



Stated-Preference Survey Design and Testing in Health Applications

Deborah A. Marshall¹ · Jorien Veldwijk^{2,3} · Ellen M. Janssen⁴ · Shelby D. Reed⁵

Accepted: 27 December 2023

© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2024

Abstract

Following the conceptualization of a well-formulated and relevant research question, selection of an appropriate stated-preference method, and related methodological issues, researchers are tasked with developing a survey instrument. A major goal of designing a stated-preference survey for health applications is to elicit high-quality data that reflect thoughtful responses from well-informed respondents. Achieving this goal requires researchers to design engaging surveys that maximize response rates, minimize hypothetical bias, and collect all the necessary information needed to answer the research question. Designing such a survey requires researchers to make numerous interrelated decisions that build upon the decision context, selection of attributes, and experimental design. Such decisions include considering the setting(s) and study population in which the survey will be administered, the format and mode of administration, and types of contextual information to collect. Development of a survey is an interactive process in which feedback from respondents should be collected and documented through qualitative pre-test interviews and pilot testing. This paper describes important issues to consider across all major steps required to design and test a stated-choice survey to elicit patient preferences for health preference research.

Key Points for Decision Makers

Stated-preference survey design and testing in health applications requires careful planning and development, with multiple decisions influenced by the decision context, study attributes, the study population and setting, technical aspects of the survey and iterative testing.

The goal of survey design is to develop a survey that provides high-quality stated-preference data reflecting thoughtful responses from well-informed and engaged respondents to inform a health-related decision problem.

The guidance in this paper covers the essential elements of designing and testing a stated-preference survey that reflects evidence from the literature and practical experience to increase the quality of the resulting data.

✉ Ellen M. Janssen
ejansse5@its.jnj.com

¹ Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

² Erasmus School of Health Policy and Management, Erasmus University Rotterdam, Rotterdam, The Netherlands

³ Erasmus Choice Modeling Centre, Erasmus University Rotterdam, Rotterdam, The Netherlands

⁴ Janssen Research and Development, Titusville, NJ, USA

⁵ Preference Evaluation Research Group, Duke Clinical Research Institute and Department of Population Health Sciences, Duke University, Durham, NC, USA

1 Introduction

Stated preferences in health preferences research can be described as qualitative or quantitative assessments of the relative desirability or acceptability of features that differ

among alternative health states, health interventions, or health services [1, 2]. This paper focuses on the design of quantitative, stated-preference surveys rather than surveys designed to measure revealed-choice behavior or preferences elicited with rating scales or qualitative methods. A major goal of designing a stated-preference survey is to elicit high-quality stated-preference data that reflect thoughtful responses from well-informed and engaged respondents to inform a given decision problem in healthcare. To attain this goal, researchers should aim to design a survey that is easily understood by respondents and encourages choice making wherein hypothetical bias is minimized. This requires researchers to make numerous interrelated decisions that build upon the decision context, selection of attributes, and experimental design for the study.

After having clearly defined the research question of the preference study at hand, there are many design considerations that are essential to develop a high-quality preference survey, ranging from refining the research question, conducting qualitative research in order to define the attributes and levels, and creating an experimental design through to analyzing data and interpreting results. These steps and considerations have been captured in good research practice guidance documents in the literature [3–6]. We are building on this foundational work to focus on preference survey and instrument design more specifically.

We differentiate between the preference survey instrument (the actual text, choice tasks, and graphics used to elicit preference data) and the entire survey, which includes

the setting and study population considerations, format and mode of administration, and contextual information. The overall survey may also include multiple other survey instruments in addition to the preference survey instrument. This is because the ‘quality’ or appropriateness of the survey instrument can only be evaluated in the context of these other factors. We have focused on self-completed, online and electronic preference surveys since this is currently the most common mode of administration. However, the principles related to preference survey design still apply to most other modes of administration. One exception of note is surveys that are administered by an interviewer or that are interview assisted, whether by virtual means, phone, or in-person, rather than self-completed.

There are several aspects of survey design to consider in advance of the more technical aspects of survey instrument development, including the setting(s) and respondent characteristics, format options and mode of administration, and metadata or paradata (information to contextualize the survey instrument and respondents) and technology platforms (Fig. 1). In the next stage (Fig. 2), researchers will begin to build the survey instrument by preparing respondents to answer the preference-elicitation questions, creating the layouts for the preference-elicitation questions, and assembling or developing items to collect all necessary respondent information. This includes drafting content to clearly communicate all study-specific background information, questions, and instructions in non-technical language for respondents. After the survey is drafted, researchers must

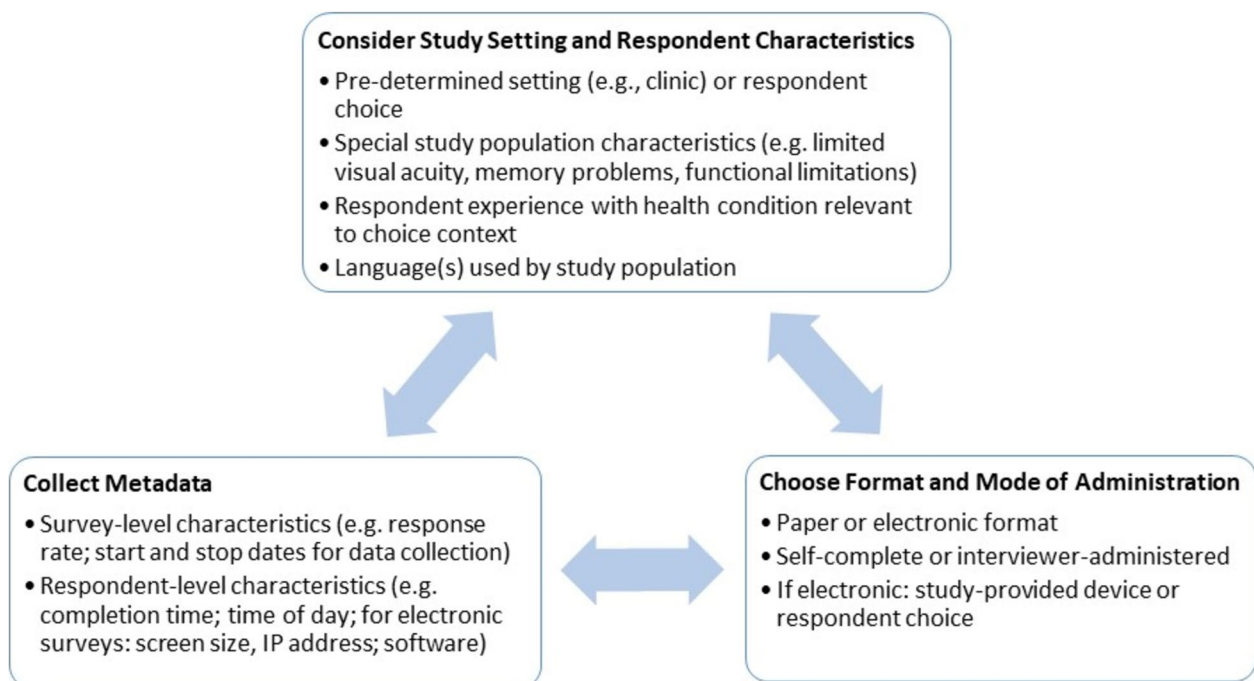


Fig. 1 Overview of survey design considerations

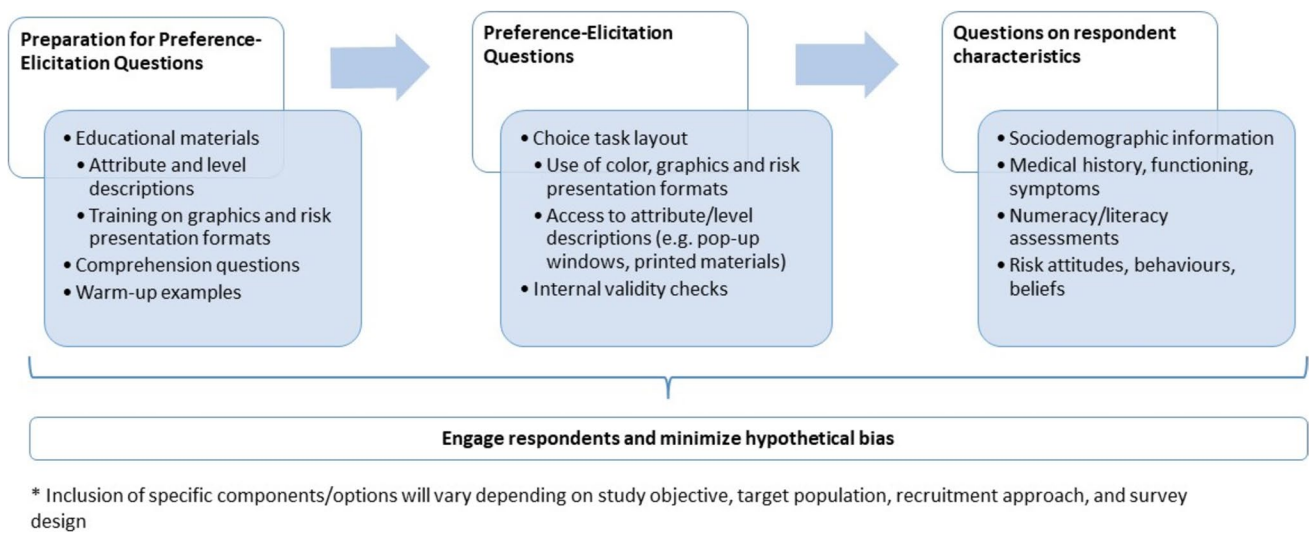


Fig. 2 Survey instrument content

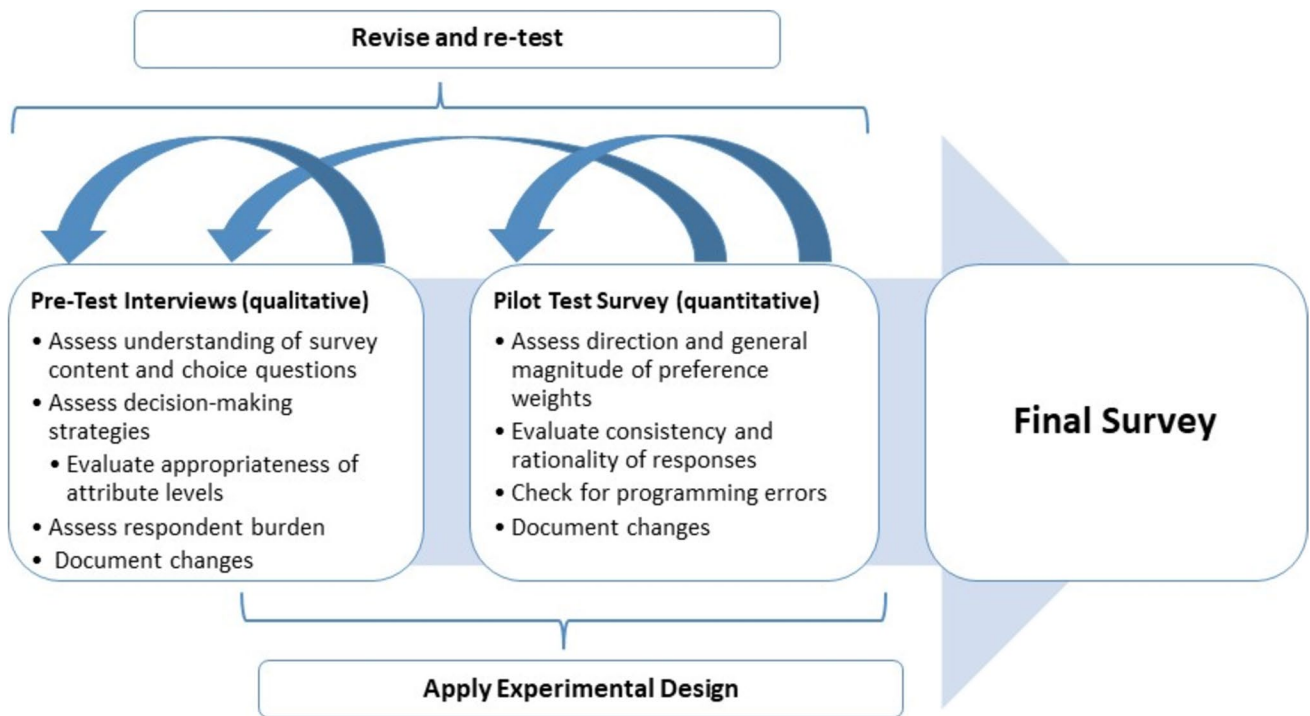


Fig. 3 Testing, revising, and finalizing survey

take care to conduct qualitative and quantitative testing of the survey (Fig. 3). In qualitative pretest interviews, they must inquire and listen to feedback from individuals that resemble those that might complete the final survey version to guide and document revisions to the survey. In quantitative pilot tests, responses of a small sample of respondents are examined to identify choice inconsistencies or other stumbling blocks. Testing of the survey is in an iterative

process with potentially multiple rounds of pretesting and pilot testing before full launch of the final survey informed by the experimental design.

This paper provides guidance based on the available evidence in the literature as well as practical experience of the authors in conducting preferences studies on how to design and test a stated-preference survey given the decision context and selected attributes. This paper focuses on the design

of stated-preference survey instrument to elicit preference data from patients, specifically. However, many of the learnings can be applied to the design of stated-preference survey instruments targeted towards other types of respondents (e.g., caregivers or clinicians).

2 Survey Design Considerations

In general, a preference survey provides insights on how respondents approach a decision as described by certain information, and the purpose of a preferences study is fundamentally about how the preference information from the study will be used. Clarifying the purpose of the preferences study, the relevance of the study to decision makers, and the potential for informing a decision are central aspects of survey design (Fig. 1) [7].

2.1 Setting and Respondent Characteristics

Before drafting a survey, researchers should give careful consideration to the setting in which it will be administered and who is being surveyed.

1. *Setting*: Will the respondent complete the survey on their own time in a setting of their choosing, or will the respondent complete the survey in specific setting such as a clinic waiting room?
2. *Level of assistance*: Will the respondent be capable of completing the survey with no supervision, or will the respondent need assistance to initiate and/or complete the survey?
3. *Special characteristics*: Does the target population have any special characteristics and accessibility considerations (e.g., children, individuals with cognitive limitations, impaired vision or hearing, low literacy, language choices) that will influence survey design?
4. *Experience*: To what extent will respondents have experience with the health condition and attributes selected for the study?

2.2 Format Options and Mode of Administration

First, researchers must decide whether the survey is intended to be self-completed by the respondent or whether it will be interviewer administered. They should consider the logistics required, including the need for an interview guide and training, as well as the potential for social-response bias with having an interviewer. However, for specific types of respondents, such as young children or people with cognitive impairment, interviewer-assisted administration may be necessary to acquire high-quality data. Next, researchers must further decide whether the survey will be paper-based or

electronic. Although not definitive, a comparison of results from a discrete choice experiment (DCE) survey administered online and by paper found no differences in preferences. However, the response rate was lower and the cost was higher for the paper survey versus the online survey [7]. Although paper-based surveys may sometimes be necessary (e.g., respondents who experience difficulties with electronic devices or with particular visual impairments), they often impose additional burden on the study team in terms of distributing surveys, collecting surveys, and entering data. Electronic surveys provide researchers with some advantages since they directly collect responses and can incorporate logic rules to identify potential data-entry errors and maximize data completion. Additionally, electronic surveys facilitate the inclusion of technical advances such as video clips to convey information and allow respondents' choices about language, font size, and use of voice overs to use with, or instead of, reading text.

A particular advantage of using electronic surveys is the opportunity to collect data using smartphones, which offer the possibility of efficiently reaching specific populations such as youth and populations without computer access [8]. However, when developing survey content, researchers should be mindful how well it is displayed on smaller devices such as tablets or smartphones. In addition, survey respondents using smartphones are more likely to multitask and be around other people, although this may not have a significant impact on data quality [8–10].

2.3 Metadata and Technology Platforms for Completing the Survey

Metadata in the context of survey research can be broad, encompassing many aspects of research data and processes, such as features at the survey level (e.g., response rate), at the respondent level (e.g., IP address), and the survey-item level (e.g., time required to respond to each question). Some metadata are automatically generated using software and often available at no additional cost. For example, survey software often can identify the mode or screen size used when the survey was completed, which allows the researcher to identify respondents who viewed the survey on different types of devices, platforms, or formats. Other metadata must be manually collected; thus, researchers need to specify in advance which information they would like to collect. Researchers should consider at the outset which contextual factors may provide them with valuable insights about preference patterns, dropouts, or other behaviors of respondents. As technology advances, the ability to collect alternative types of metadata will likely evolve and offer researchers opportunities for new insights to improve response rates, improve respondent engagement, and identify factors that may influence survey responses.

3 Survey Instrument Development

3.1 Engaging Respondents and Minimizing Hypothetical Bias

An overarching aim of survey development is to build an engaging instrument that encourages respondents to internalize information about study attributes and the decision context, and to carefully consider each choice task (Fig. 2).

Unless respondents consider their participation important, they will more likely rely on decision heuristics when selecting preferred alternatives across choice tasks [11–13]. Hypothetical bias is frequently cited as a limitation of stated-preference research [14–16]. However, there are several approaches to reduce hypothetical bias. *Ex-ante* approaches, or approaches implemented before the respondent starts answering survey questions, include administering an honesty oath, ‘cheap talk,’ emphasizing consequentiality, and minimizing social desirability bias.¹⁶ For more information, see the electronic supplementary materials.

To better engage respondents, aesthetic design principles can be applied throughout a survey to minimize unnecessary cognitive burden. These include use of clean, easy-to-read font for text, and minimalist styles and backgrounds. Each survey page/screen should include plenty of white space devoid of text and figures. Use of loading bars showing progress through the survey might help increase the proportion of completed surveys.

3.2 Preparing Respondents for Preference Survey Instrument Elicitation Questions

3.2.1 Educational Materials

To prepare respondents for completing the preference-elicitation task, researchers are advised to provide respondents with training related on the content and meaning of all attributes and levels that are included in the study as well as how to complete the choice questions.¹⁷ Such training is essential to ensure that respondents understand and consistently interpret the attributes and attribute levels to contribute to the validity of a preference study’s results.

3.2.1.1 Attribute and Level Descriptions Attributes should be described in non-technical language. Descriptions should be clear and specific so that respondents interpret attributes and levels in a comparable manner [8, 17]. Wherever possible, standard clinical definitions or descriptions that have been used in previous preference studies should be critically reviewed. When previously

used attribute descriptions are deemed to be unambiguous and appropriate, they should be considered for use to improve comparability across studies; thus, facilitating systematic literature reviews and benefit-transfer studies [18]. Researchers should be aware of framing effects [19] and present attributes and levels in a consistent manner.

While it is important to provide clear and complete descriptions, it is also critical to limit the length of descriptions to that which is sufficient and necessary for respondents to complete the choice tasks. Providing respondents with too much information might cause (1) respondents to miss the essential pieces, (2) a drop in attention, and (3) increased dropout rates. When each attribute level has a distinct meaning, all levels should be described separately. This is typically the case for nominal (e.g., modes of administration) or ordinal (e.g., mild, moderate, severe) levels. Separate descriptions often can be avoided for attribute levels measured on interval or ratio scales (e.g., risk, out-of-pocket cost). In such instances, one might consider providing an example or describing only selected levels in detail to avoid repetition and reduce burden for respondents.

3.2.1.2 Risk Grid Training Preference studies in healthcare and public health frequently include one or multiple risk attributes. The interpretation and understanding of risks are known to be complicated for respondents [20]. Because few studies have experimentally compared alternative risk communication approaches in the context of DCEs, researchers are advised to consider current best practices for presenting risks. One expert panel recently recommended the use of consistent denominators and risk formats across outcomes, avoiding ‘1 in x’ formats, and specifying the relevant time period [21, 22]. Some experts recommend presenting risk levels in more than one format to increase understanding across respondents [21] such as presenting a risk as a percentage as well as a frequency [22], but combining formats is not always advantageous [23].

3.2.1.3 Use of Videos Use of videos has shown both positive and negative effects [24–26]. Using videos to explain attributes and attribute levels and the choice context can lead to increased choice consistency and understanding among respondents [27, 28]. However, educational materials should be carefully designed, as simply providing a video does not necessarily lead to better engagement [29]. Besides relatively ‘simple’ instructional videos, researchers could opt to develop a serious game (i.e., interactive learning tool to offer experimental and engaging learning of complex skills and information). Although the use of serious games has been shown to be effective in several knowledge or behavioral interventions [30], their applicability and effectiveness in preference research is in need of further research.

3.2.2 Comprehension Questions

It is recommended that researchers consider how to assess respondents' attention to and understanding of a study's attributes and levels as well as its choice tasks. One option is to include quiz questions about attribute descriptions or information represented in an example choice task. Incorrect responses can serve as a potential indicator that the respondent did not understand the content or did not read the content. Jointly evaluating respondents' performance on quiz questions with response times can be useful to identify individuals who performed poorly on quiz questions and sped through the survey, suggesting inattentiveness rather than difficulty understanding the information. Although there are numerous ways to assess attentiveness and understanding, there are no clear decision rules or thresholds about what is considered acceptable or not. In the pre-test and pilot-test phases of survey development, such assessments provide a basis for researchers to identify any concerns and refine the survey instrument. Inclusion of quiz questions in the fielded survey can also be useful in the analysis phase to test whether respondents providing incorrect responses expressed different preference patterns or greater error variance compared to respondents with correct responses.

3.2.3 Warm-up Examples

Providing warm-up examples of the choice task can be useful in preparing respondents. If the choice task is complex, the researcher can include multiple warm-up examples that use a subset of the attributes to build up to the full choice task. Also, with ordered attributes, researchers can design a warm-up example in which one profile dominates the alternative profile (i.e., greater benefits and lower risks) to teach respondents how to interpret and answer the choice tasks and to identify respondents who may not understand the choice task.

3.3 Layouts for Preference Elicitation Questions

3.3.1 Format (Paper or Electronic)

Several decisions have to be made in regard to the layout of preference elicitation tasks. First, researchers present only one choice task per page (paper surveys) or screen (electronic survey), since preference surveys require that respondents independently consider each choice task and consider all information presented in those tasks [31]. To provide a natural mapping between the instructions, choice profiles, and the response option [32], preferably all elements of the choice task are presented on the same page or screen and need to be fully visible to respondents. For electronic surveys, this means that the entire choice task, including the

instructions and response options, needs to be visible on the screen without scrolling required. At the same time, font size needs to be large enough for the choice task to be easily readable.

However, with a variety of modes of administration now available to complete surveys electronically, there may be differences in how the choice tasks appear to the respondent on different platforms. Researchers should be mindful of this and aim to present questions in formats that are easy to use and read on small touch screens [8, 33]. We recommend that researchers carefully consider the positive and negative aspects of making a preference survey mobile-device compatible [34, 35]. At a minimum, surveys should be tested on a variety of web browsers and mobile devices to ensure that choice tasks appear as intended. When a larger screen size is required to view the entire choice task, researchers can restrict respondents to (or request they use) tablets or computers when completing the survey.

In addition, researchers should think about how to access attribute descriptions and levels while respondents complete choice tasks. In paper-based surveys, respondents can simply be reminded that they can refer back to the full attribute descriptions as they complete the choice tasks. It can also be helpful to include one or more supplemental pages that include attribute definitions for quick reference. In electronic surveys, pop-up windows can be programmed to appear when a respondent hovers or clicks on an attribute label in the choice task to provide full attribute and level descriptions. Ideally, links to these pop-up windows are embedded within each choice task.

3.3.2 Choice Task Layout (Color-Coding and Use of Graphics)

If researchers anticipate that respondents need additional assistance in reading choice tasks and quickly flagging differing attribute levels, color-coding or shading techniques can be considered. Color-coding of choice tasks can help respondents quickly identify differences between choice profiles and might reduce dropout rate and improve attribute attendance [36, 37]. Different color-coding strategies include highlighting differences in attribute levels between profiles [38] or, when attribute levels are strictly ordinal, via intensity color-coding [39, 40]. Researchers should make sure that respondents with color vision deficiency can correctly interpret the choice tasks. Color schemes, both in graphics and when used for the color-coding of attributes, should be optimized for individuals with color vision deficiency; red and green should be avoided [55] and specialized color maps [41], shades of one color, should be used.

Adding graphics can also be helpful to explain the meaning of an attribute or to represent different attribute levels in choice tasks [23], but restraint is recommended. Adding

graphics for reasons other than describing an attribute (e.g., aesthetics) can be distracting and overwhelming. There is conflicting evidence related to the extent to which graphics add to the understanding of respondents regarding (risk) attributes [42, 43]. However, most empirical risk-communication studies recommend the use of icon arrays to improve understanding of risk information [44], but it is unclear the extent to which findings from the risk-communication literature apply to choice questions that require comparisons and tradeoffs across alternatives.

3.4 Other Information to Collect

3.4.1 Demographic and Medical Background Information

Many researchers initiate a survey by asking respondents to report select demographic and other background information. In many cases, these questions are used to determine whether an individual is eligible to participate (screening questions) and to start with an ‘easy’ set of questions. Whether these questions can be asked prior to or following the provision of informed consent, if required, will vary according to local and institutional policies. To minimize responder fatigue during the preference-elicitation portion of the survey, it can be advantageous to ask background questions not essential for determining inclusion and exclusion criteria until after the preference-elicitation questions have been completed.

It is generally recommended that researchers use standardized questions and response options to collect demographic information to facilitate comparisons across studies and to evaluate representativeness of the study sample. Standardized questions and response options will depend mainly on the study regions but may include census surveys collected by governments around the world or surveys conducted by international bodies such as the World Health Organization. Researchers should be mindful not to include too many of such questions before the preference elicitation section of the survey to avoid dropouts at that stage. In addition to collecting demographic information to describe the respondents, collecting information on respondent characteristics that might affect preferences is also important to enable the assessment of preference heterogeneity in the analysis phase of the study [45]. Questions related to attitudes and behavior that might impact preferences should only be positioned after the preference-elicitation tasks, along with any other questions or assessments that could impact responses to preference questions.

3.4.2 Internal Validity Checks

There are a number of validity tests that can be included to measure respondent comprehension and consistency. For

comprehensive reviews of tests for validity and reliability, please refer to Janssen et al. [46] and Johnson et al. [47]. When considering the inclusion of validity tests, researchers need to balance the number of tests versus respondent burden since such tests will add additional time to complete the survey. Researchers will also have to decide how to interpret the findings since, to date, there are no specific rules about what constitutes ‘high-quality’ or ‘low-quality’ respondents.

3.4.3 Numeracy and Literacy Checks

Since choice tasks often include interpreting complex health information and numerical information, preference researchers may choose to assess levels of health literacy and numeracy of respondents as these factors may affect their understanding of the tasks, which may contribute to preference heterogeneity [48]. Individuals with higher levels of numeracy understand health-related numerical information more accurately and make ‘better’ health decisions based on that information [49]. There are several validated instruments that can be used to assess objective or subjective numeracy [50–54]. Similarly, there are several instruments that can be used to assess health literacy [55, 56] and risk literacy [57]. If the survey includes graphs, it might also be useful to assess graph literacy, which predicts comprehension independently from other types of numeracy [58, 59].

3.4.4 Risk Attitudes, Behaviors, and Beliefs

Researchers may also consider including some questions about the respondents’ attitudes, risk behaviors, experiences, and beliefs [60]. These questions could be used in the analysis of results to explore subgroups based on hypothesized relationships or in latent-class analysis to explore heterogeneity and identify classes of respondents based on their preferences. These questions might be tailored to reflect specific decision-making contexts such as vaccination where it might be important to assess vaccine hesitancy.

3.4.5 Respondent Feedback

Researchers can also ask respondents directly about the survey with respect to difficulty, understanding and certainty. To ascertain this information, some researchers include direct questions in the survey or conduct post-survey debriefing and ‘think aloud’ studies [61, 62].

4 Testing the Survey

Once the attributes and levels (or objects) are finalized, the choice tasks of the stated-preference survey are generated using an experimental design that can be generated using

a variety of catalogues and software [5, 63]. Although an experimental design will be statistically efficient, consideration also needs to be given to response efficiency. Response efficiency can be assessed through pretesting and pilot testing to assess the number of choices that are feasible for respondents to complete and check attention levels.

Pretesting and pilot testing are important steps in the design and refinement of both the DCE choice tasks and the overall survey. Pretesting and pilot testing can serve to identify potential issues with the survey before and after it is programmed for administration. If issues are identified during the testing process, the survey should be updated and the testing process repeated in an iterative process (Fig. 3).

4.1 Pretesting

Pretesting generally includes qualitative interview methods to test respondent comprehension and areas of potential misunderstanding [17, 64] of the survey. In general, the goal of pretesting is to assess if the respondent understands the question in the way the researcher intended. Pretesting can also be helpful to reduce response burden, narrow down a longer list of potential attributes to be included in the choice tasks, or assess if any questions are inappropriate or too sensitive in nature [17]. Pretesting can also be helpful to assess the range of levels for an attribute in the stated-preference survey [65, 66].

There are no standardized methods for pretesting stated-preference surveys in health, but Ruel et al. provide a pretesting checklist for surveys in general [67]. Two common approaches used in pretesting are think-aloud interviews [64] and cognitive debrief interviews [67]. Using this approach, the researcher can collect feedback from a respondent as they answer each question and use probing methods to ask specific questions about how the respondent answered a question.

Ideally, pretesting should continue until the point of saturation when researchers are no longer identifying any new issues for revision to the survey. A general rule of thumb for surveys (not specific to stated-preference surveys) is between 12 and 50 respondents [67]. It is suggested that the researcher carefully documents changes made to the survey in the process of pretesting and the rationale for why the changes were made.

4.2 Pilot Testing

The final survey may be pilot tested with between 30 and 100 respondents [67]. Pilot tests can serve as a trial run of survey logistics [67] and be used to assess responses to the choice tasks [67]. The pilot data can be used to explore the direction and size of attribute coefficients [17], assess the consistency and rationality of responses in the survey [17],

evaluate whether respondents are making tradeoffs amongst attribute levels [17], examine data collection and coding procedures [67], and, when applicable, generate priors for parameter estimates for use in the experimental design of the main stated-preference survey, which can be critical when the potential sample is small [57].

5 Concluding Remarks

For the health preference research field to develop more systematic evidence on the impact of survey-design choices, researchers must be transparent and thorough in reporting on survey development and testing. Researchers should always (1) report how attributes and levels were communicated to respondents, (2) report on the educational materials assisting respondents with completing the choice tasks, (3) show an example choice task, and (4) make the survey instrument available, either on request or, preferably, in supplemental online materials. Systematic reporting will allow others to learn from positive and failed attempts at improving respondents' understanding, reducing drop-out, and minimizing measurement error associated with preference estimates.

Since most preference surveys are currently administered online, additional studies should investigate when and how best to use technological advances. For now, it is unclear what electronic features might assist in eliciting high-quality stated-preference data that reflect thoughtful responses from well-informed and engaged respondents to inform a given research question. Further insights are warranted for, amongst others, developing preference survey content for small screens (i.e., mobile phones). Related to this, further studies need to provide clarity on how best to ensure respondents understand the provided information on attributes, levels, and choice tasks as well as how to make sure respondents engage with the information and completion of the choices. In that respect, measures to assess if respondents understood presented information and/or were engaged, should be developed. Online features might be instrumental to advance this line of research to more effectively minimize respondent dropout and burden. Furthermore, studies examining the influence of various approaches to minimizing hypothetical bias in health-related stated-preference research are needed. Future research should also focus on best approaches for presenting complex attributes such as benefits and risks as well as choice-task layouts (e.g., the use of graphics). Future study registries of preference studies in health and their corresponding survey instruments could be particularly valuable in assisting the field in continuous learning and advancement in survey design, development, and comparability.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40271-023-00671-6>.

Declarations

Funding No financial support was provided to support this manuscript.

Conflict of interest Dr. Janssen is an employee of Johnson & Johnson and holds stock in Johnson & Johnson. Dr. Reed maintains a detailed listing of research funding and financial disclosures at: <https://scholars.duke.edu/person/shelby.reed>. Dr. Marshall reports personal fees from Analytica and Novartis and expenses from Illumina. Dr. Veldwijk has nothing to declare.

Author contributions All authors have been fully involved in the study design, collection, analysis, and interpretation of data, and writing of this article. All authors have approved the manuscript and agreed with its submission.

Data availability statement Not applicable.

References

- U.S. Department of Health and Human Services Food and Drug Administration, Center for Devices and Radiological Health and Center for Biologics Evaluation and Research. Patient Preference Information—Voluntary Submission, Review in Premarket Approval Applications, Humanitarian Device Exemption Applications, and De Novo Requests, and Inclusion in Decision Summaries and Device Labeling: Guidance for Industry, Food and Drug Administration Staff, and other Stakeholders. <http://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm446680.pdf> Accessed 24 Aug 2016.
- Medical Device Innovation Consortium. A framework for incorporating information on patient preferences regarding benefit and risk into regulatory assessments of new medical technology. 2015. http://mdic.org/wp-content/uploads/2015/05/MDIC_PCBR_Framework_Web1.pdf.
- Bridges JFP, de Bekker-Grob EW, Hauber AB, Heidenreich S, Janssen E, Bast A, Hanmer J, Danyliv A, Low E, Bouvy JC, Marshall DA. A roadmap for increasing the usefulness and impact of patient-preference studies in decision making in health: a good practices report of an ISPOR task force. *Value Health*. 2023;26(2):153–62.
- Hauber AB, Marcos Gonzales J, Groothuis-Oudshoorn CGM, Prior T, Marshall DA, Cunningham C, IJzerman MJ, Bridges JFP. 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):300–15.
- Johnson FR, Lancsar E, Marshall DA, Kilambi V, Mühlbacher A, Regier DA, Bresnahan BW, Kanninen B, Bridges JFP. Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task force. *Value Health*. 2013;16:3–13.
- Bridges JFP, Hauber AB, Marshall DA, Lloyd A, Prosser LA, Regier DA, Johnson FR, Mauskopf J. Conjoint analysis applications in health—a checklist: a report of the ISPOR good research practices for Conjoint Analysis Task Force. *Value Health*. 2011;14(4):403–13.
- Determann D, Lambooi MS, Steyerberg EW, de Bekker-Grob EW, de Wit GA. Impact of survey administration mode on the results of a health-related discrete choice experiment: online and paper comparison. *Value Health*. 2017;20(7):953–60.
- Antoun C, Couper MP, Conrad FG. Effects of mobile versus PC web on survey response quality: a crossover experiment in a probability web panel. *Public Opin Q*. 2017;81(S1):280–306.
- Skeje AM, Lindhjem H, Skjelfo S, Navrud S. Smartphone and tablet effects in contingent valuation web surveys—no reason to worry? *Ecol Econ*. 2019;165: 106930.
- Bruijne M De, Oudejans M. Online surveys and the burden of mobile responding. In: *Survey Measurement: Techniques, Data Quality and Sources of Error* Engel (Ed). Campus Verlag Frankfurt; 2015.
- Kahneman D, Tversky A. *Choices, values and frames*. New York: Cambridge University Press; 2000.
- Gigerenzer G, Todd P. *ABC Research Group*. New York: Simple heuristics make us smart. Evolution and Cognition. Oxford University Press; 1999.
- Veldwijk J, Marceta SM, Swait JD, Lipman SA, de Bekker-Grob EW. Taking the shortcut: simplifying heuristics in discrete choice experiments. *Patient*. 2023;16(4):301–15.
- Rakotonarivo OS, Schaafsma M, Hockley N. A systematic review of the reliability and validity of discrete choice experiments in valuing non-market environmental goods. *J Environ Manage*. 2016;183:98–109.
- Hanley N, Mourato S, Wright RE. Choice modelling approaches: a superior alternative for environmental valuation? *J Econ Surv*. 2001;15:435–62.
- Loomis JB. Strategies for overcoming hypothetical bias in stated preference surveys. *J Agric Resour Econ*. 2014;39:34–46.
- Hensher D, Rose JM, Greene WH. *Applied choice analysis: second edition*. Cambridge: Cambridge University Press; 2015.
- Johnston RJ, Boyle KJ, Loureiro ML, Navrud S, Rolfe J. Guidance to enhance the validity and credibility of environmental benefit transfers. *Environ Resour Econ*. 2021;79:575–624.
- Veldwijk J, Essers BAB, Lambooi MS, Dirksen CD, Smit HA, de Wit GA. Survival or mortality: does risk attribute framing influence decision-making behavior in a discrete choice experiment? *Value Health*. 2016;19(2):202–9.
- Slovic P. Perception of risk. *Science*. 1987;236:280–5.
- Bonner C, Trevena LJ, Gaissmaier W, et al. Current best practice for presenting probabilities in patient decision aids: fundamental principles. *Med Decis Mak*. 2021;41(7):821–33.
- Lipkus IM. Numeric, verbal, and visual formats of conveying health risks: suggested best practices and future recommendations. *Med Decis Mak*. 2007;27(5):696–713.
- Woloshin S, Schwartz LM. Communicating data about the benefits and harms of treatment: a randomized trial. *Ann Intern Med*. 2011;155(2):87–96.
- Bateman JJ, Day BH, Jones AP, Jude S. Reducing gain-loss asymmetry: a virtual reality choice experiment valuing land use change. *J Environ Econ Manag*. 2009;58:106–18.
- Hoffmann S, Winter J, Caro FG, Gottlieb AS. Effects of video enhancement in a stated-choice experiment on medical decision making. *Gerontology Institute Publications*; 2014. Paper 107.
- Charvin M, Launoy G, Berchi C. The effect of information on prostate cancer screening decision process: a discrete choice experiment. *BMC Health Serv Res*. 2020;20(1):467. <https://doi.org/10.1186/s12913-020-05327-x>.
- Vass CM, Davison NJ, Vander Stichele G, Payne K. A picture is worth a thousand words: the role of survey training materials in stated-preference studies. *Patient*. 2020;13(2):163–73. <https://doi.org/10.1007/s40271-019-00391-w>. PMID:31565784;PMCID:PMC7075825.
- Lim SL, Yang JC, Ehrisman J, Havrilesky LJ, Reed SD. Are videos or text better for describing attributes in stated-preference surveys? *Patient*. 2020;13(4):401–8.

29. Smith IP, Whichello CL, de Bekker-Grob EW, Mølken MPMHR, Veldwijk J, de Wit GA. The impact of video-based educational materials with voiceovers on preferences for glucose monitoring technology in patients with diabetes: a randomised study. *Patient*. 2023;16(3):223–237. <https://doi.org/10.1007/s40271-022-00612-9>.
30. Westera W, Nadolski R, Hummel H, et al. Serious games for higher education: a framework for reducing design complexity. *J Comput Assist Learn*. 2008;24(5):420–32.
31. McFadden D. Conditional logit analysis of qualitative choice behaviour. In: Zarembka P (ed) *Frontiers in econometrics*. Academic Press: New York; 1974.
32. Dillman DA, Gertseva A, Mahon-Haf T. Achieving usability in establishment surveys through the application of visual design principles. *J Off Stat*. 2005;21:183–214.
33. Keusch F, Yan T. Web versus mobile web: an experimental study of device effects and self-selection effects. *Soc Sci Comput Rev*. 2017;35(6):751–69.
34. Hartman JD, Craig BM. Does device or connection type affect health preferences in online surveys? *Patient*. 2019;12(6):639–50.
35. Vass CM, Boeri M. Mobilising the next generation of stated-preference studies: the association of access device with choice behaviour and data quality. *Patient*. 2021;14(1):55–63.
36. Jonker MF, Donkers B, de Bekker-Grob EW, Stolk EA. Effect of level overlap and color coding on attribute non-attendance in discrete choice experiments. *Value Health*. 2018;21(7):767–71.
37. Jonker MF, Donkers B, de Bekker-Grob E, Stolk EA. Attribute level overlap (and color coding) can reduce task complexity, improve choice consistency, and decrease the dropout rate in discrete choice experiments. *Health Econ*. 2019;28(3):350–63.
38. Norman R, Viney R, Aaronson NK, Brazier JE, Cella D, Costa DS, Fayers PM, Kemmler G, Peacock S, Pickard AS, Rowen D, Street DJ, Velikova G, Young TA, King MT. Using a discrete choice experiment to value the QLU-C10D: feasibility and sensitivity to presentation format. *Qual Life Res*. 2016;25(3):637–49.
39. Jonker MF, Attema AE, Donkers B, Stolk EA, Versteegh MM. Are health state valuations from the general public biased? A test of health state reference dependency using self-assessed health and an efficient discrete choice experiment. *Health Econ*. 2017;26(12):1534–47. <https://doi.org/10.1002/hec.3445>.
40. Jonker MF, Donkers B, De Bekker-Grob EW, Stolk EA. Advocating a paradigm shift in health-state valuations: The estimation of time-preference corrected QALY tariffs. *Value Health*. 2018;21(8):993–1001.
41. Nuñez JR, Anderton CR, Renslow RS. Optimizing colormaps with consideration for color vision deficiency to enable accurate interpretation of scientific data. *PLoS ONE*. 2018;13(7): e0199239.
42. Vass C, Rigby D, Campbell S, Tate K, Stewart A, Payne K. Investigating the framing of risk attributes in a discrete choice experiment: an application of eye-tracking and think aloud. *Med Decis Mak*. 2015;35(1):E99.
43. Veldwijk J, Lambooy MS, van Til JA, Groothuis-Oudshoorn CG, Smit HA, de Wit GA. Words or graphics to present a discrete choice experiment: does it matter? *Patient Educ Couns*. 2015;98(11):1376–84.
44. Richter R, Jansen J, Bongaerts I, Damman O, Rademakers J, van der Weijden T. Communication of benefits and harms in shared decision making with patients with limited health literacy: a systematic review of risk communication strategies. *Patient Educ Couns*. 2023;17(116): 107944.
45. Vass C, Boeri M, Karim S, Marshall DA, Craig B, Ho KA, Mott D, Ngorsuraches S, Badawy SM, Muhlbacher A, Gonzalez MJ, Heidenreich S. Accounting for preference heterogeneity in discrete-choice experiments: an ISPOR special interest group report. *Value Health*. 2022;25(5):685–94.
46. Janssen EM, Marshall DA, Hauber AB, Bridges JFP. Improving the quality of discrete-choice experiments in health: how can we assess validity and reliability? *Expert Rev Pharmacoecon Outcomes Res*. 2017;17(6):531–42.
47. Johnson FR, Yang J-C, Reed SD. The internal validity of discrete choice experiment data: a testing tool for quantitative assessments. *Value Health*. 2019;22(2):157–60.
48. Veldwijk J, van der Heide I, Rademakers J, Schuit AJ, de Wit GA, Uiters E, Lambooy MS. Preferences for vaccination: does health literacy make a difference? *Med Decis Mak*. 2015;35(8):948–58.
49. Reyna VF, Brainerd CJ. The importance of mathematics in health and human judgment: numeracy, risk communication, and medical decision making. *Learn Individ Differ*. 2007;17(2):147–59.
50. Dykhuis MAKE, Slowik PDL, Bryce PDK, Hyde-Nolan PDME, Eshelman PDA, Miller-Matero PDLR. A new measure of health numeracy: brief medical numbers test (BMNT). *Psychosomatics*. 2018.
51. Schapira MM, Walker CM, Cappaert KJ, Ganschow PS, Fletcher KE, McGinley EL, Del Pozo S, Schauer C, Tarima S, Jacobs EA. The numeracy understanding in medicine instrument: a measure of health numeracy developed using item response theory. *Med Decis Mak*. 2012;32(6):851–65.
52. Schapira MM, Walker CM, Miller T, Fletcher KE, Ganschow PS, Jacobs EA, Imbert D, O'Connell M, Neuner JM. Development and validation of the numeracy understanding in Medicine Instrument short form. *J Health Commun*. 2014;19(Suppl 2):240–53.
53. Fagerlin A, Zikmund-Fisher BJ, Ubel PA, Jankovic A, Derry HA, Smith DM. Measuring numeracy without a math test: development of the Subjective Numeracy Scale (SNS). *Med Decis Making*. 2007;27:672–80.
54. McNaughton CD, Cavanaugh KL, Kripalani S, Rothman RL, Wallston KA. Validation of a Short, 3-Item Version of the Subjective Numeracy Scale. *Med Decis Mak*. 2015;35(8):932–6.
55. Health Literacy Tool Shed. Health literacy tool shed: a database of health literacy measures. 2018. <http://healthliteracy.bu.edu/>. Accessed 21 July 2023.
56. Chew LD, Griffin JM, Partin MR, et al. Validation of screening questions for limited health literacy in a large VA outpatient population. *J Gen Intern Med*. 2008;23(5):561–6.
57. Cokely ET, Galesic M, Schulz E, Ghazal S, Garcia-Retamero R. Measuring risk literacy: the Berlin Numeracy Test. *Judgm Decis Mak*. 2012;7(1):25–47.
58. Galesic M, Garcia-Retamero R. Graph literacy: a cross-cultural comparison. *Med Decis Mak*. 2011;31:444–57.
59. Okan Y, Janssen E, Galesic M, Waters EA. Using the short graph literacy scale to predict precursors of health behavior change. *Med Decis Mak*. 2019;39(3):183–95.
60. Russo S, Jongerius C, Faccio F, Pizzoli SFM, Pinto CA, Veldwijk J, Janssens R, Simons G, Falahee M, de Bekker-Grob E, Huys I, Postmus D, Kihlbom U, Pravettoni G. Understanding patients' preferences: a systematic review of psychological instruments used in patients' preference and decision studies. *Value Health*. 2019;22(4):491–501.
61. Whitty JA, Walker R, Golenko X, Ratcliffe J. A think aloud study comparing the validity and acceptability of discrete choice and best worst scaling methods. *PLoS ONE*. 2014;9(4): e90635.
62. Bridges J, Oakes A, Renhart C, Byard E, O'Donoghue B. Developing and piloting an instrument to prioritize the worries of patients with acute myeloid leukemia. *Patient Prefer Adherence*. 2018;12:647–55.
63. Flynn TN, Marley AAJ, Louviere JJ. Best-worst scaling: theory, methods and applications. Cambridge: Cambridge University Press; 2015.
64. Collins D. Pretesting survey instruments: an overview of cognitive methods. *Qual Life Res*. 2003;12(3):229–38.

65. Mitchell RC, Carson RT. Using surveys to value public goods—the contingent valuation method. New York: Resources for the Future; 1993.
66. Klose T. The contingent valuation method in health care. *Health Policy*. 1999;47(2):97–123.
67. Ruel E, Wagner WI, Gillespie B. Pretesting and pilot testing. In: Ruel E, editor. *The practice of survey research: theory and applications*. SAGE Publications, Inc.; 2015.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.