* 2020;124(6):647-658. doi:10.1016/j.healthpol.2020.04.007
87. McNamara S, Tsuchiya A, Holmes J. Does the UK-public's aversion to inequalities in health differ by group-labelling and health-gain type? A choice-experiment. *Soc Sci Med 1982.* 2021;269(ut9, 8303205):113573. doi:10.1016/j.socscimed.2020.113573
88. Shmueli A, Golan O, Paolucci F, Mentzakis E. Efficiency and equity considerations in the preferences of health policy-makers in Israel. *Isr J Health Policy Res.* 2017;6(101584158):18. doi:10.1186/s13584-017-0142-7
89. Norman R, Hall J, Street D, Viney R. Efficiency and equity: a stated preference approach. *Health Econ.* 2013;22(5):568-581. doi:10.1002/hec.2827
90. Li M, Colby H, Fernbach P. Efficiency for Lives, Equality for Everything Else: How Allocation Preference Shifts Across Domains. *Soc Psychol Personal Sci.* 2019;10(5):697-707. doi:10.1177/1948550618783709
91. Lim MK, Bae EY, Choi SE, Lee EK, Lee TJ. Eliciting public preference for health-care resource allocation in South Korea. *Value Health J Int Soc Pharmacoeconomics Outcomes Res.* 2012;15(1 Suppl):S91-4. doi:10.1016/j.jval.2011.11.014

92. Gibbs N, Powell PA, Tsuchiya A. Equal access for equal need: Eliciting public preferences for access to health treatment by employment status. *Soc Sci Med* 1982. 2019;222(ut9, 8303205):246-255. doi:10.1016/j.socscimed.2019.01.003
93. Paolucci F, Mentzakis E, Defechereux T, Niessen LW. Equity and efficiency preferences of health policy makers in China--a stated preference analysis. *Health Policy Plan*. 2015;30(8):1059-1066. doi:10.1093/heapol/czu123
94. Mentzakis E., Garcia-Goni M., Sequeira A.R., Paolucci F. Equity and efficiency priorities within the Spanish health system: A discrete choice experiment eliciting stakeholders preferences. *Health Policy Technol*. 2019;8(1):30-41. doi:10.1016/j.hlpt.2019.01.003
95. Lal A, Mohebi M, Sweeney R, Moodie M, Peeters A, Carter R. Equity Weights for Socioeconomic Position: Two Methods-Survey of Stated Preferences and Epidemiological Data. *Value Health J Int Soc Pharmacoeconomics Outcomes Res*. 2019;22(2):247-253. doi:10.1016/j.jval.2018.07.006
96. van den Broek-Altenburg EM, Benson JS, Atherly AJ. Examining alignment of community health teams' preferences for health, equity, and spending with state all-payer waiver priorities: A discrete choice experiment. *Health Serv Res*. 2023;(g2l, 0053006). doi:10.1111/1475-6773.14257
97. Ratcliffe J, Bekker HL, Dolan P, Edlin R. Examining the attitudes and preferences of health care decision-makers in relation to access, equity and cost-effectiveness: a discrete choice experiment. *Health Policy Amst Neth*. 2009;90(1):45-57. doi:10.1016/j.healthpol.2008.09.001
98. McHugh N, Pinto-Prades JL, Baker R, Mason H, Donaldson C. Exploring the relative value of end of life QALYs: Are the comparators important?. *Soc Sci Med* 1982. 2020;245(ut9, 8303205):112660. doi:10.1016/j.socscimed.2019.112660
99. Green C, Gerard K. EXPLORING THE SOCIAL VALUE OF HEALTH-CARE INTERVENTIONS: A STATED PREFERENCE DISCRETE CHOICE EXPERIMENT. *Health Econ*. 2009;18(8):951-976. doi:10.1002/hec.1414
100. Lindholm L.A., Emmelin M.A., Rosen M.E. Health maximization rejected: The view of Swedish politicians. *Eur J Public Health*. 1997;7(4):405-410. doi:10.1093/eurpub/7.4.405
101. Dolan P, Tsuchiya A. Health priorities and public preferences: the relative importance of past health experience and future health prospects. *J Health Econ*. 2005;24(4):703-714. doi:10.1016/j.jhealeco.2004.11.007
102. Franken M, Koolman X. Health system goals: a discrete choice experiment to obtain societal valuations. *Health Policy Amst Neth*. 2013;112(1-2):28-34. doi:10.1016/j.healthpol.2012.12.013
103. Lindholm L, Rosen M, Emmelin M. How many lives is equity worth? A proposal for equity adjusted years of life saved. *J Epidemiol Community Health*. 1998;52(12):808-811.
104. Steinert JI, Sternberg H, Veltri GA, Buthe T. How should COVID-19 vaccines be distributed between the Global North and South: a discrete choice experiment in six European countries. *eLife*. 2022;11(101579614). doi:10.7554/eLife.79819

105. Hadian M, Jouyani Y, Asadi H, Palangi H, Rahimnia R. Identifying the Criteria Affecting Appropriate Allocation of Health System Resources to Different Diseases in Iran: A Qualitative Inquiry. *Health SCOPE*. 2019;8(4). doi:10.5812/jhealthscope.83968
106. Yang F., Katumba K.R., Greco G., et al. INCORPORATING CONCERN FOR HEALTH EQUITY INTO RESOURCE ALLOCATION DECISIONS: DEVELOPMENT OF A TOOL AND POPULATIONBASED VALUATION FOR UGANDA. *BMJ Glob Health*. 2022;7(Supplement 2):A13. doi:10.1136/bmjgh-2022-ISHP.36
107. Pinho M, Botelho A. Inference Procedures to Quantify the Efficiency-Equality Trade-Off in Health from Stated Preferences: A Case Study in Portugal. *Appl Health Econ Health Policy*. 2018;16(4):503-513. doi:10.1007/s40258-018-0394-6
108. Gyrd-Hansen D. Investigating the social value of health changes. *J Health Econ*. 2004;23(6):1101-1116.
109. 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
110. Ubel P, Richardson J, Prades J. Life-saving treatments and disabilities - Are all QALYs created equal? *Int J Technol Assess Health CARE*. 1999;15(4):738-748. doi:10.1017/S0266462399154138
111. Ottersen T, Maestad O, Norheim OF. Lifetime QALY prioritarianism in priority setting: quantification of the inherent trade-off. *Cost Eff Resour Alloc CE*. 2014;12(1):2. doi:10.1186/1478-7547-12-2
112. Youngkong S, Teerawattananon Y, Tantivess S, Baltussen R. Multi-criteria decision analysis for setting priorities on HIV/AIDS interventions in Thailand. *Health Res Policy Syst*. 2012;10(101170481):6. doi:10.1186/1478-4505-10-6
113. Perneger T, Martin D, Bovier P. Physicians' attitudes toward health care rationing. *Med Decis Making*. 2002;22(1):65-70. doi:10.1177/02729890222062928
114. Jelsma J, Shumba D, Kristian H, De Weerd W, De Cock P. Preferences of urban Zimbabweans for health and life lived at different ages. *Bull World Health Organ*. 2002;80(3):204-209.
115. Erdem S, Thompson C. Prioritising health service innovation investments using public preferences: a discrete choice experiment. *BMC Health Serv Res*. 2014;14. doi:10.1186/1472-6963-14-360
116. Olsen JA. Priority preferences: "end of life" does not matter, but total life does. *Value Health J Int Soc Pharmacoeconomics Outcomes Res*. 2013;16(6):1063-1066. doi:10.1016/j.jval.2013.06.002
117. Otim ME, Kelaher M, Anderson IP, Doran CM. Priority setting in Indigenous health: assessing priority setting process and criteria that should guide the health system to improve Indigenous Australian health. *Int J Equity Health*. 2014;13(101147692):45. doi:10.1186/1475-9276-13-45

118. Meusel V, Mentzakis E, Baji P, Fiorentini G, Paolucci F. Priority setting in the German healthcare system: results from a discrete choice experiment. *Int J Health Econ Manag.* 2023;23(3):411-431. doi:10.1007/s10754-023-09347-y
119. Wouters S, van Exel J, Baker R, B F Brouwer W. Priority to End of Life Treatments? Views of the Public in the Netherlands. *Value Health J Int Soc Pharmacoeconomics Outcomes Res.* 2017;20(1):107-117. doi:10.1016/j.jval.2016.09.544
120. Lee JS, Kim S, Do YK. Public Preferences for Allocation Principles for Scarce Medical Resources in the COVID-19 Pandemic in Korea: Comparisons With Ethicists' Recommendations. *J Prev Med Public Health Yebang Uihakhoe Chi.* 2021;54(5):360-369. doi:10.3961/jpmp.21.333
121. Luyten J, Kessels R, Goos P, Beutels P. Public preferences for prioritizing preventive and curative health care interventions: a discrete choice experiment. *Value Health J Int Soc Pharmacoeconomics Outcomes Res.* 2015;18(2):224-233. doi:10.1016/j.jval.2014.12.007
122. Bae EY, Lim MK, Lee B, Bae G, Hong J. Public preferences in healthcare resource allocation: A discrete choice experiment in South Korea. *Health Policy Amst Neth.* 2023;138(8409431, hep):104932. doi:10.1016/j.healthpol.2023.104932
123. Luyten J, Dorgali V, Hens N, Beutels P. Public preferences over efficiency, equity and autonomy in vaccination policy: An empirical study. *Soc Sci Med.* 2013;77:84-89. doi:10.1016/j.socscimed.2012.11.009
124. Darvishi A, Daroudi R, Yaseri M, Sari AA. Public preferences regarding the priority setting criteria of health interventions for budget allocation: results of a survey of Iranian adults. *BMC Public Health.* 2022;22(1):2038. doi:10.1186/s12889-022-14404-1
125. Blacksher E, Rigby E, Espey C. Public values, health inequality, and alternative notions of a "fair" response. *J Health Polit Policy Law.* 2010;35(6):889-920. doi:10.1215/03616878-2010-033
126. Luyten J, Beutels P, Vandermeulen C, Kessels R. Social preferences for adopting new vaccines in the national immunization program: A discrete choice experiment. *Soc Sci Med 1982.* 2022;303(ut9, 8303205):114991. doi:10.1016/j.socscimed.2022.114991
127. McKie J, Richardson J. Social preferences for prioritizing the treatment of severely ill patients: The relevance of severity, expected benefit, past health and lifetime health. *Health Policy Amst Neth.* 2017;121(8):913-922. doi:10.1016/j.healthpol.2017.05.010
128. Chim L, Salkeld G, Kelly P, Lipworth W, Hughes DA, Stockler MR. Societal perspective on access to publicly subsidised medicines: A cross sectional survey of 3080 adults in Australia. *PloS One.* 2017;12(3):e0172971. doi:10.1371/journal.pone.0172971
129. Skedgel C, Wailoo A, Akehurst R. Societal preferences for distributive justice in the allocation of health care resources: a latent class discrete choice experiment. *Med Decis Mak Int J Soc Med Decis Mak.* 2015;35(1):94-105. doi:10.1177/0272989X14547915

130. Linley WG, Hughes DA. Societal views on NICE, cancer drugs fund and value-based pricing criteria for prioritising medicines: a cross-sectional survey of 4118 adults in Great Britain. *Health Econ*. 2013;22(8):948-964. doi:10.1002/hec.2872
131. Arroyos-Calvera D, Covey J, Loomes G, McDonald R. The efficiency-equity trade-off, self-interest, and moral principles in health and safety valuation. *Soc Sci Med* 1982. 2019;238(ut9, 8303205):112477. doi:10.1016/j.socscimed.2019.112477
132. Skedgel C. The prioritization preferences of pan-Canadian Oncology Drug Review members and the Canadian public: a stated-preferences comparison. *Curr Oncol Tor Ont*. 2016;23(5):322-328.
133. Bae E.Y., Lim M.K., Choi S.E., Lee T.J. The public's preference on the priorities in health care. *Value Health*. 2010;13(7):A534.
134. Mason H, Baker R, Donaldson C. Understanding public preferences for prioritizing health care interventions in England: does the type of health gain matter? *J Health Serv Res POLICY*. 2011;16(2):81-89. doi:10.1258/jhsrp.2010.010039
135. Ratcliffe J, Lancsar E, Walker R, Gu Y. Understanding what matters: An exploratory study to investigate the views of the general public for priority setting criteria in health care. *Health Policy Amst Neth*. 2017;121(6):653-662. doi:10.1016/j.healthpol.2017.03.003
136. Nicolet A, van Asselt A, Vermeulen K, Krabbe P. Value judgment of new medical treatments: Societal and patient perspectives to inform priority setting in The Netherlands. *PLOS ONE*. 2020;15(7). doi:10.1371/journal.pone.0235666
137. Shah KK, Tsuchiya A, Wailoo AJ. Valuing health at the end of life: a stated preference discrete choice experiment. *Soc Sci Med* 1982. 2015;124(ut9, 8303205):48-56. doi:10.1016/j.socscimed.2014.11.022
138. 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
139. Lal A, Siahpush M, Moodie M, Peeters A, Carter R. Weighting Health Outcomes by Socioeconomic Position Using Stated Preferences. *PharmacoEconomics - Open*. 2018;2(1):43-51. doi:10.1007/s41669-017-0036-1
140. Card K, Adshade M, Hogg R, Jollimore J, Lachowsky N. What public health interventions do people in Canada prefer to fund? A discrete choice experiment. *BMC PUBLIC Health*. 2022;22(1). doi:10.1186/s12889-022-13539-5
141. Jain V, Atun R, Hansen P, Lorgelly P. Which countries need COVID-19 vaccines the most? Development of a prioritisation tool. *BMC Public Health*. 2022;22(1):1518. doi:10.1186/s12889-022-13948-6
142. Bae EY, Lim MK, Lee B, Bae G. Who should be given priority for public funding?. *Health Policy Amst Neth*. 2020;124(10):1108-1114. doi:10.1016/j.healthpol.2020.06.010

143. Reckers-Droog V, van Exel J, Brouwer W. Who should receive treatment? An empirical enquiry into the relationship between societal views and preferences concerning healthcare priority setting. *PLoS One*. 2018;13(6):e0198761. doi:10.1371/journal.pone.0198761
144. Quaipe M, Terris-Prestholt F, Di Tanna GL, Vickerman P. How well do discrete choice experiments predict health choices? A systematic review and meta-analysis of external validity. *Eur J Health Econ*. 2018;19(8):1053-1066. doi:10.1007/s10198-018-0954-6
145. Pearce A, Harrison M, Watson V, et al. Respondent Understanding in Discrete Choice Experiments: A Scoping Review. *The Patient*. 2021;14(1):17-53. doi:10.1007/s40271-020-00467-y
146. Lancsar E, Swait J. Reconceptualising the External Validity of Discrete Choice Experiments. *Pharmacoeconomics*. 2014;32(10):951-965. doi:10.1007/s40273-014-0181-7
147. Ryan M, Scott DA, Reeves C, et al. Eliciting public preferences for healthcare: a systematic review of techniques. *Health Technol Assess Winch Engl*. 2001;5(5):1-186. doi:10.3310/hta5050
148. Squires JE, Estabrooks CA, O'Rourke HM, Gustavsson P, Newburn-Cook CV, Wallin L. A systematic review of the psychometric properties of self-report research utilization measures used in healthcare. *Implement Sci*. 2011;6(1):83. doi:10.1186/1748-5908-6-83
149. Crump RT, Lau R, Cox E, Currie G, Panepinto J. Testing the feasibility of eliciting preferences for health states from adolescents using direct methods. *BMC Pediatr*. 2018;18(1):199. doi:10.1186/s12887-018-1179-7
150. de Bekker-Grob EW, Swait JD, Kassahun HT, et al. Are Healthcare Choices Predictable? The Impact of Discrete Choice Experiment Designs and Models. *Value Health*. 2019;22(9):1050-1062. doi:10.1016/j.jval.2019.04.1924
151. Zangaro GA. Importance of Reporting Psychometric Properties of Instruments Used in Nursing Research. *West J Nurs Res*. 2019;41(11):1548-1550. doi:10.1177/0193945919866827
152. Louviere JJ, Lancsar E. Choice experiments in health: the good, the bad, the ugly and toward a brighter future. *Health Econ Policy Law*. 2009;4(Pt 4):527-546. doi:10.1017/S1744133109990193