SYSTEMATIC REVIEW

Current Practices for Accounting for Preference Heterogeneity in Health-Related Discrete Choice Experiments: A Systematic Review

Suzana Karim[1] · Benjamin M. Craig[1] · Caroline Vass[2,3] · Catharina G. M. Groothuis-Oudshoorn[4]

Abstract

Background Accounting for preference heterogeneity is a growing analytical practice in health-related discrete choice experiments (DCEs). As heterogeneity may be examined from different stakeholder perspectives with different methods, identifying the breadth of these methodological approaches and understanding the differences are major steps to provide guidance on good research practices.

Objectives Our objective was to systematically summarize current practices that account for preference heterogeneity based on the published DCEs related to healthcare.

Methods This systematic review is part of the project led by the Professional Society for Health Economics and Outcomes Research (ISPOR) health preference research special interest group. The systematic review conducted systematic searches on the PubMed, OVID, and Web of Science databases, as well as on two recently published reviews, to identify articles. The review included health-related DCE articles published between 1 January 2000 and 30 March 2020. All the included articles also presented evidence on preference heterogeneity analysis based on either explained or unexplained factors or both.

Results Overall, 342 of the 2202 (16%) articles met the inclusion/exclusion criteria for extraction. The trend showed that analyses of preference heterogeneity increased substantially after 2010 and that such analyses mainly examined heterogeneity due to observable or unobservable factors in individual characteristics. Heterogeneity through observable differences (i.e., explained heterogeneity) is identified among 131 (40%) of the 342 articles and included one or more interactions between an attribute variable and an observable characteristic of the respondent. To capture unobserved heterogeneity (i.e., unexplained heterogeneity), the studies largely estimated either a mixed logit ($n = 205$, 60%) or a latent-class logit ($n = 112$, 32.7%) model. Few studies ($n = 38$, 11%) explored scale heterogeneity or heteroskedasticity.

Conclusions Providing preference heterogeneity evidence in health-related DCEs has been found as an increasingly used practice among researchers. In recent studies, controlling for unexplained preference heterogeneity has been seen as a common practice rather than explained ones (e.g., interactions), yet a lack of providing methodological details has been observed in many studies that might impact the quality of analysis. As heterogeneity can be assessed from different stakeholder perspectives with different methods, researchers should become more technically pronounced to increase confidence in the results and improve the ability of decision makers to act on the preference evidence.

1 Introduction

Systematic literature reviews of discrete choice experiments (DCEs) in healthcare have shown increases in both the number of applications and the complexity of models; however, these reviews have disregarded the detailed information on the modeling of preference heterogeneity, specifically [1–3]. Understanding heterogeneity in preferences for health and healthcare-related goods and services motivates researchers to apply alternative methods to estimate preferences across individuals. Most of the preference heterogeneity analyses in the stated choice experiments have a tendency to explore

✉ Suzana Karim
 suzanakarim@usf.edu

[1] University of South Florida, 4202 E Fowler Ave, Tampa, FL 33620, USA

[2] RTI Health Solutions, Manchester, UK

[3] The University of Manchester, Manchester, UK

[4] University of Twente, Enschede, The Netherlands

Key Points for Decision Makers

The current practices in healthcare discrete choice experiments (DCEs) incorporate multiple methodologies to account for preference heterogeneity.

Understanding unexplained preference heterogeneity through the use of mixed models (mixed logit/random parameter logit or latent class) is more prevalent in the health-related DCE literature than investigations into explained differences through the use of interaction terms.

Few researchers have attempted to show alternative specifications (i.e., distributional assumptions) to understand the evidence of unexplained preference heterogeneity or to apply more complex models (such as using correlated distribution for preference parameters or controlling scale to account for preference heterogeneity) to enrich the interpretation of the results.

heterogeneity from different angles, which are guided by the choice modeling research questions, fundamental assumptions of choice behavior (i.e., whether choice heterogeneity varies systematically or randomly, whether heterogeneity varies through observable characteristics or latent characteristics), and model specifications (e.g., model identification, distributional assumption, model fitting and selection criteria, convergence aspects) [4–6]. Furthermore, the confounding nature of the preference parameter with the scale parameter motivates many researchers to account for scale heterogeneity (i.e., heteroskedasticity) while estimating preference/taste heterogeneity (i.e., differences in the relative effects of attribute levels) [7]. These make the existing research fragmented, and no systematic reviews have yet been conducted to summarize the existing preference heterogeneity approaches applied in healthcare studies. As published in *Pharmacoeconomics* [1], Zhou and colleagues conducted a systematic review specifically on the latent class (LC) models applied in healthcare preference studies, and many other systematic reviews on healthcare DCEs have long existed that mostly provided descriptions of the general state of the science [2–4, 8]. Hence, the objectives of this manuscript are to systematically summarize the current practices that account for preference heterogeneity, focusing on the published DCEs related to healthcare and identifying methodological gaps to inform future researchers in the DCE health preference heterogeneity research field.

This systematic review was conducted by the authors in collaboration with a strategic working group within The Professional Society for Health Economics and Outcomes

Research (ISPOR) Health Preference Research Special Interest Group (HPR SIG), which was established in early 2020. The acknowledgment section of the paper duly credited the contributors of the HPR SIG members. In order to provide a comprehensive overview of the established practices, the working group also performed an online survey among health preference researchers and methodological experts in choice modeling. Excerpts from this literature review were published alongside the survey evidence to provide context for the scientific community [9]. As the thorough discussion on the publication trend was beyond the scope of the HPR SIG's report, the aim of the current paper is to document the findings of the systematic literature review in detail, emphasizing noteworthy patterns and trends in the examination of preference heterogeneity within health-related DCEs. For example, this systematic review includes in-depth descriptions of the research questions, the samples recruited, and an evaluation of the various approaches to experimental design. The paper also included a critical review of the analytical models applied in analyzing preference heterogeneity. Finally, the results are synthesized into a narrative summary that shows methodological gaps in current research practices for future healthcare DCEs to investigate preference heterogeneity.

2 Methods

2.1 Search Strategy

The literature review employed the search strategy in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist [10] and earlier published systematic reviews of healthcare DCEs [1, 2, 8, 11]. The literature search was performed in the PubMed, OVID (EconLit, Medline, Embase), and Web of Science electronic databases using the joint search terms shown in electronic supplementary material (ESM) Appendix 1, Fig. A1). In addition, supplementary citation searches were performed on two of the latest published systematic reviews on health DCEs [1, 2]. The date of the search period for published articles was from 1 January 2000 to 31 March 2020, as preference heterogeneity is little discussed in articles before 2000 [4]. During the search process, records were removed if (1) author name or abstract was missing; (2) published in non-English journals; (3) published as proceedings or conference paper, abstract only, or as a book section; and (4) systematic reviews and published protocols. Finally, duplicates from the identified articles were removed by hand and with Endnote X9 tools [12]. The final list was shared within ISPOR's health preference research SIG [13] to identify any of the missing articles on preference heterogeneity that were not retrieved earlier.

2.2 Inclusion and Exclusion Criteria

Articles were included if (1) a DCE was used to elicit preferences for health and healthcare, such as health valuation studies, treatment studies, structure/policy studies (e.g., examine job preferences of health workers) [1, 2]; (2) the analysis was based on a random utility model (RUM) under microeconomic theory [14]; and (3) included analyses of preference heterogeneity either observable difference (e.g., included interaction terms between attributes/attribute levels and respondents' characteristics in a pooled data set) [15] or unobservable differences (e.g., finite and continuous mixture models with and without covariates effects). For the purposes of this review, informative heterogeneity (i.e., the systematic correlation between unobservable factors associated with choices, e.g., nested logit) is included within unobservable differences; however, we recognize that such correlations may be unrelated to individual factors.

Articles were excluded if (1) studies analyzed preference of food (e.g., high sugar), transportation (e.g., road safety), and environment (e.g., air quality control) that may or may not be related to health unless addressing health and healthcare audiences (published in health, health economics, or methodological journal); (2) studies focused on choice heterogeneity only to evaluate heuristics (e.g., attribute non-attendance), information processing (i.e., differences in utility function), and data mining perspectives; and (3) studies that stratified the data with separate analyses (e.g., multiple countries or studies compared patients and physicians unless a random parameter logit [RPL] ran in subsample). To concentrate the discussion of analytical methods only on DCEs, studies with adaptive or scale-based preference-elicitation tasks, such as best-worst scaling (case 1, case 2, or case 3) [16], conjoint analysis (where respondents were required to rate or rank alternatives or attribute levels) [17], kaizen [18], and ranking [19] were also excluded.

2.3 Study Selection

At the first stage of article selection, all abstracts were examined by a pair of trained reviewers. Sixteen members from the ISPOR preference heterogeneity working group voluntarily participated in the training and screening process and were paired into eight groups. The pooled articles from the initial search were distributed among the eight groups with no overlapping. The detail of the working group is mentioned in the contributor section. Based on the inclusion/exclusion criteria, each reviewer independently classified the abstracts into accepted, rejected, or remained undecided. If both reviewers accepted an abstract, the article advanced to the next step for full-text review; however if reviewers differentiated in opinion, a third reviewer (CV) independently arbitrated the disputed abstract.

After the abstract review, the articles were collected, curated, and distributed again. As with the abstract review, the steps of the dual review were repeated in the full-text screening stage. If disagreement remained after third-party arbitration, the study team (BC, CV, SH, SK) discussed the articles and achieved a consensus-based decision. Once the full list of included articles was finalized, the study team sought feedback from the ISPOR HPR Special Interest Group [13] and the heterogeneity working group to include any missed articles.

2.4 Data Extraction

Given the articles that passed the full-test screening stage, the reviewers extracted data using a Microsoft Excel tool developed in Visual Basic for Applications (VBA). This VBA tool (ESM Appendix 2) was developed through an iterative consultation process with working group members, and each reviewer participated in the beta testing to clarify its terminology and to ensure systematic extraction of the key study features of interest. The VBA tool promoted consistency in extraction across the working group members who had different backgrounds and varying levels of expertise.

Using the VBA tool, reviewers extracted background information such as date of publication, objective (e.g., preferences for treatments, health outcomes, jobs), country of origin, sample type and size, number of alternatives and attributes, and experimental design. Detailed consideration was given to analytical models such as articles that conducted an RPL, data were extracted on number and type or draws, and the type of distribution specified for the random parameters. In studies reporting LC analysis, details about the number of classes and approaches used to select classes (e.g., Bayesian, Akaike information criteria [AIC]) were extracted. Other relevant extracted information was the statistical software name, estimation approach, statistical tests for the presence of heterogeneity, post-estimation calculations (e.g., willingness to pay, maximum acceptable risk, relative importance [RI]), and subjective categorization of the extent of discussion on preference heterogeneity into the article. Inferences were discouraged (for example, if an article did not state certain details, it was not inferred from the software used but rather marked as 'unreported'). Apart from differences between reviewers, data extraction may vary between articles due to differences in research practices and reporting quality.

2.5 Analysis

To understand the change in the characteristics of publications over time, we analyzed the articles in three different time frames: 2002–2010, 2011–2015, and 2016–2020

(March)[1]. The background information of articles was summarized through counts, percentages, means, and standard deviations (SDs). The major background information was the year of publication, country, study objective, study field and population. Data were also extracted on the study design, including sample size, number of attributes, number of alternatives, and experimental design approach. Finally, information on the presentation of results was also recorded. The review summarized the methodology of each study to understand how preference heterogeneity was addressed. Key outcomes related to methods of each study comprising modeling approaches such as observable and unobservable heterogeneity, consideration of scale, statistical specifications such as details of the distribution of estimated parameters and estimation techniques, software use, application of alternative models, and the interpretation of heterogeneous preferences.

3 Overview of Healthcare Discrete Choice Experiments (DCEs) with Preference Heterogeneity Analyses

The initial search from the electronic databases and citation of relevant articles identified 1202 records, of which 975 were accepted after the first screening. Chronological screening steps of abstracts and then the full texts identified 379 articles for data extraction. During the data extraction phase, another 45 studies were red-flagged and expert guidance incorporated another eight articles. Thus, the final database contained 342 articles (Fig. 1); the full list of articles can be found in ESM Appendix 1.

3.1 Trends and Sample Size

The trend of published healthcare DCEs shows that preference heterogeneity analyses increased substantially after 2010 (i.e., yearly publication increased by more than 10 since 2011) (Fig. 2). The average number of published preference heterogeneity analyses was 3.6 per year until 2010, increasing to 29.2 between 2011 and 2015, and 33.4 between 2016 and 2020. In 2015, the number was highest ($n = 65$). Among the 342 articles, the sample sizes ranged from 35 to 4600, with a mean of 583.9 (SD 567.83) and a median of 406 (interquartile range [IQR] 213.8–722.8) (Fig. 2). Only around 16% ($n = 59$) of the analyses had over 1000 respondents.

3.2 Areas of Application, Regions, and Population

Among the 342 articles, Table 1 shows that the preference heterogeneity analyses were largely focused on the preferences for treatment, or medicines ($n = 163$, 47.8%), followed by preferences for devices, screening, or procedures ($n = 64$, 18.8%) and health policies ($n = 46$, 13.5%). Around 7% ($n = 24$) of the studies were focused on job or education choices, 3% on health states/outcomes (within the quality-adjusted life-year framework), 1% on preferences for trial or program participation (e.g., blood donation, smoking cessation program), and around 9% had other types of objectives (e.g., preference for insurance, lifestyle, services, learning, and information).

The majority of the preference heterogeneity analyses were performed in Europe ($n = 193$, 56.4%), followed by North America ($n = 94$, 27.5%) and Oceania ($n = 38$, 11.1%), namely Australia (Table 1). Within Europe, 25% of the articles were from the UK ($n = 48$), 16% were from Germany ($n = 31$), and others were from other countries, such as France ($n = 15$), Spain ($n = 8$) and Italy ($n = 12$). In North America, all studies are from the US ($n = 55$, 16.0%) and Canada ($n = 38$, 11.1%) except one from Central America. In Asia ($n = 24$, 7.0%), almost half of the studies are from China ($n = 11$) and 8% of the total extracted papers are from Africa ($n = 27$). South America produced the least number of studies across the world ($n = 3$) [20–23].

Within these areas of application and regions, most studies were conducted among patients ($n = 163$, 47.5%), followed by the general population ($n = 94$, 27.4%) and healthcare workers ($n = 52$, 15.2%) [e.g., physicians, nurses, pharmacists].

3.3 Attributes, Alternatives, Designs, and Estimates

Around half of the reviewed papers ($n = 173$, 50.58%) had five or six attributes, with the minimum and maximum number of attributes being 2 and 20, respectively (Fig. 3, Table 2). Furthermore, some studies described the choice tasks by presenting a combination of subsets of attributes rather than using the standard practice of presenting all attributes in each choice task [24–32]. The majority of studies had two alternatives ($n = 271$, 80.2%), followed by three alternatives ($n = 42$, 12.4%), excluding the reference case (e.g., opt-out, status quo). In most of the papers, the alternatives ($n = 279$, 81.4%) were unlabeled and around 35% of the studies included a reference case in the choice scenarios (Table 2).

In order to retrieve information on the experimental design, we collected data on specific terms that the authors mentioned explaining the study design. In their description, nearly all articles reported the number of choice tasks per respondent ($n = 334$, 97.6%), which ranged from 3 to 48,
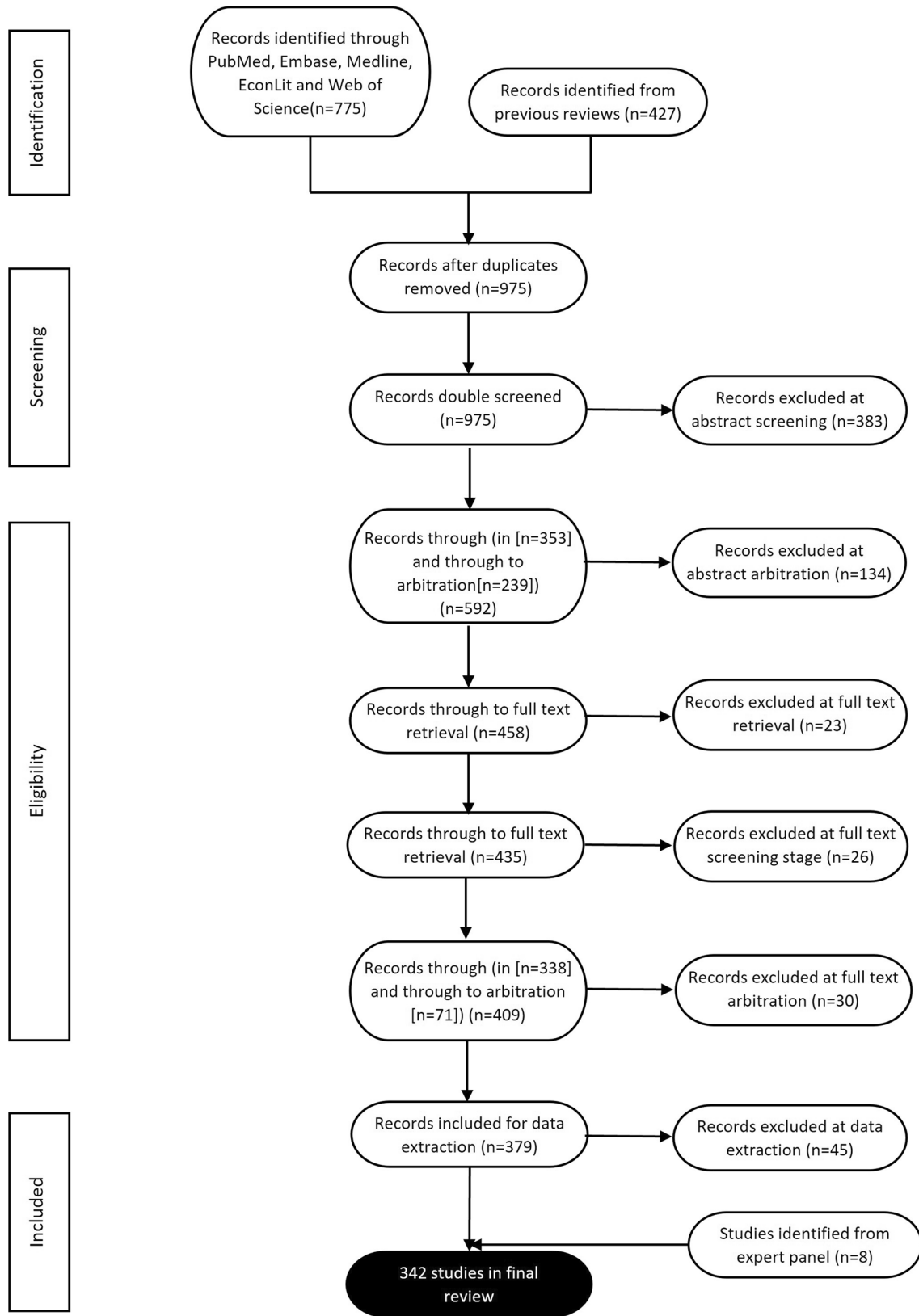
---

[1] Within 2000–2001, there were no studies among the curated articles

Identification

Records identified through PubMed, Embase, Medline, EconLit and Web of Science(n=775)

Records identified from previous reviews (n=427)

↓

Records after duplicates removed (n=975)

Screening

↓

Records double screened (n=975) → Records excluded at abstract screening (n=383)

↓

Records through (in [n=353] and through to arbitration[n=239]) (n=592) → Records excluded at abstract arbitration (n=134)

Eligibility

↓

Records through to full text retrieval (n=458) → Records excluded at full text retrieval (n=23)

↓

Records through to full text retrieval (n=435) → Records excluded at full text screening stage (n=26)

↓

Records through (in [n=338] and through to arbitration [n=71]) (n=409) → Records excluded at full text arbitration (n=30)

Included

↓

Records included for data extraction (n=379) → Records excluded at data extraction (n=45)

↓

342 studies in final review ← Studies identified from expert panel (n=8)

Fig. 1 PRISMA diagram [9]. *PRISMA* Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Fig. 2 Trend of publications from 2000 to 31 March 2020 (left); distribution of sample size of the reviewed articles (right). *IQR* interquartile range
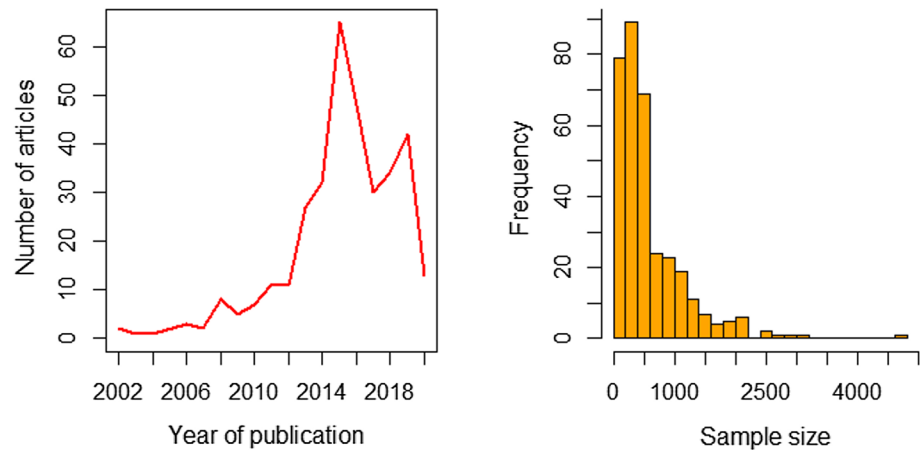

**Table 1** Background of discrete choice experiment studies

	2002–2010 N = 29 n (%)	2011–2015 N = 146 n (%)	2016–2020 N = 167 n (%)	Total N = 342 n (%)
Main objective				
Developing health policies	4 (13.8)	23 (15.8)	19 (11.4)	46 (13.5)
Device/process/screening/procedure	6 (20.7)	23 (15.8)	35 (21.1)	64 (18.8)
Health states/outcomes	2 (6.9)	2 (1.4)	5 (3.0)	9 (2.6)
Job or education choices	2 (6.9)	11 (7.5)	11 (6.6)	24 (7.0)
Treatment/medicine	10 (34.5)	72 (49.3)	81 (48.8)	163 (47.8)
Trial or other participation	0 (0.0)	2 (1.4)	1 (0.6)	3 (0.9)
Other	5 (17.2)	13 (8.9)	14 (8.4)	32 (9.4)
Study country				
Asia	0 (0.0)	9 (6.2)	15 (9.0)	24 (7.0)
Africa	1 (3.4)	3 (8.9)	13 (7.8)	27 (7.9)
Oceania	2 (6.9)	20 (13.7)	16 (9.6)	38 (11.1)
Australia	2 (6.9)	20 (13.7)	16 (9.6)	38 (11.1)
Europe	15 (51.7)	83 (56.8)	95 (56.9)	193 (56.4)
Germany	0 (0.0)	11 (7.5)	20 (12.0)	31 (9.1)
UK	6 (20.7)	21 (14.4)	21 (12.6)	48 (14.0)
North America	10 (34.5)	47 (32.2)	37 (22.2)	94 (27.5)
Canada	6 (20.7)	21 (14.4)	11 (6.6)	38 (11.1)
USA	4 (13.8)	25 (17.1)	26 (15.6)	55 (16.1)
South America	0 (0.0)	0 (0.0)	3 (1.8)	3 (0.9)
Sample type				
General population	7 (24.1)	40 (27.4)	46 (27.5)	93 (27.2)
Caregiver	2 (6.9)	10 (6.8)	16 (9.6)	28 (8.2)
Healthcare workers	3 (10.3)	22 (15.1)	27 (16.2)	52 (15.2)
Policy makers	1 (3.4)	4 (2.7)	4 (2.4)	9 (2.6)
Patients	12 (41.4)	74 (50.7)	77 (46.1)	163 (47.7)
Students	2 (6.9)	7 (4.8)	6 (3.6)	15 (4.4)
Others	3 (10.3)	7 (4.8)	10 (6.0)	20 (5.8)

Data are expressed as *n* (%)

with a median of 12 (IQR 8–16) (Table 2). Most studies (*n* = 304, 88.9%) reported information on set selection, and among them, 61% of the articles (*n* = 208) reported using some form of efficient design. The most prevalent design type is D-efficient (*n* = 185, 54.1%) (Table 2). Other prominent design types are fractional (*n* = 88, 25.7%), orthogonal

Fig. 3 Number of attributes in the included studies (left; *n* = 341); number of alternatives in the included studies excluding the reference case (e.g., opt-out, status-quo) [right; *n* = 333]. *IQR* interquartile range
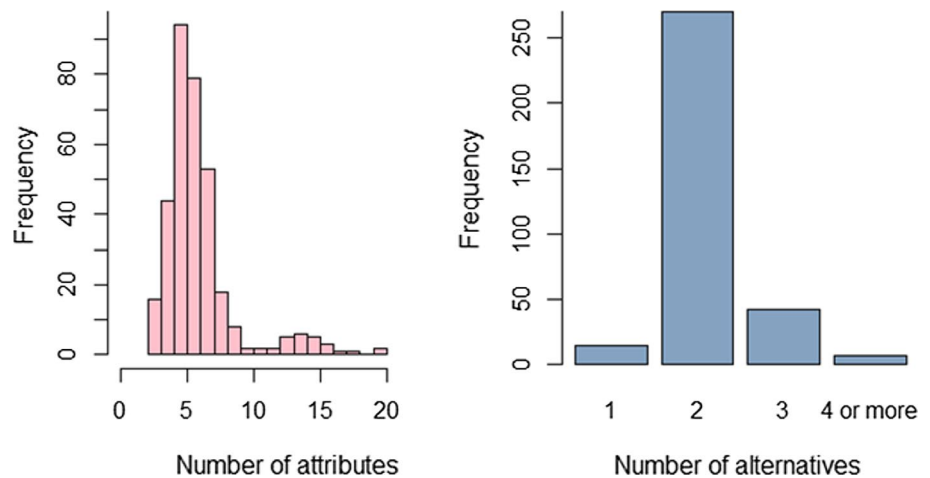
(*n* = 45, 21.9%), and balanced (*n* = 35, 10.2%); however, these terms may not distinctively describe their approach (e.g., an efficient design is fractional too). Thirty-four percent of the studies reported constructing their design accounting for the main effects only, 6% accounted for two-way or higher-order effects, and most did not report the particular information (59%).

The review demonstrated the diversity of software used in experimental design and estimation. The most prevalent software package to design the experiment was Ngene (*n* = 112, 32.7%), followed by Sawtooth (*n* = 63, 18.4%) and SAS (*n* = 43, 12.6%) (Table 2). Around 76% (*n* = 261) of the reviewed articles reported the software package used for the analysis. Among them, the highest used software package was NLOGIT (*n* = 95), followed by Stata (*n* = 93), Sawtooth/Lighthouse Studio (*n* = 27), and Latent GOLD (*n* = 23).

In their description of the results, 80.8% reported coefficient estimates, and 28.6% of the studies reported willingness-to-pay (WTP). After 2010, studies also focused on other representation formats such as odds ratio (OR) and maximum/minimum acceptable risk (MAR). In addition, the application of RI and preference share has increased substantially in recent studies after 2016 (Table 2). In addition, a few studies reported the valuation space as WTP space [33–38].

4 Current Practices to Explore Preference Heterogeneity in Healthcare DCEs

4.1 Overview

To describe alternative methods used in preference heterogeneity analysis, we classified the models based on whether heterogeneity was explained by observable factors/covariates through subgroup analysis using interaction terms

(explained preference heterogeneity) [39] and/or considered heterogeneity as unobserved (i.e., latent), which cannot be explained by observable correlates (unexplained preference heterogeneity). Accounting for unexplained preference heterogeneity needs further assumptions about the structure of latent heterogeneity [40, 41]. The major models identified to explore the unexplained heterogeneity are mixed logit (MXL) models with a continuous distribution of preference parameters (also known as RPL models), LC logit model with a discrete distribution of parameters[2], and scale heterogeneity models where scale parameter is random either with the continuous or discrete distribution [40, 42].

Among the 342 articles, the most common applied models are the MXL/RPL (*n* = 205, 59.9%), LC (*n* = 112, 32.7%), and conditional logit (CL; *n* = 90, 26.3%) models. Among the 205 analyses using the MXL/RPL, 42% (*n* = 86) were published between 2011 and 2015, and 50% (*n* = 104) were published between 2016 and 2020 (Fig. 4). A similar trend was observed in LC models, where 33% (*n* = 37) were published between 2011 and 2015, and 61% (*n* = 68) were published between 2015 and 2020 (Fig. 4). The application of LC models increased by 15% between 2011 and 2015 and 2016 and 2020, whereas the application of the MXL/RPL increased by 4%. The review also identified that the application of a particular model is often associated with software usage. Such as, classical MXL models were mostly performed in Stata (*n* = 62) or NLOGIT (*n* = 72); Bayesian mixed logit (hierarchical Bayes) models were mostly

---

[2] Both MXL/RPL and LC are mixed models that control for unexplained preference heterogeneity by estimating a distribution of preference around each estimated coefficient. We separated mixed logit/random parameter logits and latent class models to be persistent with the health preference literature and previous systematic reviews on healthcare DCEs (Soekhai et al. [2], de Bekker-Grob et al. [8])

Table 2 Alternatives, attribute and designs of discrete choice experiment studies

	2002–2010 N = 29 n (%)	2011–2015 N = 146 n (%)	2016–2020 N = 167 n (%)	Total N = 342 n (%)
Number of attributes [median (IQR)]	6 (5–7)	6 (5–7)	6 (5–7)	6 (5–7)
Number of alternatives [mode]	2 (69.0)	2 (81.1)	2 (81.2)	2 (80.1)
Labeled alternative				
Yes	5 (17.2)	25 (17.5)	20 (12.0)	50 (14.8)
No	22 (75.9)	114 (79.7)	141 (84.9)	277 (82.0)
Not sure/unclear	2 (6.9)	4 (2.8)	5 (3.0)	11 (3.3)
Reference case (opt-out, fixed status quo)				
Yes	12 (41.4)	46 (31.9)	61 (37.2)	119 (35.3)
No	16 (55.2)	95 (66.0)	94 (57.3)	205 (60.8)
Not sure/unclear	1 (3.4)	3 (2.1)	9 (5.5)	13 (3.9)
Max choice tasks/respondent [median (IQR)]	10.5 (7–16)	12 (8–16)	12 (9–15)	12 (8–16)
Experimental design				
Balanced	7 (24.1)	14 (9.6)	14 (8.4)	35 (10.2)
D-efficient (Bayesian and non-Bayesian)	5(17.2)	80 (54.8)	100 (59.9)	185 (54.1)
Foldover	2 (6.9)	7 (4.8)	1 (0.6)	10 (2.9)
Full factorial	0 (0.0)	7 (4.8)	4 (2.4)	11 (3.2)
Fractional (unspecified)	10 (34.5)	44 (30.1)	33 (19.8)	87 (25.4)
Optimal	2 (6.9)	14 (9.6)	7 (4.2)	23 (6.7)
Orthogonal	15 (51.7)	41 (28.1)	19 (11.4)	75 (21.9)
Random	1 (3.4)	6 (4.1)	7 (4.2)	14 (4.1)
Other designs	4 (13.8)	6 (4.1)	12 (7.2)	22 (6.4)
Not reported	5 (17.2)	13 (8.9)	20 (12.0)	38 (11.1)
Design software				
Ngene	0 (0.0)	34 (23.4)	77 (46.1)	111 (32.6)
SAS	3 (10.3)	23 (15.9)	17 (10.2)	43 (12.6)
Sawtooth	5 (17.2)	28 (19.3)	30 (18.0)	63 (18.5)
Other	6 (20.7)	23 (15.8)	14 (8.4)	43 (12.6)
Not reported	15 (51.7)	38 (26.0)	29 (17.4)	82 (24.0)
Estimates				
Coefficient	23 (79.3)	117 (79.6)	137 (82.0)	227 (81.0)
Odds ratio	0 (0.0)	8 (5.4)	9 (5.4)	17 (4.97)
Max/min acceptable risk	0 (0.0)	6 (4.1)	9 (5.4)	15 (4.38)
Relative importance	2 (6.9)	34 (23.1)	59 (35.3)	95 (27.77)
Willingness to pay	11 (37.9)	43 (29.3)	44 (26.3)	38 (28.66)
Preference share/choice share	4 (13.8)	27 (18.4)	30 (18.0)	61 (17.84)
Other results	2 (6.9)	33 (22.4)	36 (21.6)	71 (20.76)

Data are expressed as *n* (%) unless otherwise specified

max maximum, *min* minimum

performed through Sawtooth ($n = 18$)[3], and models were mostly analyzed by using Latent GOLD ($n = 21$), Sawtooth ($n = 23$), or NLOGIT ($n = 31$).

Around 38% ($n = 126$) of the reviewed paper explored heterogeneity using multiple models. Among them, 72%

($n = 91$) of the studies used two models, 24.4% ($n = 31$) used three different models, and only four studies used four different types of models [43–46]. Among the papers that reported estimation with two different models, 25% ($n = 23$) of them applied MXL/RPL and LC, 24% ($n = 22$) used CL and LC, and 16.8% ($n = 15$) used CL with MXL/RPL. The combination of CL, MXL/RPL, and LC is also the highest (35.4 %, $n = 11$) among the papers that used three different models for estimation. Among the papers that applied a

---

[3] Detailed discussion about the classical and Bayesian estimation approach of MXL/RPL can be found in the paper by Huber and Train [98].
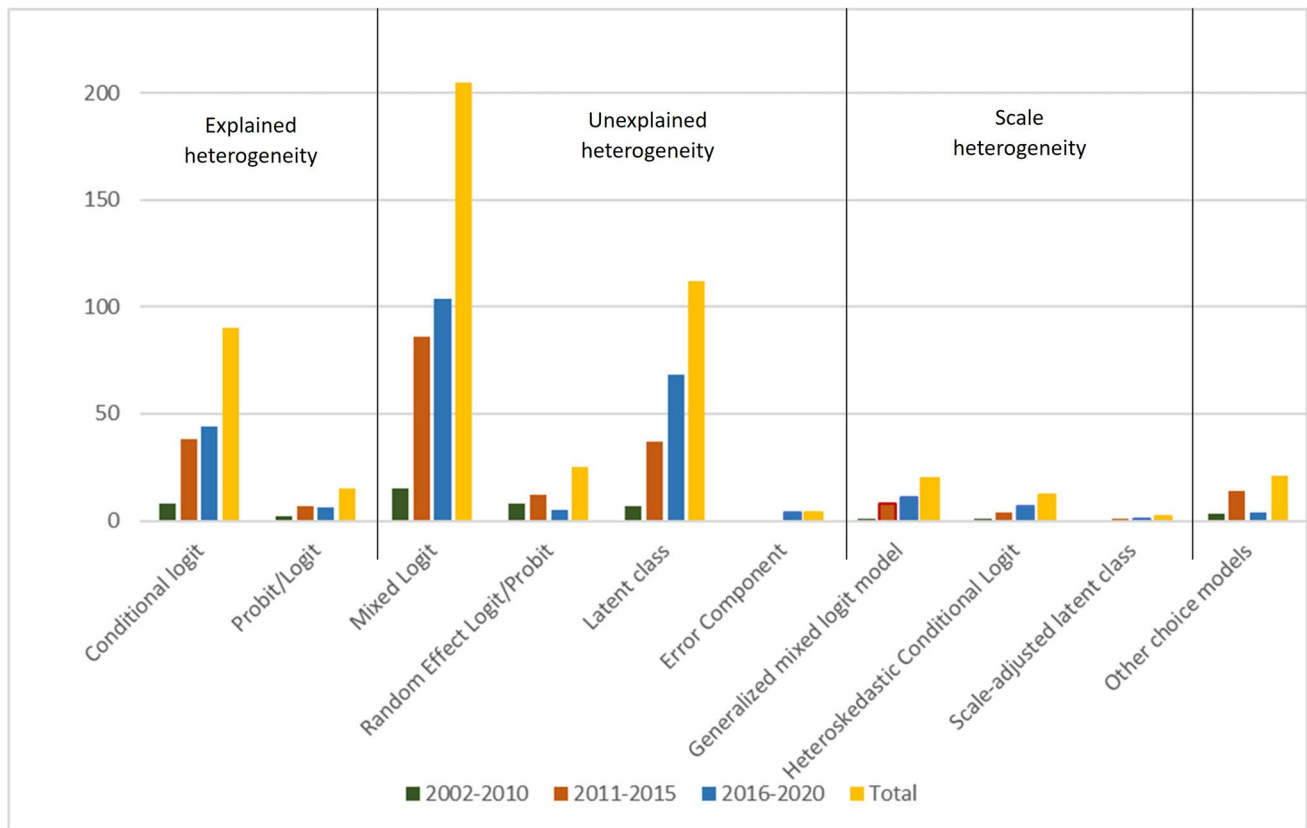
Fig. 4 Alternative models applied in the included studies

single model ($n = 216$), 59.4% ($n = 128$) of the papers used any form of the MXL model (i.e., classical, Bayesian, or generalized), followed by the LC model ($n = 43$), CL model ($n = 16$), and random-effect probit/logit ($n = 17$).

4.2 Explained Heterogeneity

By holding the assumption that preferences are correlated with observable characteristics, differences in preferences can be explained by observable (e.g., clinical and socioeconomic) factors. In this case, heterogeneity is usually modeled from the pooled sample by estimating a coefficient on the interaction between an individual-level characteristic variable and an attribute variable. The CL/multinomial logit (MNL) models have most frequently been used to identify heterogeneity through observable differences [15]. Around 25% ($n = 90$) of the reviewed papers used CL/MNL models with such an interaction [47–51]. Some other studies also included interaction within MXL/RPL and random-effects models [52–55]. Among the total reviewed articles, 38.2% ($n = 131$) of studies included subgroup analysis (e.g., where an individual-level characteristic variable has interacted with all attribute variables), and 13% ($n = 45$) of studies also included interaction between attributes.

4.3 Unexplained Heterogeneity

Rather than limiting the source of heterogeneity by deterministic segmentation, the underlying assumption can be that preferences vary across individuals due to latent factors that may or may not be easily observed. Based on this key principle, different applications capture unexplained heterogeneity by specifying the model with random preference coefficients with specific distributional assumptions and correlation structures [56]. In its most simple form, the MXL/ RPL model accounts for unexplained heterogeneity through individual deviations from the mean estimate in a continuous distribution specification. Among those reviewed, the majority of the MXL/RPL models ($n = 177$, 86.3%) were estimated using the classical estimation procedure (i.e., maximum likelihood), and the remainder used the Bayesian estimation technique (commonly known as hierarchical Bayes [HB]). Majority of the studies reported SD or the spread of the random parameters to identify the degree of preference heterogeneity, whereas some studies introduced an interaction between the mean estimate of the random parameter and a covariate to understand preference heterogeneity around the mean on the basis of the observed covariates. Of models that are required to specify distributional assumptions (e.g.,

MXL, HB, generalized multinomial logit model [GMNL]; $n = 212$), just over half ($n = 117$, 55.2%) have reported the specific distribution. Among them, most of the papers reported specifying a normal distribution ($n = 110$), followed by lognormal ($n = 15$) and triangular ($n = 2$) distributions. Many studies also specified distribution(s) for parameters that approximate reality as closely as possible (lognormal distribution for non-negative parameters, imposing constraints to control the spread/SD of the parameters, fixing coefficient[s]) [38, 57–63]. Only nine of the reported studies explicitly mentioned using the correlated distribution of preference parameters [22, 64–70]. Two studies had infused the error component model with mixed logit as a model specification [71, 72].

Some studies that used classical MXL/RPL reported further details in the specification; for example, 39.3 % ($n = 71$) reported the number and types of draws for simulation. Seventy of these used Halton draws with an average of 1528 draws (minimum 50, maximum 10,000), two studies used pseudo-random draws [73, 74], and one study used Modified Latin Hypercube Sampling (MHLS) draw with 500 draws [75]. One study also reported to have multidraws to test the stability and precision of the estimated coefficient [57]. Two studies estimated MXL using both the classical and Bayesian estimation techniques [76, 77]. Other types of random parameter models are the random-effect logit/probit model (applied in 7%, $n = 25$ studies).

In contrast to the MXL/RPL, the LC model specifies unexplained heterogeneity as classes (clusters) of individuals with similar (but unobserved) preference structures. LC models replace the continuous distribution assumption of MXL/RPL with a discrete distribution in which preference heterogeneity is captured by membership in distinct classes or segments [40]. The assumption of preference homogeneity holds within a class. Among the papers that estimated the LC model (32.2%, $n = 110$) (Fig. 4), 98% of the studies reported the class size of the best-fitted model. The range of reported best-fitted class sizes varied between two and six. Around 80% ($n = 88$) of the LC papers mentioned the model selection criteria, and the most common approach to selecting the optimal number of classes was the AIC (58.1%) [78], the Bayesian information criterion (BIC; 48.2%), and the consistent AIC (CAIC; 9.0%) [78]. Forty-one (37.2%) of the studies reported using one model selection criterion, 23 (26.1%) reported two criteria, and 22 papers reported using more than two criteria to select the best-fitted model. Studies that found conflicting recommendations for best-fitted classes by different information criteria tend to choose BIC and CAIC over AIC, as AIC was reported to overfit the model and BIC imposes more stringent criteria [24, 30, 79, 80]. In addition to information criteria, interpretability of class coefficients, sample size, adjusted R^2, entropy score, and bootstrapping procedure to obtain the log-likelihood

difference between classes have also been used to rationalize the optimal number of segments in LC models [24, 28, 30, 81–83].

In addition to the wide applications of the above models, around 7% ($n = 25$) of the studies used random effect probit/logit models to control for unobserved heterogeneity using panel specification of the data [32, 55, 84–87].

4.4 Scale Heterogeneity

In addition, to explore heterogeneity only through taste variation among respondents, some papers included the source of heterogeneity from the scale parameter (which relates to variation from the choice behavior and creates differences in the absolute effects of all attributes) and allowed for observable or unobservable differences in scale between individuals (i.e., scale heterogeneity) [7, 42]. For example, the heteroskedastic conditional logit (HCL) model ($n = 12$) allows for observable differences in scale between individuals [15]. The scale-adjusted LC model ($n = 2$) specifies scale heterogeneity as classes (clusters) of individuals with similar (but unobserved) scale structures [88, 89]. Overall, 11.7% ($n = 38$) of the articles explored scale heterogeneity in the data, and the generalized mixed logit model (GMNL) is identified as the most common model in this regard ($n = 20$).

5 Discussion

Our findings agree with the earlier systematic reviews; before 2012, 15% of the published healthcare DCEs dealt with heterogeneity, which increased to 67% from 2013 to 2017 [2]. This study provides an overview of how preference heterogeneity has been incorporated into healthcare DCEs. Most of the studies were focused on identifying heterogeneous preferences for different treatments/medications followed by preferences for different medical procedures or screening devices. A large number of studies also included heterogeneous preferences in developing health policies. In most of the studies, the targeted population was patients and the general population, followed by healthcare workers and informal caregivers. We have also identified that preference heterogeneity analyses in DCEs are becoming increasingly common across regions. The studies are conducted over a broad range of sample sizes, numbers of attributes, and numbers and types of alternatives. Furthermore, we have found a number of software packages used in analysis where STATA and NLOGIT were the most common. This can be because both software can handle CL, MXL/RPL, and LC models, which are the most frequently used models. On the other hand, some of the software are associated with a particular type of model, such as Latent GOLD with LC and Sawtooth with the hierarchical Bayes.

The trend of shifting heterogeneity analysis through random variation (i.e., discrete distribution or continuous distribution) of taste parameters rather than identifying deterministic segmentation (explained preference heterogeneity) has increased the scope of investigation by capturing a wider extent of preference heterogeneity [56]. This may be possible due to a significant increase in the performance of computational power of personal computers and straightforward implementation of different software. However, researchers should cautiously use the models by understanding the hidden assumptions of each model. Even though all the studies explicitly mentioned their model, many studies did not clearly report key details of the chosen model. For example, in analyzing heterogeneity through MXL/RPL models, a proper investigation should be done on identifying which parameters may consider as random and justification of the distributional assumption of the random parameters, whether to use multiple draws to identify the type and number of draws for simulation, and whether to account for correlation across alternatives or an individual's multiple-choice situation [90, 91]. Similarly, in the LC model, the optimization algorithm in LC logit models, the covariates to identify class membership, and the type of attribute coding. In MXL models, the typical use of Halton draws is found to be consistent with methodological papers, and even though there is no standard practice of a required number of draws, usually the number of draws increases as models become more complex (i.e., correlation of attributes and alternatives, number of attributes and alternatives) [90, 92]. Furthermore, studies with too few Halton draws might have the possibility to produce bias estimation [91]. The application of multiple models to understand the structure of heterogeneity is observed in many of the reviewed papers and is considered a good practice in different literature [1, 93]. Moreover, authors often use simple models (i.e., CL) to inform subsequent analyses (i.e., specifying utility function, select starting values). In addition, comparing with simple base-case models (i.e., CL) with heterogeneous model output, often facilitates researchers to validate the model (i.e., sign on coefficients) and to identify the merit of the presence of heterogeneity across different attributes [94–96].

Although we did not systematically assess how the preference heterogeneity analysis was reflected in discussing its implication at the practical context or policy level, the subjective categorization of reviewers identified 40% of the reviewed papers extensively discussed heterogeneity, another 40% discussed it to some extent, and the remaining 20% did not discuss heterogeneity and its implications. Those papers that discussed the heterogeneity issue extensively were also likely to do further post-estimations such as obtaining individual-level parameters, regression analysis to examine effects of individual characteristics, and estimated

distribution of MAR, WTP, or preference shares. However, the literature contained few confirmatory studies [15], even though, in the wider choice modeling field, there is mixed support regarding the ability of researchers to correctly identify and account for scale heterogeneity in their estimations. Data from this review identified a small number of papers that controlled for scale while identifying preference/taste heterogeneity, which is consistent with views from the earlier published review [3]. The usual motivation for accounting for scale heterogeneity is to either improve model fit or to avoid biased results [7, 9, 71, 97]; however, identifying the exact tools to account for scale while identifying preference/taste heterogeneity and the success of those models to separate taste and scale heterogeneity is an area under development.

5.1 Limitations

The study is limited by several weaknesses. First, we have curated only the papers that analyzed preference heterogeneity and extracted information primarily on preference heterogeneity relevant issues. Hence, it was beyond the scope of this study to directly compare any model that does not capture preference heterogeneity. Multiple reviewers were involved at the extraction stage, which may arise the possibility of different interpretations of selected literature, and which might have affected the data extraction and, as a consequence, the results presented. Furthermore, this paper summarized what is reported in reviewed papers, not necessarily what was done. To improve the consistency, the reviewers met biweekly during the data collection period to discuss any confusion, completed the beta test with the extraction tool, and each paper was reviewed by two extractors. Second, if a study included multiple sample types, the sample size showed the aggregate number. Similarly, if multiple models were used in the analysis, each paper is counted as one application of the model. Furthermore, we may have had duplicate studies where different models have been published on the same data set for different journals/audiences, therefore there is the potential for double-counting as it was too difficult to identify and remove. Lastly, the review focused on DCEs, the most popular stated preference method; however, the findings are likely to be applicable to other preference elicitation methods, such as best-worst scaling, which utilizes similar models. Future research may be conducted to understand more complex issues regarding preference heterogeneity, such as the effects of experimental design on heterogeneity analysis, covariate association to identify LC membership or subgroup analysis, accounting for correlation between choice situations or across parameters, and the influence of preference heterogeneity results on practitioners and regulators in decision making.

6 Conclusion

This systematic review has shown that preference heterogeneity analysis is a growing trend in healthcare DCEs. This study provides the current practices of a wider variety of models to account for preference heterogeneity. A large number of studies not reporting methodological details were identified, suggesting the need for greater guidance on good research practice. As heterogeneity can be assessed from different stakeholder perspectives with different methods, the research should focus more on providing complete information on analytical methods, for example, a theoretical framework for model selection, behavioral assumptions for variable selection, distributional assumption, convergence criteria, and other statistical details. The inclusion of preference heterogeneity analysis in healthcare DCEs signifies that developing guidance specifically for the analysis of unexplained heterogeneity, such as MXL/RPL and LC models, might positively impact the quality of health preference research, increase confidence in the results, and improve the ability of decision makers to act on the preference evidence.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40273-022-01178-y.

Declarations

Funding/Support No external funding was used.

Conflicts of interest/competing interest The authors declare that they have no conflict of interest.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent to publication All authors provide this consent.

Availability of data and material Please contact to the corresponding author.

Code availability Please contact the corresponding author for any requests and queries for any study materials including codes.

References

1. Zhou M, Thayer WM, Bridges JFP. Using latent class analysis to model preference heterogeneity in health: a systematic review. Pharmacoeconomics. 2018;36:175–87.
2. Soekhai V, de Bekker-Grob EW, Ellis AR, et al. Discrete choice experiments in health economics: past, present and future. Pharmacoeconomics. 2019;37:201–26.
3. Wright SJ, Vass CM, Sim G, et al. Accounting for scale heterogeneity in healthcare-related discrete choice experiments when comparing stated preferences: a systematic review. Patient. 2018;11:475–88.
4. 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;2:55–64.
5. Colombo S, Hanley N, Louviere JJ. Modeling preference heterogeneity in stated choice data: an analysis for public goods generated by agriculture. Agric Econ. 2009;40:307–22.
6. Hensher DA, Rose JM, Greene WH. Applied choice analysis. 2nd ed. Cambridge: Cambridge University Press; 2015.
7. Groothuis-Oudshoorn CGM, Flynn TN, Yoo HI, et al. Key issues and potential solutions for understanding healthcare preference heterogeneity free from patient-level scale confounds. Patient. 2018;11:463–6.
8. de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. Health Econ. 2012;21:145–72.
9. Vass C, Boeri M, Karim S, et al. Accounting for Preference Heterogeneity in Discrete-Choice Experiments: An ISPOR Special Interest Group Report. Value in Health. 2022;25:685–94.
10. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372: n71.
11. Clark MD, Determann D, Petrou S, et al. Discrete choice experiments in health economics: a review of the literature. Pharmacoeconomics. 2014;32:883–902.
12. The EndNote Team. EndNote. EndNote X9 ed. Philadelphia: Clarivate; 2013.
13. Mansfield C, Sikirica MV, Pugh A, et al. Patient preferences for attributes of type 2 diabetes mellitus medications in Germany and Spain: an online discrete-choice experiment survey. Diabetes Therapy. 2017;8:1365–78.
14. Manski CF. The structure of random utility models. Theor Decis. 1977;8:229.
15. Craig BM, de Bekker-Grob EW, González JM, et al. A guide to observable differences in stated preference evidence. Patient. 2022;15(3):329–39.
16. Louviere JJ, Flynn TN, Marley AAJ. Best-worst scaling: theory, methods and applications. Cambridge: Cambridge University Press; 2015.
17. Orme BK. Getting started with conjoint analysis: strategies for product design and pricing research. Madison: Research Publishers; 2010.
18. Craig BM, Rand K, Hartman JD. Preference paths and their Kaizen tasks for small samples. Patient. 2022;15(2):187–96.
19. Craig BM, Busschbach JJ, Salomon JA. Modeling ranking, time trade-off, and visual analog scale values for EQ-5D health states: a review and comparison of methods. Med Care. 2009;47:634–41.
20. Wijnen BFM, Van Engelen RPLB, Ostermann J, et al. A discrete choice experiment to investigate patient preferences for HIV testing programs in Bogota, Colombia. Expert Rev Pharmacoecon Outcomes Res. 2019;19:195–201.
21. Goossens AJ, Cheung KL, Sijstermans E, et al. A discrete choice experiment to assess patients' preferences for HIV treatment in the rural population in Colombia. J Med Econ. 2020;8:1.
22. 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;109:633–40.
23. Sijstermans E, Cheung KL, Goossens AJ, et al. A discrete choice experiment to assess patients' preferences for HIV treatment in the urban population in Colombia. J Med Econ. 2020;23(8):803–11.
24. Cunningham CE, Barwick M, Short K, et al. Modeling the mental health practice change preferences of educators: a discrete-choice conjoint experiment. School Ment Health. 2014;6:1–14.
25. Cunningham CE, Kostrzewa L, Rimas H, et al. Modeling organizational justice improvements in a pediatric health service. Patient. 2013;6:45–59.

26. Cunningham CE, Walker JR, Eastwood JD, et al. Modeling mental health information preferences during the early adult years: a discrete choice conjoint experiment. J Health Commun. 2014;19:413–40.

27. Cunningham CE, Henderson J, Niccols A, et al. Preferences for evidence-based practice dissemination in addiction agencies serving women: a discrete-choice conjoint experiment. Addiction. 2012;107:1512–24.

28. Cunningham CE, Barwick M, Rimas H, et al. Modeling the decision of mental health providers to implement evidence-based children's mental health services: a discrete choice conjoint experiment. Adm Policy Ment Health. 2018;45:302–17.

29. Cunningham CE, Deal K, Rimas H, et al. Modeling the information preferences of parents of children with mental health problems: a discrete choice conjoint experiment. J Abnorm Child Psychol. 2008;36:1123–38.

30. Cunningham CE, Rimas H, Chen Y, et al. Modeling parenting programs as an interim service for families waiting for children's mental health treatment. J Clin Child Adolesc Psychol. 2015;44:616–29.

31. Cunningham CE, Deal K, Neville A, et al. Modeling the problem-based learning preferences of McMaster University undergraduate medical students using a discrete choice conjoint experiment. Adv Health Sci Educ Theory Pract. 2006;11:245–66.

32. Ratcliffe J, Van Haselen R, Buxton M, et al. Assessing patients' preferences for characteristics associated with homeopathic and conventional treatment of asthma: a conjoint analysis study. Thorax. 2002;57:503–8.

33. Ozdemir S, Johnson FR, Hauber AB. Hypothetical bias, cheap talk, and stated willingness to pay for health care. J Health Econ. 2009;28:894–901.

34. Michaels-Igbokwe C, Terris-Prestholt F, Lagarde M, et al. Young people's preferences for family planning service providers in Rural Malawi: a discrete choice experiment. PLoS ONE. 2015;10(12): e0143287.

35. Marshall DA, MacDonald KV, Heidenreich S, et al. The value of diagnostic testing for parents of children with rare genetic diseases. Genet Med. 2019;21:2798–806.

36. Kjaer T, Bech M, Kronborg C, et al. Public preferences for establishing nephrology facilities in Greenland: estimating willingness-to-pay using a discrete choice experiment. Eur J Health Econ. 2013;14:739–48.

37. Grisolía JM, Longo A, Boeri M, et al. Trading off dietary choices, physical exercise and cardiovascular disease risks. Soc Sci Med. 2013;93:130–8.

38. Hole AR, Kolstad JR. Mixed logit estimation of willingness to pay distributions: a comparison of models in preference and WTP space using data from a health-related choice experiment. Empirical Econ. 2012;42:445–69.

39. Craig BM, de Bekker-Grob EW, González Sepúlveda JM, et al. A guide to observable differences in stated preference evidence. Patient. 2022;15(3):329–39.

40. Greene WH, Hensher DA. A latent class model for discrete choice analysis: contrasts with mixed logit. Transport Res Part B Methodol. 2003;37:681–98.

41. Wedel M, Kamakura W, Arora N, et al. Discrete and continuous representations of unobserved heterogeneity in choice modeling. Mark Lett. 1999;10:219–32.

42. Vass CM, Wright S, Burton M, et al. Scale heterogeneity in healthcare discrete choice experiments: a primer. Patient. 2018;11:167–73.

43. Mulhern B, Norman R, Lourenco RD, et al. Investigating the relative value of health and social care related quality of life using a discrete choice experiment. Soc Sci Med. 2019;233:28–37.

44. Kaambwa B, Ratcliffe J, Shulver W, et al. Investigating the preferences of older people for telehealth as a new model of health care service delivery: a discrete choice experiment. J Telemed Telecare. 2017;23:301–13.

45. Kaambwa B, Lancsar E, McCaffrey N, et al. Investigating consumers' and informal carers' views and preferences for consumer directed care: a discrete choice experiment. Soc Sci Med. 2015;140:81–94.

46. Ride J, Lancsar E. Women's preferences for treatment of perinatal depression and anxiety: a discrete choice experiment. PLoS ONE. 2016;11(6): e0156629.

47. Luyten J, Kessels R, Goos P, et al. Public preferences for prioritizing preventive and curative health care interventions: a discrete choice experiment. Value Health. 2015;18:224–33.

48. Koopmanschap MA, Stolk EA, Koolman X. Dear policy maker: have you made up your mind? A discrete choice experiment among policy makers and other health professionals. Int J Technol Assess Health Care. 2010;26:198–204.

49. Sawamura K, Sano H, Nakanishi M. Japanese public long-term care insured: preferences for future long-term care facilities, including relocation, waiting times, and individualized care. J Am Med Dir Assoc. 2015;16(4):350.e9-20.

50. Paolucci F, Mentzakis E, Defechereux T, et al. Equity and efficiency preferences of health policy makers in China—a stated preference analysis. Health Policy Plan. 2015;30:1059–66.

51. Moise N, Wood D, Cheung YKK, et al. Patient preferences for personalized (N-of-1) trials: a conjoint analysis. J Clin Epidemiol. 2018;102:12–22.

52. Walker R, Morton R, Palmer S, et al. Patient preferences for dialysis modalities: a discrete-choice experiment. Nephrology. 2017;22:48.

53. Arbiol J, Yabe M, Nomura H, et al. Using discrete choice modeling to evaluate the preferences and willingness to pay for leptospirosis vaccine. Hum Vaccin Immunother. 2015;11:1046–56.

54. Kruk ME, Riley PL, Palma AM, et al. How can the health system retain women in HIV treatment for a lifetime? A discrete choice experiment in Ethiopia and Mozambique. PLoS ONE. 2016;11(8): e0160764.

55. Chu L-W, So JC, Wong L-C, et al. Community end-of-life care among Chinese older adults living in nursing homes. Geriatr Gerontol Int. 2014;14:273–84.

56. Hess S. Latent class structures: taste heterogeneity and beyond. In: Hess S, Daly AJ, editors. Handbook of choice modelling. New York: Edward Elgar Publishing; 2014.

57. Van Puyvelde S, Caers R, Du Bois C, et al. Does organizational ownership matter? Objectives of employees in public, nonprofit and for-profit nursing homes. Appl Econ. 2015;47:2500–13.

58. Kjae T, Gyrd-Hansen D. Preference heterogeneity and choice of cardiac rehabilitation program: results from a discrete choice experiment. Health Policy. 2008;85:124–32.

59. Ikenwilo D, Heidenreich S, Ryan M, et al. The best of both worlds: an example mixed methods approach to understand men's preferences for the treatment of lower urinary tract symptoms. Patient. 2018;11:55–67.

60. Vennedey V, Derman SHM, Hiligsmann M, et al. Patients' preferences in periodontal disease treatment elicited alongside an IQWiG benefit assessment: a feasibility study. Patient Prefer Adherence. 2018;12:2437–47.

61. Mankowski C, Ikenwilo D, Heidenreich E, et al. Men's preferences for the treatment of lower urinary tract symptoms associated with benign prostatic hyperplasia: a discrete choice experiment. Patient Prefer Adherence. 2016;10:2407–17.

62. Deidda M, Meleddu M, Pulina M. Potential users' preferences towards cardiac telemedicine: a discrete choice experiment investigation in Sardinia. Health Policy Technol. 2018;7:125–30.

63. Hole AR. Modelling heterogeneity in patients' preferences for the attributes of a general practitioner appointment. J Health Econ. 2008;27:1078–94.

64. Michaels-Igbokwe C, Lagarde M, Cairns J, et al. Designing a package of sexual and reproductive health and HIV outreach services to meet the heterogeneous preferences of young people in Malawi: results from a discrete choice experiment. Health Econ Rev. 2015. https://doi.org/10.1186/s13561-015-0046-6.

65. Campbell HE, Gray AM, Watson J, et al. Preferences for interventions designed to increase cervical screening uptake in non-attending young women: how findings from a discrete choice experiment compare with observed behaviours in a trial. Health Expect. 2020;23:202–11.

66. Brown DS, Poulos C, Johnson FR, et al. Adolescent girls' preferences for HPV vaccines: a discrete choice experiment. Adv Health Econ Health Serv Res. 2014;24:93–121.

67. Hall J, Fiebig DG, King MT, et al. What influences participation in genetic carrier testing? Results from a discrete choice experiment. J Health Econ. 2006;25:520–37.

68. Tayyari Dehbarez N, Raun Morkbak M, Gyrd-Hansen D, et al. Women's preferences for birthing hospital in Denmark: a discrete choice experiment. Patient. 2018;11:613–24.

69. van de Wetering L, van Exel J, Bobinac A, et al. Valuing QALYs in relation to equity considerations using a discrete choice experiment. Pharmacoeconomics. 2015;33:1289–300.

70. Howard K, Salkeld GP, Patel MI, et al. Men's preferences and trade-offs for prostate cancer screening: a discrete choice experiment. Health Expect. 2015;18:3123–35.

71. de Bekker-Grob EW, Donkers B, Bliemer MCJ, et al. Can healthcare choice be predicted using stated preference data? Soc Sci Med. 2020;246: 112736.

72. Boeri M, Szegvari B, Hauber B, et al. From drug-delivery device to disease management tool: a study of preferences for enhanced features in next-generation self-injection devices. Patient Prefer Adherence. 2019;13:1093–110.

73. Wong CKH, Man KKC, Ip P, et al. Mothers' preferences and willingness to pay for human papillomavirus vaccination for their daughters: a discrete choice experiment in Hong Kong. Value Health. 2018;21:622–9.

74. Abiiro GA, Torbica A, Kwalamasa K, et al. What factors drive heterogeneity of preferences for micro-health insurance in rural Malawi? Health Policy Plan. 2016;31:1172–83.

75. Klojgaard ME, Hess S. Understanding the formation and influence of attitudes in patients' treatment choices for lower back pain: testing the benefits of a hybrid choice model approach. Soc Sci Med. 2014;114:138–50.

76. Negrin MA, Pinilla J, Leon CJ. Willingness to pay for alternative policies for patients with Alzheimer's Disease. Health Econ Policy Law. 2008;3:257–75.

77. Johnson FR, Ozdemir S, Phillips KA. Effects of simplifying choice tasks on estimates of taste heterogeneity in stated-choice surveys. Soc Sci Med. 2010;70:183–90.

78. Horby PW, Roddick A, Spata E, et al. Azithromycin in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial. Lancet. 2021;397:605–12.

79. Peyron C, Pelissier A, Bejean S. Preference heterogeneity with respect to whole genome sequencing. A discrete choice experiment among parents of children with rare genetic diseases. Soc Sci Med. 2018;214:125–32.

80. Ammi M, Peyron C. Heterogeneity in general practitioners' preferences for quality improvement programs: a choice experiment and policy simulation in France. Health Econ Rev. 2016;6:44.

81. Cunningham CE, Hutchings T, Henderson J, et al. Modeling the hospital safety partnership preferences of patients and their families: a discrete choice conjoint experiment. Patient Prefer Adherence. 2016;10:1359–72.

82. Jarvis W, Pettigrew S. The relative influence of alcohol warning statement type on young drinkers' stated choices. Food Qual Prefer. 2013;28:244–52.

83. Sadler A, Shi L, Bethge S, et al. Incentives for blood donation: a discrete choice experiment to analyze extrinsic motivation. Transfus Med Hemother. 2018;45:116–24.

84. Zweifel P, Telser H, Vaterlaus S. Consumer resistance against regulation: The case of health care. J Regul Econ. 2006;29:319–32.

85. Muhlbacher AC, Junker U, Juhnke C, et al. Chronic pain patients' treatment preferences: a discrete-choice experiment. Eur J Health Econ. 2015;16:613–28.

86. Murchie P, Norwood PF, Pietrucin-Materek M, et al. Determining cancer survivors' preferences to inform new models of follow-up care. Br J Cancer. 2016;115:1495–503.

87. Norman R, Hall J, Street D, et al. Efficiency and equity: a stated preference approach. Health Econ. 2013;22:568–81.

88. Erdem S, Thompson C. Prioritising health service innovation investments using public preferences: a discrete choice experiment. BMC Health Serv Res. 2014;14:360.

89. Vass CM, Rigby D, Payne K. Investigating the heterogeneity in women's preferences for breast screening: does the communication of risk matter? Value Health. 2018;21:219–28.

90. Hensher DA, Greene WH. The mixed logit model: the state of practice. Transportation. 2003;30:133–76.

91. Ellis A, de Bekker-Grob E, Howard K, et al. Number of Halton draws required for valid random parameter estimation with discrete choice data. Patient. 2019;12:432.

92. Czajkowski M, Budziński W. Simulation error in maximum likelihood estimation of discrete choice models. J Choice Model. 2019;31:73–85.

93. Lancsar E, Fiebig DG, Hole AR. Discrete choice experiments: a guide to model specification, estimation and software. Pharmacoeconomics. 2017;35:697–716.

94. Meads DM, O'Dwyer JL, Hulme CT, et al. Patient Preferences for Pain Management in Advanced Cancer: Results from a Discrete Choice Experiment. Patient. 2017;10:643–51.

95. Copsey B, Buchanan J, Fitzpatrick R, et al. Duration of treatment effect should be considered in the design and interpretation of clinical trials: results of a discrete choice experiment. Med Decis Making. 2019;39:461–73.

96. Miners A, Nadarzynski T, Witzel C, et al. Preferences for HIV testing services among men who have sex with men in the UK: a discrete choice experiment. PLoS Med. 2019;16: e1002779.

97. Milte R, Ratcliffe J, Chen G, et al. Cognitive overload? An exploration of the potential impact of cognitive functioning in discrete choice experiments with older people in health care. Value Health. 2014;17:655–9.

98. Huber J, Train K. On the similarity of classical and Bayesian estimates of individual mean partworths. Mark Lett. 2001;12:259–69.