Prediction accuracy of discrete choice experiments in health-related research: a systematic review and meta-analysis

Ying Zhang,[a,b] Thi Quynh Anh Ho,[c] Fern Terris-Prestholt,[d] Matthew Quaife,[e] Esther de Bekker-Grob,[f,g,h] Peter Vickerman,[i] and Jason J. Ong[a,b,j,*]

[a]School of Translational Medicine, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Victoria, Australia
[b]Melbourne Sexual Health Centre, Alfred Health, Melbourne, Victoria, Australia
[c]Deakin Health Economics, Institute for Health Transformation, School of Health and Social Development, Faculty of Health, Deakin University, Melbourne, Victoria, Australia
[d]Warwick Medical School, University of Warwick, Coventry, United Kingdom
[e]Patient Centered Research, Evidera, London, United Kingdom
[f]Erasmus School of Health Policy & Management, Erasmus University, Rotterdam, the Netherlands
[g]Erasmus Choice Modelling Centre, Erasmus University, Rotterdam, the Netherlands
[h]Erasmus Centre for Health Economics Rotterdam, Erasmus University, Rotterdam, the Netherlands
[i]Bristol Medical School, University of Bristol, Bristol, United Kingdom
[j]London School of Hygiene and Tropical Medicine, London, United Kingdom

eClinicalMedicine
2025;79: 102965
Published Online xxx
https://doi.org/10.1016/j.eclinm.2024.102965

Summary

Background Discrete choice experiments (DCEs) are increasingly used to inform the design of health products and services. It is essential to understand the extent to which DCEs provide reliable predictions outside of experimental settings in real-world decision-making situations. We aimed to compare the prediction accuracy of stated preferences with real-world choices, as modelled from DCE data.

Methods We searched six databases for health-related studies that used DCE to assess external validity and reported on predicted versus real-world choices, up to July 2024. A generalised linear mixed model was used for a meta-analysis to jointly pool the sensitivity and specificity. Heterogeneity was assessed using the I^2 statistic, and sources of heterogeneity using meta-regression. This study is registered with PROSPERO (CRD42023451545).

Findings We identified 14 relevant studies, of which 10 were included in the meta-analysis. Most studies were conducted in high-income countries (11/14, 79%) from the European region (9/14, 64%) and analysed using mixed logit models (5/14, 36%). Pooled sensitivity and specificity estimates were 89% (95% CI:77–95, I^2 = 97%) and 52% (95% CI:32–72, I^2 = 95%), respectively. The area under the SROC curve (AUC) was 0.81 (95% CI:0.77–0.84). Our meta-regression found that DCEs for prevention-related choices had higher sensitivity than treatment-related choices. DCEs conducted under clinical settings and analysed using the heteroskedastic multinomial logit model, incorporating systematic preference heterogeneity and random opt-out utility, had higher specificity than non-clinical settings and alternative models.

Interpretation DCEs are valuable for capturing health-related preferences and possess reasonable external validity to predict health-related behaviours, particularly for opt-in choices. Contextual factors (e.g., type of intervention, study setting, analysis method) influenced the predictive accuracy.

Funding JJO is supported by an Australian National Health and Medical Research Council Emerging Leadership Investigator Grant (GNT1193955). EBG is supported by the Dutch Research Council (NWO-Talent-Scheme-Vidi-Grant No, 09150171910002). YZ is supported by an Australian Government Research Training Program (RTP) scholarship.

Keywords: Discrete choice experiment; External validity; Predictive accuracy; Stated preferences; Revealed preferences

*Corresponding author. Melbourne Sexual Health Centre, 580 Swanston Street, Carlton, Victoria 3053, Australia.
E-mail address: Jason.ong@monash.edu (J.J. Ong).

Research in context

Evidence before this study

To identify review papers on prediction accuracy of discrete choice experiments (DCE) in health-related research, we searched PubMed in English on September 13, 2023, using the terms "discrete choice experiment*" AND "external validity" AND "health" with no date restrictions. There was only one systematic review by Quaife et al. which captured the literature up to August 2015 and summarised six papers using meta-analysis. Results from that study showed that pooled sensitivity and specificity estimates were 88% (95% CI: 81–92%) and 34% (95% CI: 23–46%), respectively, and the area under the summary receiver operating characteristic curve (AUC) was 0.60 (95% CI: 0.55–0.64). As the field of health research continues to evolve, an increasing number of studies incorporate external validity evaluations, this study updates and expands upon existing evidence to capture recent developments and advancements.

Added value of this study

The aim of our study is to conduct a systematic review of health-related studies comparing stated preferences, as estimated through predicted probability models resulting from DCE data, to revealed preferences gathered through observational or survey means. We conducted a comprehensive search of six databases (Ovid Medline, Embase, PsycINFO, CINAHL, EconLit, and Web of Science) on July 23, 2024, to explore factors affecting DCE accuracy. Of the 14 included studies, we found six additional studies, beyond those reviewed by Quaife, that compared stated with revealed preferences (e.g., colorectal cancer screening, influenza vaccination). Most studies were conducted in high-income countries (11/14, 79%), primarily in Europe (9/14, 64%) and analysed using mixed logit models (5/14, 36%). Ten of 14 papers contained sufficient data for meta-analysis, revealing a pooled sensitivity and specificity of 89% (95% CI: 77–95, I^2 = 97%) and 52% (95% CI: 32–72, I^2 = 95%). The AUC was 0.81 (95% CI: 0.77–0.84) indicating reasonable accuracy of DCEs in predicting health-related choices. Meta-regression showed that DCEs focused on prevention (e.g., vaccinations) had higher sensitivity than treatment-related choices (e.g., treatment for sleep apnoea). DCEs conducted in clinical settings (e.g., hospitals) had higher specificity than non-clinical settings (e.g., community). Further, DCEs analysed using the heteroskedastic multinomial logit (HMNL) model, incorporating systematic preference heterogeneity or random opt-out utility, had higher specificity than alternate models. For studies with aggregate-level prediction data, we used narrative synthesis to provide an overview, focusing on how DCEs were conducted and their main results.

Implications of all the available evidence

Consistent with past findings, our study demonstrates that DCEs are valuable for capturing health-related preferences and possess reasonable external validity to predict health-related behaviours, particularly for opt-in choices (i.e., someone will uptake a product or service). Analysis method and contextual factors such as the type of intervention and the setting in which choices are presented influenced the predictive accuracy of DCEs. DCEs have the potential to guide resource allocation, prioritise interventions, and enhance patient-centred care, shaping the future of evidence-based healthcare decision-making.

Introduction

Discrete choice experiments (DCEs) are a valuable instrument in health research for understanding individual preferences and guiding decision-making in complex healthcare settings.[1] DCEs can be focused on better understanding normative decision-making (such as regulatory benefit risk) and used to help people form preferences and/or aid in shared decision making. Alternatively, DCEs can be used to predict preferences and behaviours such as willingness to pay, predicted uptake rates of healthcare interventions, trade-offs that individuals are willing to make, and overall utility scores for different health states. In addition to predictive validity for forecasting demand, DCEs serve a broader purpose by offering rich insights into the multidimensional nature of preferences. For instance, DCEs can provide valuable insights by identifying latent classes and exploring preference heterogeneity, revealing distinct subgroups with shared preferences. By correlating these subgroups with external factors—such as demographic, behavioural, or clinical characteristics — further validation and interpretation of these preference structures can be achieved. Such applications enhance the value of DCEs beyond mere choice modelling, supporting an understanding of the acceptability and desirability of services, exploring benefit-risk trade-offs, and fostering patient-centred decision-making in various contexts. In DCEs, respondents are presented with hypothetical scenarios representing the attributes of services or goods and asked to choose their preferred scenario. Resulting choice data are then analysed using choice models to determine the relative importance of different attributes to respondent choices. By quantifying and analysing preferences, results from choice data can inform different healthcare attributes (e.g., cost, mode of delivery (oral medication versus injection), frequency of treatment), services, and policies – for example, health technology assessments and regulatory risk–benefit assessments.[2,3] DCEs are not solely valuable for predicting real-world behaviour; they also offer critical insights in areas where revealed preferences fall short, particularly when studying non-market goods, hypothetical scenarios, or new product attributes. One particularly useful feature of DCEs is their ability to

include products or attributes that do not yet exist in reality or are challenging to implement within a clinical trial[1] or identify preferences in situations where a good or service is not yet available. By capturing underlying preference structures, DCEs enable researchers to assess normative trade-offs and explore decisions essential for policy and strategic development. This broader application supports informed decision-making in contexts where real-world data are either unavailable or limited in scope.

Despite the growing popularity of DCEs, concerns persist regarding the external validity (i.e., predictive accuracy) of their findings – that is, how well DCE predicts choice behaviours (e.g., decision to adopt a product or service, or decision not to use a product or service) beyond the specific experimental setting.[4] Prediction accuracy is one way to test the level of external validity. Assessing external validity is crucial to examine if the insights derived from DCEs are reliable and applicable to real-world situations, and can robustly generate findings to inform healthcare policy and practice. Because respondents are not obligated to follow through with their choices in the DCE, the potential for hypothetical bias exists, which may diminish the usefulness of DCE results. While significant contributions have been made to the methodological literature on DCEs, the focus has been primarily on summarising DCE applications,[2,5,6] collating preference research on a particular health or disease area,[7,8] or synthesising methodological innovations to maximise internal validity.[9,10] Although these aspects are important, there has been a scarcity of empirical research evaluating whether the choices made in DCEs accurately reflect real-life decisions or the situations in which they may provide more or less reliable predictions. In a meta-analysis by Quaife et al., collating six studies up to August 2015, DCEs have a pooled sensitivity and specificity of 88% and 34%, respectively.[11] Nevertheless, the analysis was hindered by the scarcity of studies, resulting in an underpowered assessment that could not discern the specific scenarios in which DCEs exhibited superior or inferior performance. As the field of health research continues to evolve, with an increasing number of studies incorporating evaluations of external validity, it is imperative to update and expand upon existing knowledge to capture recent developments and advancements. Moreover, as more relevant studies are published, a better understanding of what factors may affect the prediction accuracy of DCEs is valuable for enhancing the relevance and generalisability of DCE findings in informing decisions that impact public health and individual patient care.

This systematic review aims to assess the prediction accuracy of DCEs in the literature through synthesizing data from health-related studies comparing stated preferences, as modelled through predicted probability models resulting from DCE data, to revealed preferences, as gathered through observational or survey means, and to assess the prediction accuracy of DCEs in the literature.

Methods

This review follows the recommendations in the Cochrane Handbook for Systematic Reviews of Interventions[12] and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for reporting.[13] This study is registered with PROSPERO (CRD42023451545).

Search strategy and selection criteria

A comprehensive electronic search of six databases (Ovid Medline, Embase, PsycINFO, CINAHL, EconLit, and Web of Science) was conducted on July 23, 2024. The search strategy was built around overarching terms, including 'discrete choice*', 'external validity', and 'choice behaviour [MeSH]' (Supplementary Table S1) and was adapted for each database. The search included all available years up to July 2024, to include any paper that might have been missed in Quaife et al.'s review. No language restrictions were set, and non-English studies were translated using Google Translate. In addition, a supplementary manual search of references cited by the included studies was completed to locate relevant papers missed by the search strategy. Studies were included if they reported primary data on the assessment of external validity using a DCE methodology and reported on predicted versus actual health-related choices. Papers that used a preference elicitation method other than a DCE or reported choices unrelated to health were excluded. Editorial letters without primary data, reviews, and protocol papers were excluded. Search results from each database were downloaded into the Covidence (Veritas Health Innovation, Australia) systematic review tool. After removing duplicates, two researchers (YZ and AH) independently screened the titles and abstracts of all articles potentially eligible for full-text retrieval, whilst a third researcher (JO) resolved any discrepancies.

Data extraction

Next, two independent reviewers (YZ and AH) extracted data from full-text articles that satisfied the inclusion criteria using a data extraction spreadsheet, which a third reviewer (JO) checked. Data collected from each study included the study design (e.g., number of attributes, number of choice tasks); population characteristics; study setting; type of intervention (e.g., screening, treatment); DCE design; use of pictures and icon arrays to communicate attribute levels (e.g., pictures, icon arrays); analysis method (e.g., HMNL model, mixed logit model); sample size; survey mode (e.g., paper-based, online); number of true and false positives; number of true and false negatives; and sensitivity and specificity, where available.

Data analysis

Given that DCE predictions are a type of binary classification test, we employed the methodologies used in assessing clinical diagnostic tests to synthesise the outcomes of the included studies. In the context of DCE predictions, high sensitivity (true-positive rate) indicates reliability in predicting opting-in behaviours. We defined an opting-in behaviour as an act of making a deliberate choice, either in a DCE or in a real-life scenario, to adopt a product or service not currently used by the respondent. High specificity (true-negative rate) indicates reliability in predicting opting-out behaviours, which we have defined as a respondent's decision not to use a product or service in the DCE or real-world context.

Statistical analyses were conducted in Stata 17 (StataCorp LP, USA) using the *midas* command. A generalised linear mixed model was used for a meta-analysis to jointly pool the sensitivity and specificity, which assumes independent binomial distributions for true positives and negatives conditional on the sensitivity and specificity in each study.[14,15] This method preserves the bivariate data structure by jointly modelling sensitivity and specificity. Studies that investigated multiple treatments or diseases were extracted and analysed as distinct data sets. Statistical heterogeneity was assessed using the I^2 statistic, and Deeks' test[16] was used to assess publication bias. We calculated the positive and negative likelihood ratios. We used random-effects meta-regression to determine whether any study-level covariates could explain the between-study heterogeneity. For studies that did not report individual-level data, despite our efforts to obtain it from the original authors, we used narrative synthesis to provide an overview, focusing on how DCEs were conducted and their main results.

Quality assessment

The studies included in the final analysis were evaluated using the PREFS checklist.[17] Quality assessment was performed by two independent researchers (YZ and AH), and discrepancies were resolved by a third reviewer (JO).

Ethical considerations

Ethical approval was not required as this is a systematic review.

Role of funding source

The funder of the study had no involvement in the study's design, data collection, data analysis, data interpretation, or the writing of the report.

Results

A total of 9414 publications were initially identified, and after reviewing the full texts of 25 studies for eligibility, 14 were deemed relevant for inclusion, of which eight were included in Quaife's review.[11] Four studies contained aggregate level information and were considered for a qualitative synthesis,[18–21] while ten presented information at the individual level and were included in the meta-analysis[22–31] (Fig. 1). Table 1 summarises all the selected studies, and Table 2 presents the data extracted to assess the predictive ability of DCEs. Most studies were conducted in high-income countries (11/14, 79%) from the European region (9/14, 64%), using paper-based survey administration (8/14, 67%; Supplementary Figs. S1 and S2). The main method of analysis was mixed logit models (5/14, 36%). Supplementary Fig. S3 shows the number of included studies published over the years. Additional details regarding the number of choice tasks in each DCE study are presented in Supplementary Fig. S4. Five (36%) of the DCE questionnaires featured more than five attributes, and 12 studies included an opt-out option as an alternative. Two studies[20,30] examined marginal rates of substitutions, but none examined the external validity of the marginal rates of substitution (e.g., trade-offs between attributes, or willingness to pay). The studies reviewed varied considerably regarding the methods adopted for choice analyses.

Quality assessment

Supplementary Table S2 shows the results from the quality assessment of included studies. All studies had a well-defined research question and clearly stated the purpose of the study in relation to preferences. None of the studies reported on the difference between responders and non-responders. Methods of assessing preferences were also clearly explained in all the studies. Thirty-six percent of the studies (5/14) included all respondents who completed the preference questions in the analysis. The remaining nine studies excluded respondents that had missing values in the survey or in the actual healthcare utilisation data. All studies used significance tests to assess preference results.

Individual-level data

We found that the pooled sensitivity and specificity estimates were 89% (95% confidence interval (CI): 77–95) and 52% (95% CI: 32–72), respectively (Fig. 2). Predictions showed higher sensitivity than specificity, indicating that DCEs are more effective in forecasting individuals who would opt-into a health-related decision rather than those who would not. High heterogeneity was observed for sensitivity (I^2 = 97%) and specificity (I^2 = 95%), suggesting substantial between-study variation. Supplementary Fig. S5 visually depicts the summary receiver operating characteristic (SROC) curve, illustrating the relationship between the true- and false-positive rates across studies using least squares regression. The area under the SROC curve (AUC) was 0.81 (95% CI: 0.77–0.84), indicating that DCEs possess

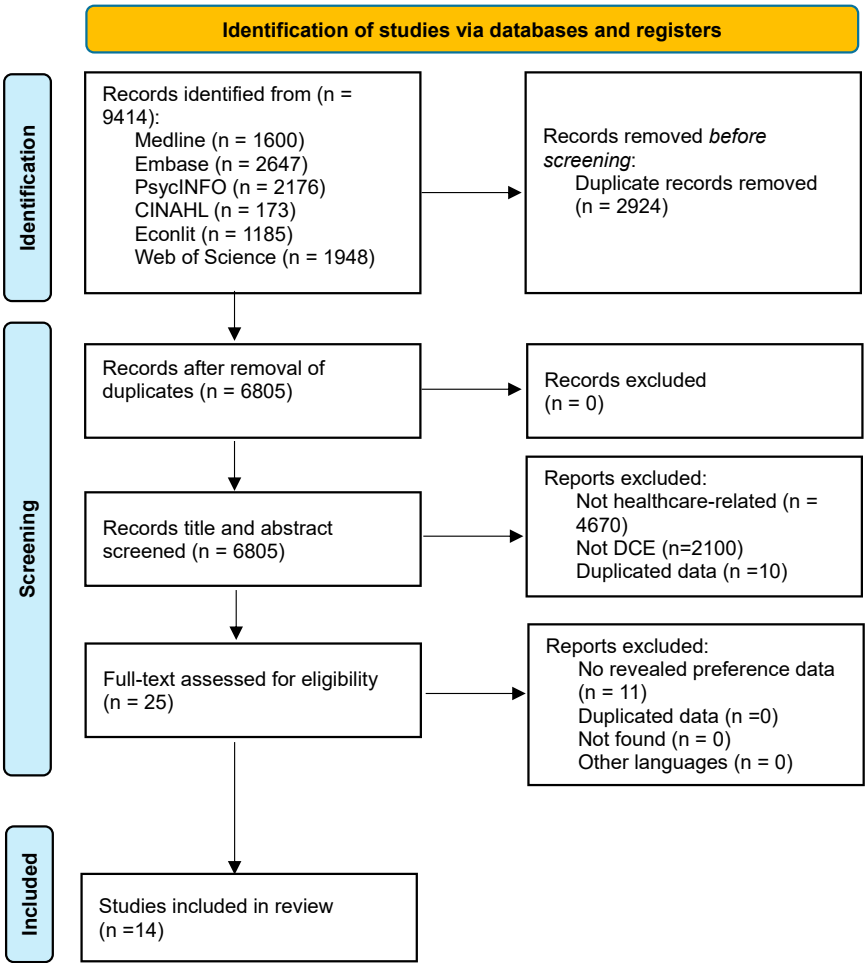

Fig. 1: PRISMA flow diagram.

reasonable ability to predict choices.[32] Our meta-regression analysis found that DCEs for prevention-related choices (n = 8 outcomes; e.g., vaccination, screening) had higher sensitivity compared to treatment-related choices (n = 4 outcomes; e.g., oral appliances for patients with obstructive sleep apnoea syndrome, alcohol brief intervention). Furthermore, DCEs conducted in clinical settings had higher specificity than non-clinical ones. Furthermore, DCEs analysed using the heteroskedastic multinomial logit (HMNL) model, incorporating systematic preference heterogeneity and random opt-out utility, also had higher specificity than alternative models (Fig. 3 and Supplementary Table S3). We also discovered no statistically significant difference in DCEs examining a single occurrence of action compared to a prolonged commitment demanded from an individual within the intervention (p = 0.59). However, as the assessment of real-world decisions was not followed-up over time and continuous actions are generally not captured, there may be changes in actual behaviour over time or attrition.

In the context of bivariate random-effects models, the pooled estimates of sensitivity and specificity are expected to be interdependent. Supplementary Fig. S6 presents a bivariate boxplot that illustrates the relationship between sensitivity and specificity. The inner oval depicts the median distribution of estimates, while the outer oval represents the 95% confidence interval. These results suggest some variation between the selected studies, with four studies falling outside the average range and one exhibiting characteristics of an outlier. No clear skew towards either sensitivity or specificity was observed. As part of the sensitivity analysis, the outlier study was removed, and the pooled sensitivity and specificity changed marginally to 90% (95% CI: 81–95) and 48% (95% CI: 26–72), respectively. Given that three of the 10 included studies were from the same author, we excluded them from our analysis. Consequently, the pooled sensitivity and specificity changed to

No	Author	Publication year	Title	Setting, Country	Outcome/ Intervention	Sample size[a]	Mode of survey administration	Source of revealed preference data	Number of choice tasks (including internal validity check)	Number of attributes	Was opt-out option given?	Analysis method used	Included/ Excluded in meta-analysis
1	Chua et al.[20]	2016	External validity of discrete choice experiment in pharmacy research: Empirical findings from a field experiment	Pharmacy, Scotland	Pharmacy-based cardiovascular health screening	SP: 423 RP: 133	Computer tablet	community pharmacy	12	6	Yes	Conditional logit model	Included
2	de Bekker-Grob et al.[21]	2020	Can healthcare choice be predicted using stated preference data?	General practice, Netherlands	Influenza vaccination	SP: 377 RP: 377	Paper-based questionnaires	General practice	16	5	Yes	HMNL + systematic preference heterogeneity + random opt-out utility (MNL, HMNL)	Included
3	de Bekker-Grob et al.[22]	2019	Are healthcare choices predictable? The impact of discrete choice experiment designs and models	General practice, Netherlands	Influenza vaccination	SP: 1261 RP: 418	Online questionnaires	General practice	16	5	Yes	HMNL + systematic preference heterogeneity + random opt-out utility (MNL, HMNL)	Included
					Colorectal cancer screening	SP:1219 RP: 406			16	5	Yes		
4	Huls and de Bekker-Grob.[23]	2022	Can healthcare choice be predicted using stated preference data? The role of model complexity in a discrete choice experiment about colorectal cancer screening	Community, Netherlands	Colorectal cancer screening	SP: 568 RP: 568	Online questionnaires	National Institute for Public Health and the Environment and local screening organizations	16	5	Yes	HMNL + systematic preference heterogeneity	Included
5	Krucien et al.[24]	2015	Empirical testing of the external validity of a discrete choice experiment to determine preferred treatment option: The case of sleep apnoea	Hospital, France	Treatment for obstructive sleep apnoea syndrome	SP: 140 RP: 138	Face-to-face interview with a trained nurse	Hospital sleep unit	10	5	Yes	G-MNL	Included
6	Kruk et al.[16]	2009	Women's preferences for place of delivery in rural Tanzania: a population-based discrete choice experiment	Community, Tanzania	Delivery	SP: 1205 RP: 1203	Face-to-face interview	Census	8	6	Yes	Hierarchical Bayesian	Excluded
7	Lambooij et al.[25]	2015	Consistency between stated and revealed preferences: a discrete choice experiment and a behavioural experiment on vaccination behaviour compared	Community, Netherlands	Hepatitis B vaccination for newborns	SP: 896 RP: 247	Paper-based questionnaires	Medical records	4	6	No	Mixed logit	Included
8	Linley and Hughes[17]	2013	Decision-makers' preferences for approving new medicines in Wales: A discrete-choice experiment with assessment of external validity	Government-funded body to appraise new medicine, Wales	New medicines	SP: 41 RP: 39	Anonymous online questionnaire	All Wales Medicines Strategy Group	7	5	Yes	Mixed logit, conditional logit model	Excluded

(Table 1 continues on next page)

No	Author	Publication year	Title	Setting, Country	Outcome/ Intervention	Sample size[a]	Mode of survey administration	Source of revealed preference data	Number of choice tasks (including internal validity check)	Number of attributes	Was opt-out option given?	Analysis method used	Included/ Excluded in meta-analysis
(Continued from previous page)													
9	Mohammadi et al.[26]	2017	Testing the external validity of a discrete choice experiment method: An application to latent tuberculosis infection treatment	Tuberculosis clinic, Canada	Latent tuberculosis infection preventive treatment (conducted several sessions)	SP: 214 RP: 204	Paper-based questionnaire	isoniazid prescription at the clinic pharmacy	12	6	Yes	Mixed logit	Included
10	Muadthong and Kessomboon[27]	2022	Preferences for participating in a new community pharmacy alcohol brief intervention in Thailand: Discrete choice experiment with assessment of external validity	Pharmacy, Thailand	Participation in a brief alcohol intervention (10–30 min conversation) led by a community pharmacist	SP: 162 RP: 162	Paper-based questionnaires	Observation by pharmacist	6	5	Yes	Mixed logit	Included
11	Ozdemir et al.[19]	2024	Getting it right with discrete choice experiments: Are we hot or cold?	Hospital, Singapore	Use of epidural analgesia for delivery	SP: 248 RP: 248	Paper-based questionnaire	Medical record	10	6	Yes	Mixed logit, latent class logit	Excluded
12	Ryan and Watson[28]	2009	Comparing welfare estimates from payment card contingent valuation and discrete choice experiments	Family Planning Clinic, United Kingdom	Chlamydia screening	SP: 130 RP: 111	Paper-based questionnaire	Family Planning Clinic	16	5	Yes	Conditional logit model	Included
13	Salampessy et al.[29]	2015	The predictive value of discrete choice experiments in public health: An exploratory application	Hospital, Netherlands	Combined lifestyle intervention (conducted over a few months) for patients with Type 2- diabetes	SP: 206 RP: 54	Paper-based questionnaires	Medical record	9	5	Yes	Mixed logit	Included
14	Salampessy et al.[18]	2022	Do patients' preferences prevail in hospital selection?: A comparison between discrete choice experiments and revealed hospital choice	Hospital, Netherlands	Choice of hospital for breast cancer surgery	SP: 631 RP: 631	Paper-based questionnaires	Claims data	10	5	No	Mixed logit	Excluded
					Choice of hospital for cataract surgery	SP: 1109 RP: 1109							

G-MNL, generalised multinominal logit; HIV, human immunodeficiency virus; HMNL; heteroskedastic multinominal logit; MNL, multinominal logit; RP, revealed preferences; SP, stated preferences. [a]Some studies may have more than one SP/RP as they included two different diseases/treatments in their papers.

Table 1: **Summary of included papers (N = 14).**

Author	Year of Publication	Outcome	Is the outcome/ intervention treatment or prevention related?	Action/ behaviour the DCE is asking about	True Positive	True Negative	False Positive	False Negative	Accuracy	Sensitivity	Specificity	PPV	NPV
Chua et al.	2016	All outcomes	Prevention-related	One-off	36	21	73	3	0.43	0.92	0.22	0.33	0.88
de Bekker-Grob et al.	2020	All outcomes	Prevention-related	One-off	227	101	32	17	0.87	0.93	0.76	0.88	0.86
de Bekker-Grob et al.	2019	Influenza vaccination	Prevention-related	One-off	246	145	10	17	0.94	0.94	0.94	0.96	0.90
de Bekker-Grob et al.	2019	Colorectal cancer screening	Prevention-related	One-off	369	25	8	4	0.97	0.99	0.76	0.98	0.86
Huls and de Bekker-Grob	2022	All outcomes	Prevention-related	One-off	510	3	45	10	0.90	0.98	0.06	0.92	0.23
Krucien et al.	2015	Continuous positive airway pressure (CPAP)	Treatment-related	Over a period of time	37	51	30	20	0.64	0.65	0.63	0.55	0.72
Krucien et al.	2015	Oral appliances (OAs)	Treatment-related	Over a period of time	36	49	12	41	0.62	0.47	0.80	0.75	0.54
Lambooij et al.	2015	All outcomes	Prevention-related	One-off	191	6	33	17	0.80	0.92	0.15	0.85	0.26
Mohammadi et al.	2017	All outcomes	Prevention-related	Over a period of time	148	22	29	5	0.83	0.97	0.43	0.84	0.81
Muadthong and Kessomboon	2022	All outcomes	Treatment-related	One-off	12	77	42	31	0.55	0.28	0.65	0.22	0.71
Ryan and Watson	2009	All outcomes	Prevention-related	One-off	88	2	2	19	0.81	0.82	0.50	0.98	0.10
Salampessy et al.	2015	All outcomes	Treatment-related	Over a period of time	36	4	9	5	0.74	0.88	0.31	0.80	0.44

CPAP, continuous positive airway pressure; NPV, negative predictive value; OA, oral appliances; PPV, positive predictive value. NB: Values for true positives, false positives, true negatives, and false negatives were calculated from stated and revealed preference data. Studies that investigated multiple treatments or diseases were extracted and analysed as distinct data sets.

Table 2: **Extracted data from included studies for meta-analysis.**

85% (95% CI: 67–94) and 39% (95% CI: 22–58), respectively. Compared to Quaife's findings, which reported pooled sensitivity and specificity estimates of 88% (95% CI: 81–92%) and 34% (95% CI: 23–46%), respectively, with an AUC of 0.60 (95% CI: 0.55–0.64), the meta-analysis of the new studies alone yielded pooled sensitivity and specificity estimates of 92% (95% CI: 75–98%) and 59% (95% CI: 27–85%), along with an AUC of 0.87 (95% CI: 0.84–0.90). These results demonstrate the significant impact of the new studies in expanding the application areas.

We assessed publication bias with Deeks' funnel plot and found no evidence of publication bias (p = 0.08; Supplementary Fig. S7). Nevertheless, the ability of this test to identify publication bias is constrained when dealing with a smaller number of studies.

Aggregate-level data

Four studies contained aggregate data that could only be narratively synthesised. In a DCE exploring women's preferences for place of delivery in rural Tanzania, Kruk et al. predicted that 56.7% of women preferred to deliver at home; however, the real-world data demonstrated that 60.8% of women delivered at home.[18] In a DCE on preferences for labour pain medicine in Singapore conducted by Ozdemir et al., both models (mixed logit and latent class logit) underestimated the probability of choosing an epidural among women admitted to a maternal institution for childbirth with the concordance rate between the predicted and actual choice ranging from 60 to 71% depending on the statistical model used.[21]

In a DCE to predict the choice of hospital in the Netherlands among patients with cataracts or breast cancer, the models overestimated which hospital patients were most (or least) likely to select for cataract or breast surgery.[20] The majority of patients with cataracts were anticipated to choose hospital 2, recognised for its high quality of care (predicted rate: 53.6%, 95% CI: 50.7–57.0), yet in reality, only 34.4% opted for it.[20] Similarly, for breast cancer patients, the model predicted that 58.5% (95% CI: 53.9–63.1) would opt for hospital 3, known for its superior quality of care, but 21.7% did so in practice. Instead, most patients preferred other hospitals labelled as "nearby located".[20] In another study, Linley and Hughes compared predicted choices obtained from forced-choice and flexible-choice models with actual choices in appraising new

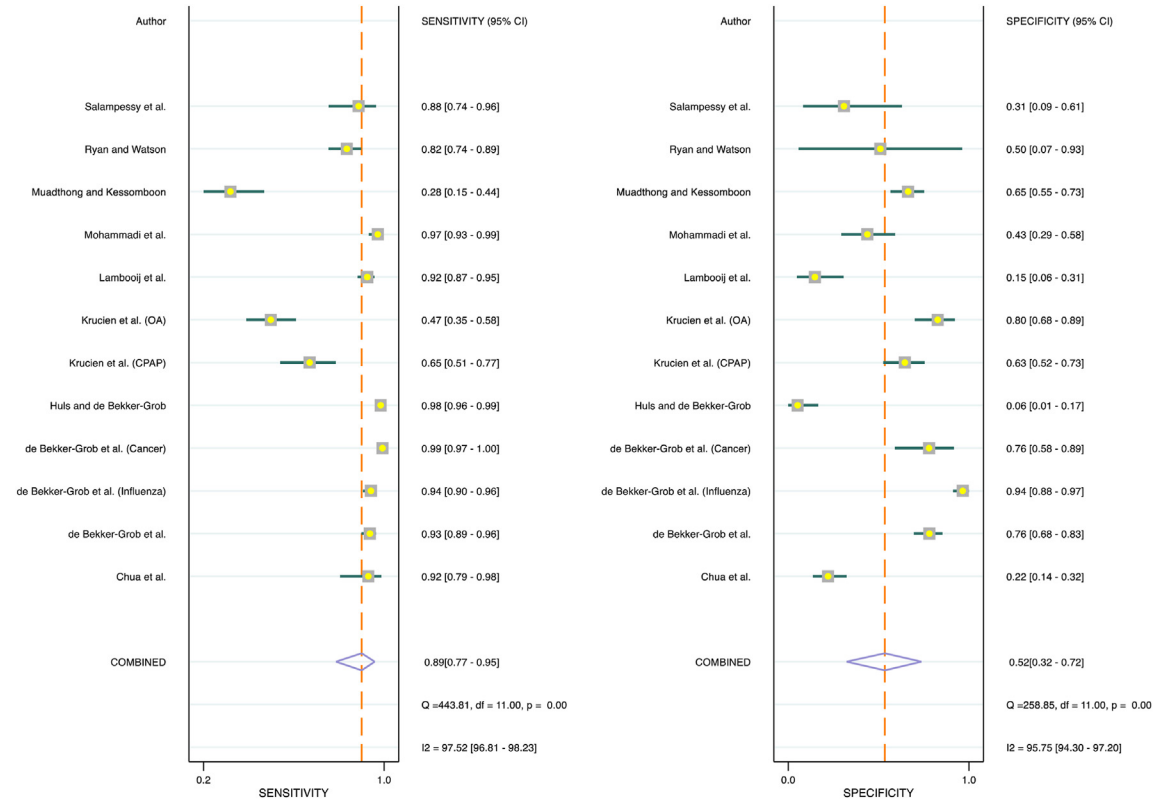

Fig. 2: Forest plot of sensitivity and specificity of DCE predictions. CPAP, continuous positive airway pressure; OA, oral appliances, NB: There were two studies that included two different diseases/treatments in their paper. Krucien et al. included oral appliances and continuous positive airway pressure as treatments for sleep apnoea. 2) de-Bekker-Grob et al. included colorectal cancer screening and influenza vaccination. Studies that investigated multiple treatments or diseases were extracted and analysed as distinct data sets.

medicines (choosing new medicine A or B based on annual number of patients to be treated), quality-adjusted life-years (QALYs) gained per treated patient, and incremental cost per QALY gained.[19] Both models correctly predicted 64% of medicine submissions, with variations in sensitivity and specificity (sensitivity of 68% and specificity of 55% for the forced-choice model and sensitivity and specificity of 64% for the flexible choice model).[19]

Discussion

This systematic review examines health-related studies comparing stated preferences, as modelled through predicted probability models from DCE choice data, to revealed preferences gathered through observational or survey means. The findings from the included studies shed light on several key aspects that warrant discussion, contributing to the understanding of the reliability and external validity of DCEs as a predictive tool in healthcare decision-making. We identified 14 studies from eight countries, employing different analytic methods and focusing on various health contexts. Consistent with past findings,[11] our study demonstrates that DCEs are valuable for capturing health-related preferences and possess good external validity to predict health-related behaviours, particularly for opt-in choices. We add new knowledge by identifying that the analysis method and contextual factors, such as the type of health intervention and the setting in which choices are made, influence the predictive accuracy of DCEs. In addition, there was no statistically significant difference in DCEs examining a single occurrence of action compared to a prolonged commitment demanded from an individual within the intervention, suggesting that DCEs predict equally well in both scenarios.

While DCEs predict health-related behaviours well, they do not perform consistently well in all health dimensions. There may be reasons why real-world choices differ from those predicted by DCEs. First, the details provided in DCE choice tasks inherently simplify reality. Even in well-executed DCEs, unobserved factors in real-life decisions may be inadequately considered. When these unobserved factors impact participants' decisions, the stated and revealed preferences may diverge. Despite efforts to create high-quality DCEs, it is unlikely

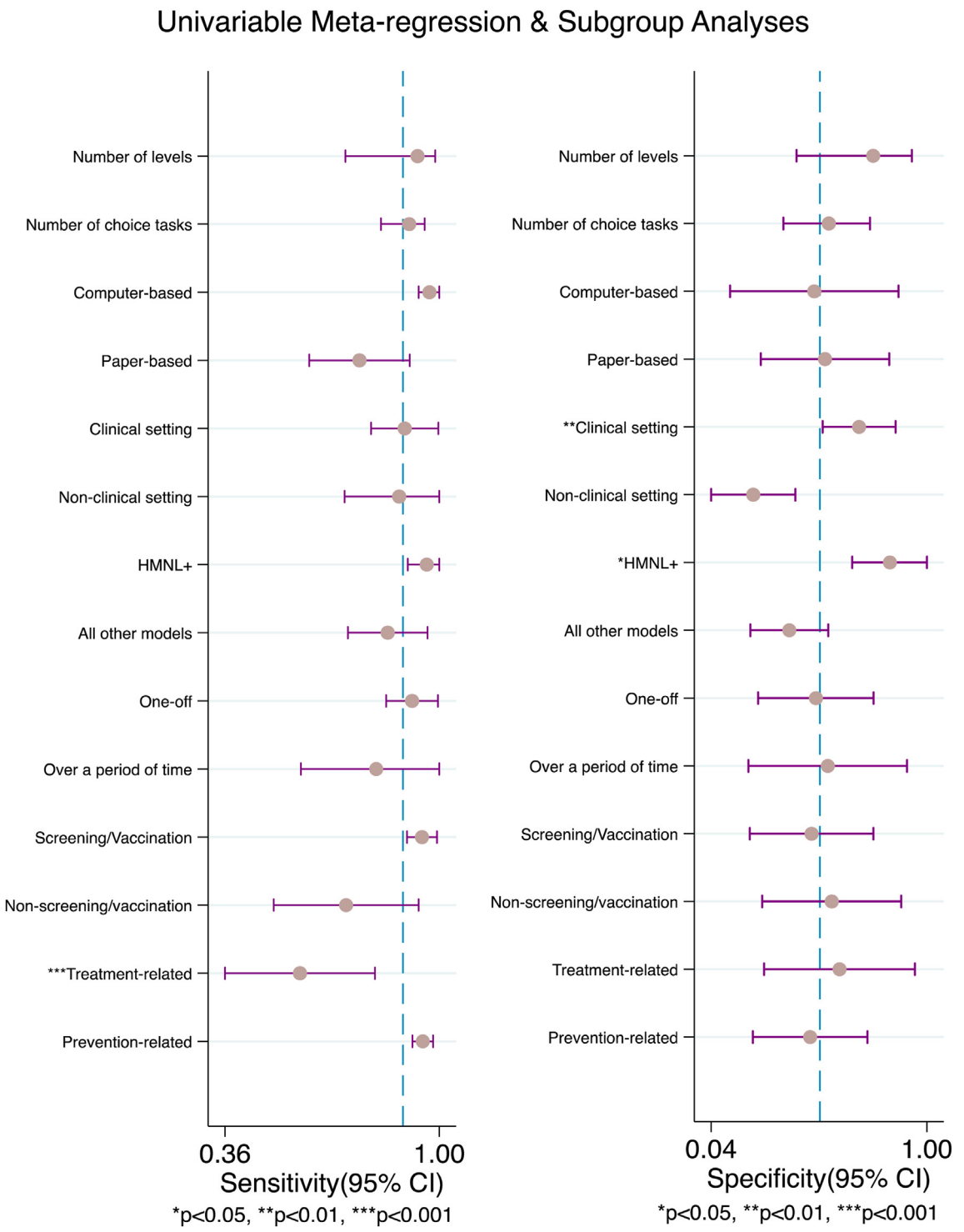

Fig. 3: Meta-regression of included studies. HMNL+, heteroskedastic multinomial logit (HMNL) with systematic preference heterogeneity and random opt-out utility.

that all pertinent choice factors will always be encompassed comprehensively. Furthermore, we observed that certain studies contained a relatively high number of attributes or choice sets for each respondent. Six (43%) of the DCE questionnaires contained more than ten choice sets, and five (36%) featured more than five attributes. However, they did not necessarily improve the predictive accuracy. This could have been a result from cognitive load or heightened task complexity,[33,34] leading respondents to simplify the task by applying heuristics, for example, ignoring certain attributes, focusing on one specific attribute or effectively making choices that they would follow on the basis of real-world scenarios. Future research should explore strategies to better account for unobserved factors and investigate methods to mitigate the impact of cognitive load in complex DCE tasks.

Second, contextual factors could have influenced the prediction accuracy of DCEs. Studies investigating specific health domains, such as treatment preferences or health service utilisation, demonstrated varying degrees of concordance between stated and revealed preferences. Our meta-regression analysis indicated that DCEs for prevention-related choices, such as vaccines or STI screening, demonstrated higher sensitivity than treatment-related choices. This difference may be attributed to a decision whether to vaccinate or undergo screening, which may be cognitively 'simpler' compared to decisions related to the complexities of managing an actual disease. Interestingly, we found that DCEs conducted in clinical settings had higher specificity than in the case of non-clinical settings. DCEs undertaken in clinical settings may increase the salience of a healthcare decision or allow participants to experience the attributes or attribute levels first-hand. This may reduce hypothetical bias – for example, DCEs conducted alongside clinical trials. Some studies considered in our review incorporated hypothetical bias mitigation strategies. In a DCE study focused on chlamydia screening, participants, after finishing the questionnaire, attended a scheduled appointment with a doctor where they were presented with the option of undergoing a chlamydia screening test.[30] The DCE in this case may have partially acted as a decision aid to weigh the different trade-offs and be prepared for the chlamydia screening discussion with the doctor, i.e., a sort of priming. Further research is needed to understand the impact of context-specific nuances on the predictive power of DCEs, which is crucial for tailoring their application to different healthcare settings.

Third, DCE predictions may suffer due to assumption of rational choice inherent in DCEs, whereas the body of work from behavioural economics suggests people may often act 'irrationally' due to cognitive biases such as anchoring bias, availability heuristic, framing effects or optimism bias.[35] For example, in the HIV prevention studies, this rational approach is being applied in a heightened emotional context of a sex act, and actual behaviour might not be a balanced evaluation of all evidence. As such, the individual's choice is no longer irrational, which violates the assumption of DCE, and thus affecting DCE's predictive accuracy. Another example is the intention–behaviour gap, resulting in individuals not always acting according to their stated intentions.[36,37] For instance, when individuals are far from making a decision, they might commit to a significant course of action, such as quitting smoking. However, as the decision moment approaches, they may opt for a small reward, such as having a cigarette. This phenomenon of hyperbolic discounting indicates that the passage of time alters the perception of the situation and choices, potentially accounting for discrepancies between DCE predictions and actual behaviour. Furthermore, the external validity of DCEs might be influenced by the hot-cold empathy gap, where individuals underestimate the impact of emotional states on their decision-making.[38] For example, decisions made in a 'cold' state when completing the survey (e.g., choice to use a condom) might be very different from a 'hot' state when the individual is confronted with the actual decision within context (e.g., before having sex). Emotional states can also distort how individuals perceive risks associated with medical interventions. Hence, this poses a challenge to the predictive accuracy of DCE findings in real-world scenarios where emotional states are subjected to fluctuations. More research is needed to expand our understanding of how these effects inherent in individuals impact the ability for DCEs to predict actual behaviours, and thus, how they can be adjusted for when collecting data within surveys.

The prediction accuracy of DCEs could be significantly impacted by people other than the decision-maker. One study conducted by one of our co-authors was stopped in the pre-testing phase and, hence, was not published.[39] This was because, in the pre-testing qualitative research (think-aloud) phase, the researchers found they could not ignore the influence that physicians and partners had on the prostate cancer treatment choices of patients. de Bekker-Grob et al. conducted an empirical examination of a choice model that integrates social influences into a DCE.[40] They observed that the influence of key influencers (e.g., partners, parents, friends, healthcare professionals) on maternal decisions regarding childhood vaccination greatly outweighed the influence of attributes associated with the vaccination itself.[40] Their DCE study found that social factors predominantly influenced maternal decisions concerning childhood vaccination.[40] Future DCEs could consider the potential impact of social influences on health-related decision-making, especially in contexts where strong underlying social norms shape the decision-makers' behaviour.

Our meta-regression revealed that the HMNL model, incorporating systematic preference heterogeneity and

random opt-out utility, demonstrated higher specificity than alternative models (i.e., multinomial logit, generalised multinomial logit). However the studies that used this analytic model were all from the same author, and there may be other confounders that could have improved external validity. Among the included papers, 29% (4/14) explored more than one model for analysis and the best-performing model was compared to the revealed preference data, while the remaining papers used a single model. Selecting an appropriate model in DCEs is a process that requires careful consideration of theoretical assumptions, methodological rigour, and practical implications. The analysis method is crucial in maximizing the predictive accuracy of DCEs as it affects how well the model captures the complexity of individual decision-making processes. Different models incorporate various assumptions about the data, such as the independence of choices or the presence of heterogeneity among respondents. Advanced models like the HMNL model can account for variations in decision-making behaviour and adapt to different scales of preferences, leading to more accurate predictions. However, as all studies that applied HMNL were led by the same study team, there may be other factors driving the strong predictive validity in addition to a well-fitting statistical model. Given the array of available models, researchers should adopt a systematic approach to model selection, leveraging a combination of goodness-of-fit metrics and predictive performance assessments in an iterative manner to identify the most appropriate model for their analyses.[41] Future guidelines should emphasise the importance of reporting the model selection process and justifying the analytic approach to ensure methodological robustness and external validity.

Our paper introduces several new insights that advance the understanding of prediction accuracy in health-related DCEs. By expanding the review to cover studies published between 2015 and 2023, it updates the field with a larger, more recent dataset than previous reviews. Notably, the study highlights the improved accuracy of DCEs, with a higher pooled sensitivity (89%) and specificity (52%) compared to earlier findings, suggesting advancements in prediction accuracy. We also uncover factors that influence these results, such as the higher sensitivity in prevention-related DCEs (e.g., vaccinations) and higher specificity in those conducted in clinical settings (e.g., hospitals). These findings not only enhance the robustness of DCE predictions across new areas but also offer clear signals for future studies: focusing on specific health interventions (e.g., preventive measures) and selection of study site to collect DCE data. This work signals pathways for refining DCE methodology to improve external validity, thus shaping future predictive modelling approaches in health care decision-making.

Our findings should be read in the light of several limitations. First, our sample size was relatively small due to the limited number of health-related DCE studies and the meta-analysis might lack statistical power. This limited our ability to explore further factors impacting the predictive accuracy of DCEs. Future DCE studies should consider incorporating an external validity component in their design where applicable. Second, some of the included studies did not clearly specify the layout of DCE questionnaires. Improved transparency in reporting and addressing these limitations is essential for strengthening the overall quality of DCE research. More consistent use of standardised reporting guidelines specific to DCEs in health-related research may facilitate comprehensive reporting and ensure methodological rigour. Third, we observed significant between-study heterogeneity, which could be attributed to the differences in study design, choice set composition, and analytical methods, reflecting the wide range of applications of DCEs in the health sector. Future research can explore variations in predictive power across different demographics, such as sex and age, to offer deeper insights and improve the generalizability of DCE results. Fourth, there could have been publication bias as we did not include studies conducted (but not peer-reviewed and published) by private sectors as industry companies may prefer to keep proprietary methodologies and insights confidential. Although publishing these calibrated DCE results will allow private industries to position themselves as leaders in innovation, build consumer trust, and demonstrate transparency in data-driven decision-making. Sharing these results can also foster academic and industry collaboration, enhance investor confidence, and meet regulatory or compliance requirements, especially in the healthcare sector. Fifth, our search strategy did not include specific terms for part-worth utility, paired comparisons, or choice-based studies, which might have provided a more comprehensive search. As a result, we may have missed some relevant papers. However, this is unlikely, as an additional search conducted in Medline did not identify any further studies reporting relevant individual-level data. Additionally, we consulted with other DCE practitioners, including our coauthors, to ensure that no other potential studies were overlooked. Sixth, no studies examined the external validity of the trade-offs between attributes, suggesting an area for future research. Finally, we acknowledge there are other methods that can be used as proxies for some predictive validity, such as repeated tasks. Future research can explore if the repeated tasks is correlated to real world predictive validity, or if inconsistencies is more reflective of learning effects or rather not paying attention.

When conducting and reporting studies comparing DCEs with market data, several best practices should be adhered to in order to ensure rigour, transparency, and validity. One of the primary considerations is to clearly define the alignment between the attributes and levels in the DCE and those observed in the real-world market

data to ensure comparability. Researchers should also employ scaling techniques to adjust for differences in preference intensities between stated (DCE) and revealed (market) data, addressing potential scale discrepancies. The use of hybrid models, which integrate both DCE and market data, is recommended as these models can provide more robust estimates by capturing the strengths of both approaches. Calibration is crucial for aligning stated preference and revealed preference data to better reflect real-world behaviours. Common strategies include rescaling stated preference data to match the magnitude of revealed preference data, incorporating correction factors or applying anchoring techniques based on known market data.[42] Of the 14 studies reviewed, five applied scaling techniques in their calibration. A critical element in this calibration process is the estimation of part-worths (or part-utilities). Part-worths reflect the relative contribution of each attribute level to an individual's overall utility, helping to quantify how changes in specific health-related attributes (e.g., treatment effectiveness, side effects) influence decision-making.[43] Accurate estimation of part-worths is essential for predicting real-world choices because they reflect the trade-offs individuals make. By aligning part-worth utilities between DCEs and real-world data, researchers can ensure that preference intensities are properly calibrated, enhancing the overall predictive power of the model.

To ensure the accuracy and generalizability of models, cross-validation techniques—such as splitting the data into training and testing sets—can be used. Sensitivity analysis and holdout sample testing also help evaluate how well the model can generalize across different datasets, thereby enhancing its predictive validity.[44] In reporting findings, it is critical to provide a detailed account of the model selection process, calibration methods, and any goodness-of-fit metrics, such as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), used for model validation. Including sensitivity analyses that explore how changes in model parameters or attributes affect outcomes is important. Transparency in discussing limitations, such as potential biases or data constraints, is also key to ensuring the external validity and interpretability of the study.

Ensuring that choice tasks closely mirror real-world scenarios is another critical factor for enhancing the face validity. Formative qualitative work, such as pilot testing, can play a valuable role in this process by ensuring that the attributes and levels are tailored to the characteristics of relevant subpopulations and effectively measure the intended factors as perceived by the researchers.[45] If attribute selection is not grounded in qualitative research, inappropriate attributes may be specified, and key attributes may be omitted, leading to omitted variable bias. Qualitative methods, such as focus groups and interviews are also useful for better understanding participants' responses to DCEs and can help avoid such biases.[5] These methods improve the overall predictive power of DCEs by ensuring the relevance and accuracy of the attributes selected. Warm-up tasks are also crucial in DCE as they help participants familiarize themselves with the task format and attribute levels. By initially using single-attribute preferences, participants can better understand each attribute individually. This approach ensures that participants are comfortable with the task and its requirements before engaging in the actual experiment. Implementing a simple warm-up choice task would not only improve participants' comprehension but also enhances the predictive validity of the experiment by reducing the likelihood of confusion or misinterpretation during the main task.

This systematic review and meta-analysis present a comprehensive overview of the prediction accuracy of DCEs in health-related research across different types of intervention. DCEs are valuable for capturing health-related preferences and possess reasonable external validity to predict health-related behaviours in the real world, particularly for opt-in choices. Contextual factors, such as the type of intervention, study setting, and analysis method, influence the predictive accuracy. By identifying key factors that improve both sensitivity and specificity, this paper contributes to refining the use of DCEs in health research, offering guidance for future studies aiming to improve the reliability of DCEs in capturing patient preferences across a range of healthcare decisions. While challenges and variations exist, the evolving landscape of methodological advancements and its potential impact on healthcare decision-making highlight the importance of continued research in this area. As predictive accuracy is improved and methodological challenges are addressed, DCEs can be used to inform improved resource allocation, prioritise interventions, and enhance patient-centred care.

Contributors

JO conceived the idea for this paper. YZ and AH did the screening and data extraction. YZ conducted the statistical analysis. YZ, AH and JO accessed and verified the data. All authors had full access to all the data in the study, contributed to the interpretation and subsequent edits of the manuscript, and had final responsibility for the decision to submit for publication.

Data sharing statement

The data will be shared on reasonable request to the corresponding author.

Declaration of interests

MQ is employed by Evidera, a consultancy who undertake preference research on behalf of pharmaceutical clients. All other authors declare no competing interests.

Acknowledgements

JJO is supported by an Australian National Health and Medical Research Council Emerging Leadership Investigator Grant (GNT1193955). EBG is supported by the Dutch Research Council (NWO-Talent-Scheme-Vidi-Grant No, 09150171910002). YZ is supported by an Australian Government Research Training Program (RTP) scholarship.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.eclinm.2024.102965.

References

1 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.
2 Clark MD, Determann D, Petrou S, Moro D, de Bekker-Grob EW. Discrete choice experiments in health economics: a review of the literature. *Pharmacoeconomics*. 2014;32(9):883–902.
3 Mühlbacher AC, Bridges JF, Bethge S, et al. Preferences for antiviral therapy of chronic hepatitis C: a discrete choice experiment. *Eur J Health Econ*. 2017;18(2):155–165.
4 Lancsar E, Swait J. Reconceptualising the external validity of discrete choice experiments. *Pharmacoeconomics*. 2014;32(10):951–965.
5 de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. *Health Econ*. 2012;21(2):145–172.
6 Mandeville KL, Lagarde M, Hanson K. The use of discrete choice experiments to inform health workforce policy: a systematic review. *BMC Health Serv Res*. 2014;14:367.
7 Lewis RA, Neal RD, Hendry M, et al. Patients' and healthcare professionals' views of cancer follow-up: systematic review. *Br J Gen Pract*. 2009;59(564):e248–e259.
8 Purnell TS, Joy S, Little E, Bridges JF, Maruthur N. Patient preferences for noninsulin diabetes medications: a systematic review. *Diabetes Care*. 2014;37(7):2055–2062.
9 Reed Johnson F, Lancsar E, Marshall D, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task force. *Value Health*. 2013;16(1):3–13.
10 Harrison M, Rigby D, Vass C, Flynn T, Louviere J, Payne K. Risk as an attribute in discrete choice experiments: a systematic review of the literature. *Patient*. 2014;7(2):151–170.
11 Quaife 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.
12 Higgins J, Thomas J, Chandler J, et al. *Cochrane Handbook for systematic reviews of interventions version 6.3*; 2022. February 2022. www.training.cochrane.org/handbook
13 Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71.
14 Harbord RM, Whiting P. metandi: meta-analysis of diagnostic accuracy using hierarchical logistic regression. *STATA J*. 2009;9(2):211–229.
15 Dwamena B. *MIDAS: Stata module for meta-analytical integration of diagnostic test accuracy studies*. Statistical Software Components; 2007.
16 Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol*. 2005;58(9):882–893.
17 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.
18 Kruk ME, Paczkowski M, Mbaruku G, de Pinho H, Galea S. Women's preferences for place of delivery in rural Tanzania: a population-based discrete choice experiment. *Am J Public Health*. 2009;99(9):1666–1672.
19 Linley WG, Hughes DA. Decision-makers' preferences for approving new medicines in Wales: a discrete-choice experiment with assessment of external validity. *Pharmacoeconomics*. 2013;31(4):345–355.
20 Salampessy BH, Ikkersheim D, Portrait FRM, Koolman X. Do patients' preferences prevail in hospital selection?: a comparison between discrete choice experiments and revealed hospital choice. *BMC Health Serv Res*. 2022;22(1):1136.
21 Ozdemir S, Gonzalez JM, Bansal P, Huynh VA, Sng BL, Finkelstein E. Getting it right with discrete choice experiments: are we hot or cold? *Soc Sci Med*. 2024;348:116850.
22 Chua GN, Bond CM, Porteous T, Ryan M. External validity of discrete choice experiment in pharmacy research: empirical findings from a field experiment. *Int J Pharm Pract*. 2016;24(SUPPL. 1):6–7.
23 de Bekker-Grob EW, Donkers B, Bliemer MCJ, Veldwijk J, Swait JD. Can healthcare choice be predicted using stated preference data? *Soc Sci Med*. 2020;246:112736.
24 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.
25 Huls SPI, de Bekker-Grob EW. Can healthcare choice be predicted using stated preference data? The role of model complexity in a discrete choice experiment about colorectal cancer screening. *Soc Sci Med*. 2022;315:115530.
26 Krucien N, Gafni A, Pelletier-Fleury N. Empirical testing of the external validity of a discrete choice experiment to determine preferred treatment option: the case of sleep apnoea. *Health Econ*. 2015;24(8):951–965.
27 Lambooij MS, Harmsen IA, Veldwijk J, et al. Consistency between stated and revealed preferences: a discrete choice experiment and a behavioural experiment on vaccination behaviour compared. *BMC Med Res Methodol*. 2015;15(100968545):19.
28 Mohammadi T, Bansback N, Marra F, et al. Testing the external validity of a discrete choice experiment method: an application to latent tuberculosis infection treatment. *Value Health*. 2017;20(7):969–975.
29 Muadthong S, Kessomboon N. Preferences for participating in a new community pharmacy alcohol brief intervention in Thailand: discrete choice experiment with assessment of external validity. *Trop J Pharm Res*. 2022;21(1):159–167.
30 Ryan M, Watson V. Comparing welfare estimates from payment card contingent valuation and discrete choice experiments. *Health Econ*. 2009;18(4):389–401.
31 Salampessy BH, Veldwijk J, Jantine Schuit A, et al. The predictive value of discrete choice experiments in public health: an exploratory application. *Patient*. 2015;8(6):521–529.
32 Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol*. 2010;5(9):1315–1316.
33 Bech M, Kjaer T, Lauridsen J. Does the number of choice sets matter? Results from a web survey applying a discrete choice experiment. *Health Econ*. 2011;20(3):273–286.
34 Louviere JJ, Islam T, Wasi N, Street D, Burgess L. Designing discrete choice experiments: do optimal designs come at a price? *J Consm Res*. 2008;35(2):360–375.
35 Blumenthal-Barby JS, Krieger H. Cognitive biases and heuristics in medical decision making: a critical review using a systematic search strategy. *Med Decis Making*. 2015;35(4):539–557.
36 Sheeran P. Intention—behavior relations: a conceptual and empirical review. *Eur Rev Soc*. 2002;12(1):1–36.
37 Ajzen I. The theory of planned behavior. *Organ Behav Hum Decis Process*. 1991;50(2):179–211.
38 Loewenstein G. Hot-cold empathy gaps and medical decision making. *Health Psychol*. 2005;24(4s):S49–S56.
39 de Bekker-Grob EW, Bliemer MCJ, Donkers B, et al. Patients' and urologists' preferences for prostate cancer treatment: a discrete choice experiment. *Br J Cancer*. 2013;109(3):633–640.
40 de Bekker-Grob EW, Howard K, Swait J. Identifying the impact of social influences in health-related discrete choice experiments. *PLoS One*. 2022;17(10):e0276141.
41 Hauber AB, González JM, Groothuis-Oudshoorn CG, et al. Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR conjoint analysis good research practices task force. *Value Health*. 2016;19(4):300–315.
42 Huang Y, Li W, Macheret F, Gabriel RA, Ohno-Machado L. A tutorial on calibration measurements and calibration models for clinical prediction models. *J Am Med Inf Assoc*. 2020;27(4):621–633.
43 Train KE. *Discrete choice methods with simulation*. Cambridge University Press; 2009.
44 Mariel P, Hoyos D, Meyerhoff J, et al. Validity and reliability. In: Mariel P, Hoyos D, Meyerhoff J, et al., eds. *Environmental valuation with discrete choice experiments: guidance on design, implementation and data analysis*. Cham: Springer International Publishing; 2021:111–123.
45 Coast J, Horrocks S. Developing attributes and levels for discrete choice experiments using qualitative methods. *J Health Serv Res Policy*. 2007;12(1):25–30.