



# How to Present a Decision Object in Health Preference Research: Attributes and Levels, the Decision Model, and the Descriptive Framework

Axel C. Mühlbacher<sup>1</sup> · Esther W. de Bekker-Grob<sup>2</sup> · Oliver Rivero-Arias<sup>3</sup> · Bennett Levitan<sup>4</sup> · Caroline Vass<sup>5</sup>

Accepted: 16 January 2024  
© The Author(s) 2024

## Abstract

In health preference research (HPR) studies, data are generated by participants’/subjects’ decisions. When developing an HPR study, it is therefore important to have a clear understanding of the components of a decision and how those components stimulate participant behavior. To obtain valid and reliable results, study designers must sufficiently describe the decision model and its components. HPR studies require a detailed examination of the decision criteria, detailed documentation of the descriptive framework, and specification of hypotheses. The objects that stimulate subjects’ decisions in HPR studies are defined by attributes and attribute levels. Any limitations in the identification and presentation of attributes and levels can negatively affect preference elicitation, the quality of the HPR data, and study results. This practical guide shows how to link the HPR question to an underlying decision model. It covers how to (1) construct a descriptive framework that presents relevant characteristics of a decision object and (2) specify the research hypotheses. The paper outlines steps and available methods to achieve all this, including the methods’ advantages and limitations.

## 1 Introduction

Health preference research (HPR) studies observe behavior to indirectly quantify preferences that determine and predict the value of healthcare goods (e.g., treatments, devices) and/or services (e.g., screening, end-of-life care). These goods or services are the “object” about which preferences are being elicited. Precisely specifying the object of research is critical, as it provides the necessary basis to quantify preferences and values and to better understand the factors that influence

people’s decision-making (subjects’ decision-making). By defining and describing specific objects<sup>1</sup>, alternatives, or items, the research team can ensure that study participants are evaluating and comparing the same things. To distinguish one object from another, objects are characterized by attributes and levels [1]. The identification and presentation of attributes and attribute levels is therefore one of the most important tasks when designing an HPR study. Because of their importance, it is critical to understand the extent to

<sup>1</sup> What are objects, alternatives, and items? *Objects* are the health conditions or treatments that participants are evaluating. Objects can be different health conditions, symptoms, or treatment effects. The use of the term “object” does not imply a preference elicitation method. Objects may not be complete substitutes in a preference elicitation task. *Alternative*, on the other hand, specifies the comparison of multiple options. The term “alternative” implies that each option being evaluated is a complete substitute for the other(s); participants are asked to make direct trade-offs. Therefore, alternatives capture the trade-offs that respondents make when comparing and evaluating different options. Alternatives are intended to be mutually exclusive and collectively exhaustive (i.e., all possible options should be included in the set of options). *Items* typically refer to the specific attributes or dimensions of health states that are being compared and evaluated. Items can be used to describe the characteristics of a health state.

Extended author information available on the last page of the article

### Key Points for Decision Makers

A decision model serves as the theoretical basis for HPR studies, describing the decision context, objects, and subjects (the individuals in the research sample) involved. It is essential to communicate the decision model clearly through narrative and mathematical specifications to ensure replication and understanding by other research teams.

Defining and describing specific objects is crucial in health preference research (HPR) to ensure consistent evaluation and comparison by participants. Precisely specifying attributes and attribute levels provides valid, reliable, and objective information within the experiment.

The identification of attributes and levels can be approached through explorative methods (e.g., literature research, interviews, focus groups) or elicitation methods (e.g., surveys, factor analysis). The choice of approach depends on the research question and available evidence. Triangulation of methods can enhance the validation of attribute selection. Setting attribute levels involves considering the scope, range, and impact of each attribute.

Building a descriptive framework involves presenting attributes and levels in a manner that aids respondents' understanding. Graphics and qualitative descriptions can supplement numerical information. Careful framing of attributes and levels can influence preferences, so researchers should consider framing effects and employ cheap talk scripts to mitigate biases.

which attributes and attribute levels provide valid, reliable, and objective information within the HPR study.

Preference data capture observable behavior in either a *stated* or a *revealed study format*. This paper focuses on stated-preference research. In stated-preference research, behavior is stimulated by a preference elicitation task, and the preferences are determined by what a participant states they would do (i.e., their stated choice). HPR elicits preferences from respondents by having them evaluate one or more objects from a set of alternatives. To elicit meaningful stated preferences, researchers must make sure that the information about objects is clearly available to and understood by respondents. The research team must decide which attributes and levels can be used to define objects to stimulate responses.

This paper provides guidance on systematically deconstructing a decision model into its component attributes and attribute levels to develop the descriptive framework (i.e.,

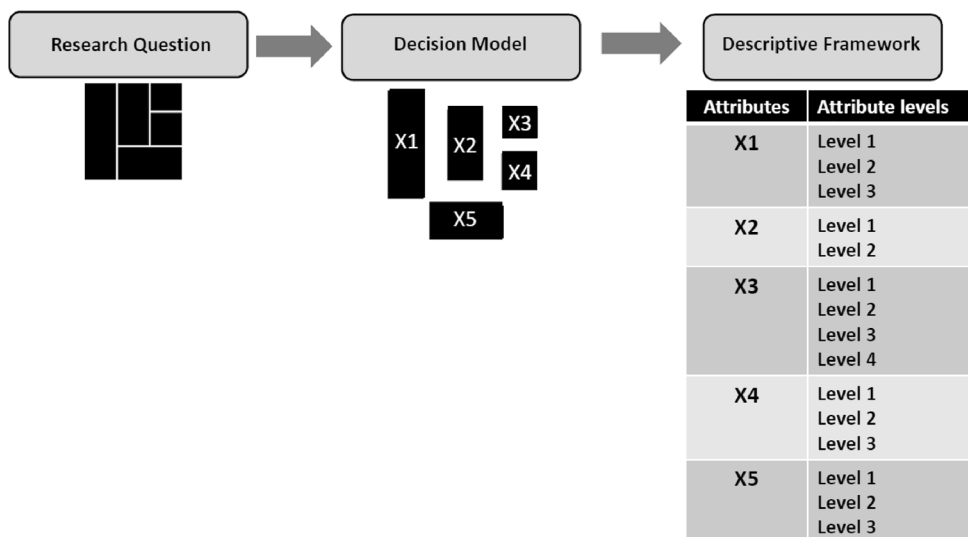
the supporting explanations of the attributes and levels) and specify the hypotheses and aims of an HPR study. The following sections define the decision model and review various methods for (1) identifying attributes and attribute levels, (2) describing decision models, (3) building descriptive frameworks, and (4) specifying hypotheses.

## 2 Decision Model

The *research question* and *decision model* are the starting point for the specification of objects (see Fig. 1). The decision model is the theoretical basis and can be described by a formal (axiomatic) system. It describes the decision context, objects, and subjects (the individuals in the research sample). *Decision context* refers to the specific situation or scenario in which the decision is being made. The decision context includes individual, social, and political perspectives, as well as any contextual factors that may influence the decision [2]. The decision-making process for individuals with late-stage lung cancer involves a complex interplay of medical facts, preferences, social dynamics, economic considerations, and the broader healthcare environment (see Box 1 in the electronic supplementary material). Understanding this decision context is crucial for healthcare providers to offer appropriate guidance and for the design of healthcare policies and support systems that align with the needs and preferences of this patient group. Objects are defined by *attributes* and their associated *attribute levels*. Attributes refer to the specific characteristics or features of a health state or treatment (e.g., if the object is a treatment, the benefits, risk of side effects, mode of administration, and dosing frequency). Level refers to the different values of each attribute. Attribute levels, when presented in a set of alternatives, are used to differentiate objects (e.g., for a mode of administration attribute, the attribute levels could be oral, injection, or infusion). Specifications of levels provide utility to the respondent and therefore stimulate behavior. Furthermore, the context of the decision and the roles of the subjects<sup>2</sup> also impact respondent behavior. Experiments and surveys are designed to observe how respondents behave

<sup>2</sup> Who are subjects, participants, respondents, and patients? Subjects, participants, respondents, and patients are individuals selected by the research team to participate in an HPR study. In HPR, preference information or data are collected from or about a person. Research subjects are referred to alternatively as research participants, respondents, and study volunteers. A subject may be a patient or a healthy individual. The term “*subject*” describes an individual who is part of the selected sample for experimental/scientific research (i.e., the subject of a quantitative study). *Participants* are individuals who voluntarily participate in an HPR study (often in the context of qualitative research [e.g., interviews]). *Respondents* are individuals who answer questionnaires. *Patients* are individuals with a medical condi-

**Fig. 1** Construct the decision model and its descriptive framework



according to a specific format. The *descriptive framework* is a visual and verbal representation of attributes and levels; it documents how attributes and levels are presented to respondents in a preference elicitation task. The research team needs a common understanding of the decision problem and model. To replicate research results, the design of the objects must be transparent to other research teams. The decision model can be communicated through a narrative version and a mathematical specification. The mathematical specification formally presents the hypotheses.

### 3 Identification of Attributes and Levels

When beginning to develop a health preference study, the research team must identify and validate which attributes are important to subjects. The research team can use *explorative* and *elicitation methods* to identify attributes and associated levels. Although we present these two methods separately in this section, researchers may find it useful to employ a comprehensive *mixed-methods approach*, which combines both explorative and elicitation methods with evidence from literature reviews. Researchers may also wish to validate

findings by using different explorative and/or elicitation methods in a process known as *triangulation*.

Ultimately, the most appropriate way of identifying attributes is likely to be driven by the research question at hand. For example, a *top-down approach*, in which the attributes are initially provided by the study design team and then fleshed out, may be required if the preference study focuses on a highly refined research question (in order, for example, to look at trade-offs for specific benefits and risks). A *bottom-up approach*, in which the attributes emerge from interviews with stakeholders, may be more appropriate in cases where there is little evidence about what drives preferences, such as with a rare disease or a new class of treatment.

Regardless of the approach by which they are developed, once the final attributes have been selected, *qualitative pre-testing* of the attributes must be conducted to test for sense, language, comprehension, and layout, as well as for causal relationships between attributes.

#### 3.1 Explorative Approaches

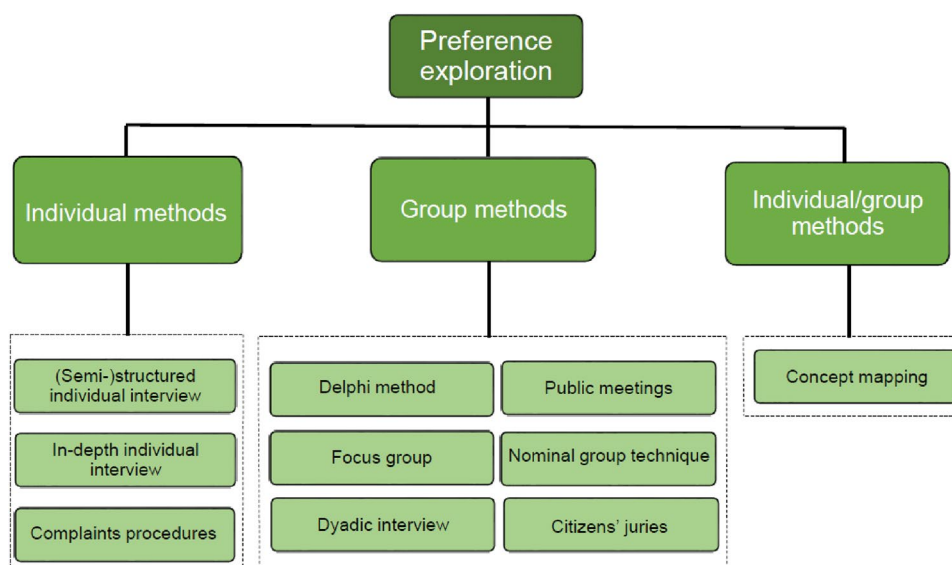
There is no strict definition for explorative approaches, as they can encompass both purely *qualitative investigations* as well as *hybrid approaches* such as *consensus methods*. In this paper, we define explorative methods loosely as *qualitative methods* that collect descriptive data through participant or phenomenon observation and that examine participants' experiences and decisions.

The term "qualitative research" generally refers to the study—using an *interpretivist paradigm*—of nonnumerical data in the absence of a clear, testable hypothesis [3]. In the case of attribute identification, explorative methods are likely to be used to investigate and detect attributes of importance in the absence of existing evidence. Such methods

Footnote 2 (continued)

tion. *Subject* and *respondent* may indicate that the individual does not understand the meaning of the information they are providing. Participant may indicate that the individual is aware of how and why he or she is providing information in a research study. The term "subject" or "human subject" is often used in regulatory guidelines, particularly in clinical research. The term "participant" is often used to replace the term "subject," as the term "subject" is sometimes considered demeaning.

**Fig. 2** Preference exploration field research methods. Source: Soekhai V, Whichello C, Levitan B, Veldwijk J, Pinto CA, Donkers B, et al. Methods for exploring and eliciting patient preferences in the medical product lifecycle: a literature review. *Drug Discov Today*. 2019 Jul;24(7):1324–1331. <https://doi.org/10.1016/j.drudis.2019.05.001>



may also be used to understand how to best define attributes in a manner that is clear and consistently interpreted by respondents.

Explorative approaches to identifying attributes could include analyzing existing literature identified through *scoping review* or *systematic review* methods. It could also involve searching for *grey literature* to identify unpublished but relevant information from sources such as websites, reports, working documents, or patient leaflets. Formal qualitative analysis of literature (including existing qualitative studies) is called “meta-synthesis” and involves combining and then interpreting textual data from multiple studies or sources.

Explorative approaches to identifying attributes could also include *field research methods*. Examples include but are not limited to the following: *interviews* with individuals (e.g., *semi-structured*, *in-depth*); *focus groups* or *dyadic interviews*; consensus methods, such as the *Delphi method*, *nominal group technique*, or *citizens' juries*; or other approaches, such as *concept mapping interviews*, *complaints procedures*, or *public meetings* (see Fig. 2 [4]).

When collecting data via field research methods, designers of preference studies must recruit participants relevant to the research question. Field research is defined as a qualitative method of data collection that aims to observe, interact with, and understand people in a natural environment [5]. In most explorative approaches to collecting field data, participants (e.g., patients, caretakers, parents, the general population) are required. Researchers may use *purposeful sampling*, which is defined as a *nonprobability sampling technique* wherein participants are chosen because of their knowledge or expertise (e.g., selecting patients with experience of a particular condition to determine the attributes of importance in a treatment). Alternatively, researchers may

choose to use *convenience sampling*, in which participants are chosen on the basis of ease of access (e.g., using patient charities or local meetings to identify potential research participants). In some instances, *snowball sampling* is used [6]. Snowball sampling is defined as a nonprobability sampling technique wherein existing study participants recruit or recommend future participants [7].

Currently, common practice for identifying attributes in HPR is a combination of both literature research (i.e., systematic reviews, scoping reviews, meta-syntheses, or other literature review types) and explorative field methods (mainly focus groups and/or individual interviews). The results of literature research are used as discussion points and/or topic guides for the field research methods.

In contrast to quantitative research (see next section), explorative research is generally not driven by sample sizes that provide statistical significance. Instead, sampling will occur until the point of *data saturation* is reached (i.e., when a degree of consensus has been achieved, or when no new themes arise from the analysis of data). Therefore, the process of explorative research may involve concurrent or repeated sampling, collection, and analysis.

In qualitative studies, recordings (voice or video) are usually made to generate the data for analysis, although these recordings may be supplemented with the facilitator's notes. Textual data (or transcriptions of voice data) are usually reduced into manageable issues of importance by coding. In its simplest form, *thematic analysis* involves identifying and recording repeated issues of interest and assigning these a code that can be grouped into themes to form attributes for the preference study. Explorative approaches, including qualitative research, are scientific, and the handling of data should therefore be rigorous, critical, and objective [8]. At a minimum, this involves an unbiased approach to analyses



and reflexive practices, whereby the researcher considers their (or society's) influence on the subject (the individual whose preferences are being investigated). *Inter-rater reliability* checks, in which independent researchers compare their analyses, can also help. See checklists and guidelines for good qualitative research practices generally (e.g., [9–11]) and for stated-preference studies specifically [12].

Although qualitative analysis focuses largely on reducing data, it may still produce more themes than can be used as attributes in an HPR study. Possible solutions to this problem are to use further expert judgement to pare down themes, or to use a formal qualitative process of reduction and synthesis without losing or changing the meaning of themes. Another possibility is to use quantitative methods (see next section) to reduce the list of potential attributes. It may also be necessary to modify the research question in order to address the number of themes that can realistically be incorporated into a health preference study.

Boxes 2 and 3 (see the electronic supplementary material) present two case study examples to illustrate the identification of attributes by using literature research followed by one or more field research methods.

### 3.2 Elicitation Approaches (Quantifiable Data)

*Factor analysis* [6] is a statistical method that can be used to quantitatively identify the most relevant factors for use as attributes. In factor analysis, the items or outcomes associated with the greatest amount of variance (thereby suggesting that they are the most important or influential) can be identified and then grouped as a single variable (or attribute). In some cases, an HPR study may have no associated patient-reported outcome or health-related quality of life data, or the research question may aim to elicit preferences in a different context. In these cases, it is valuable to collect *primary data* to identify attributes of importance. In its simplest form, this can be done with *open-ended*, *ranking*, or *Likert (rating) survey questions*.

However, Likert scale responses are prone to bunching (such as ceiling effects, when participants say all things are important), and they require respondents to answer too many questions. Similarly, ranking exercises provide an order (on a *cardinal scale*) but no *relative importance*, and therefore a cut-off for including attributes must be determined.

More advanced question formats revealing strength of importance with fewer questions could be advantageous, such as *best–worst scaling* type 1 [13], *Q-methodology* [14], *point allocation* [4], *Analytical Hierarchy Process (AHP)* [15], and *repertory grid method* [16].

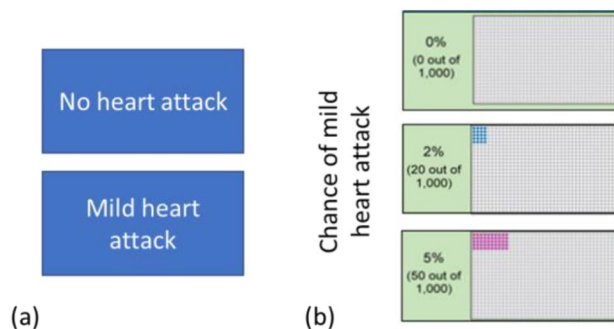
See Box 4 for a case study example (see the Electronic Supplementary Material).

### 3.3 Setting Attribute Levels

The research team must next define the scope, range, and impact of the attribute (e.g., what are the range and specific values for probabilities of a probabilistic attribute). Attributes can be defined with a fixed value or with multiple levels. Defining attribute value can be as critical to the success of an HPR survey as selecting attributes. Having too many value levels will increase the *task complexity* and *cognitive burden*, have implications for the experimental design, and possibly result in levels that are hard for respondents to distinguish between. Having too few levels may prevent the analysis from producing information about the shape of the (partial) utility function. Levels must also correspond to the range of data to which the preferences are being applied. Levels outside a meaningful range can make a survey more challenging for participants to complete and for researchers to conduct. However, the levels may also extend beyond current plausible ranges if the research question is concerned with preferences for future alternatives or if there is a need to find an unusually large maximum acceptable level of an attribute.

*Deterministic attributes* are generally easier for respondents to understand and result in less cognitive burden. An example of deterministic levels for the attribute “adverse event” could be “no heart attack” or “mild heart attack” (see Fig. 3).

*Probabilistic attributes* typically require lengthier training explanations in surveys, more testing of whether the respondents understood the training, and additional pictorial or graphical illustrations to accompany the statistic (Figure 3). For these reasons, when study designers have a choice, deterministic levels are generally a better option. However, most real-world medical decision choices are probabilistic, and probabilistic attributes better reflect the decision frame. One very common metric in HPR studies, the *maximum acceptable risk* for a harmful attribute in exchange for a given benefit, should be assessed using a



**Fig. 3** Examples of levels and attribute presentation for deterministic and probabilistic versions of an attribute. **a** Deterministic; **b** probabilistic

probabilistic attribute with several levels. A common compromise is to limit the use of probabilistic attributes just to those metrics needed to support the study's research objective. An example of probabilistic levels for the attribute "adverse event" could be "0% chance of heart attack" or "2% chance of heart attack."

*Nominal attributes* are those with different categories for levels but with no natural ordering for the levels, such as "stroke," "heart attack," and "type 2 diabetes." *Ordinal attributes* are those with different levels that have a natural order but do not have clear or equal intervals of difference between them, such as "asymptomatic," "non disabling," and "disabling" stroke. *Dichotomous attributes*—attributes with only two levels—are a special case of ordinal attributes. *Interval attributes* are those with measurable differences between values but no true zero, such as temperature. *Ratio attributes* have measurable differences and a true zero; examples include weight, cost, probabilities, and time periods.

Attribute levels refer to the specific values of attributes that are used in an HPR survey, such as distinct probabilities of a side effect, the costs of a service, or the different severities of a symptom. An attribute level is defined as one of two or more descriptors of a single attribute. When selecting levels for attributes, researchers should bear in mind the utility function that will eventually be modeled. The utility function gives the utility over the range of values for an attribute. The function may be linear, in which unit increases in the attribute result in a constant change in utility. Alternatively, utility may taper to an asymptote with increasing attribute value, or utility may have a sigmoid shape (Fig. 4). For many preference studies, a key research objective is being able to characterize the nonlinearity of the utility function.

Finally, the interval between levels is also a consideration in level design. Distinctions between levels may be important for numerous stakeholders, but these distinctions can be lost or ignored when presented in choice alternatives. For example, while the distinction between a 2% and a 4% chance of disabling stroke may be critical for clinical and regulatory decisions, patients may not consider the difference meaningful, and even if they do consider it meaningful, they may still ignore it when completing a choice task.

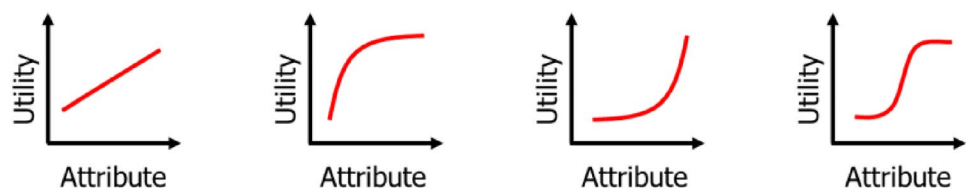
For these reasons, the range and values of levels should be refined in both preliminary scoping interviews with clinical

experts and in pretesting of survey instruments. No matter how complex a utility function is, an attribute with two levels will only provide a linear approximation of the utility function. Similarly, an attribute with two levels cannot assess when there is no change in preference below a threshold or above a threshold, as in sigmoid value functions. An attribute with two levels also cannot assess the impact of risk attitude, where uncertainty or a threshold in an attribute induces a responder to change their preferences. Multiple levels over the range of interest are needed to assess these complexities. However, too many levels will complicate the experimental design, leading either to large cognitive burden per participant (subject) or to the need for a very large sample size.

The range of the levels impacts not only the ability to characterize an attribute's utility function, but also whether the survey can assess trade-offs. Consider one deterministic benefit attribute (delaying onset of brain cancer metastasis, with levels ranging from 1 to 5 years) and one probabilistic harm attribute (chance of disabling stroke, with levels of 0% to 5%). If brain cancer patients would accept a 10% chance of disabling stroke for just a 1-year delay of metastasis, the benefits will appear to always outweigh the harm in the survey. In this case, the levels of the probabilistic harm attribute are irrelevant to patients. If the maximum probability for the harm attribute is too small, any change in the benefit will be more important than any change in harm, and the survey will show no preference sensitivity to changes in the harm. Clinical experts collaborating in the design of an HPR study may be uncomfortable with the inclusion of unrealistically large probabilities for harms, but the large probabilities may be necessary to observe trade-off behavior and thereby glean useful information from the survey. On the other hand, too high a probability of harm can appear so unrealistic that respondents stop taking the survey seriously. Fig. 4 makes clear that the location of the utility of the levels is also critical. If all the utility of the levels is below the lower curve of the sigmoid utility function, no trade-off behavior will be detected. If all the utility of the levels is in the middle of the sigmoid, the upper and lower asymptotes will not be detected.

Tables 1 and 2 present some examples of attributes and attribute levels, turning the abstract descriptive framework of Fig. 1 into concrete example attribute levels.

**Fig. 4** Linear, tapering, exponential, and sigmoid utility functions



**Table 1** Examples of continuous and categorical attributes for a health preference research study

Attributes	Description
Continuous	
Waiting time	The time that you have to wait
Out-of-pocket costs	The out-of-pocket costs you have to pay
Risk of side effects	The risk you will have of experiencing side effects
Risk of poor outcomes	The risk you will have of receiving poor outcomes
Treatment duration	The time the treatment will take you
Cost to the healthcare system	The cost to the healthcare system
Categorical	
Location of hospital	Where the hospital is located
Mode of medicine administration	The path by which a medicine is taken into the body
Type of healthcare professional	The healthcare professional that will treat you
Reputation of healthcare provider	The reputation that the healthcare provider has

**Table 2** Examples of attributes and attribute levels for a health preference research study

Attributes	Attribute levels
Continuous	
Waiting time	(Interval or ratio)
	1 week
	2 weeks
Out-of-pocket costs	3 weeks
	0 Euros
	100 Euros
Risk of side effects	200 Euros
	400 Euros
	1 out of 1000
Risk of poor outcomes	10 out of 1000
	100 out of 1000
	5%
Treatment duration	10%
	1 month
	6 months
Cost to the healthcare system	12 months
	10,000 Euros
	50,000 Euros
Categorical	100,000 Euros
	(Nominal or ordinal)
	Location of hospital
Mode of medicine administration	City center
	Rural areas
Type of healthcare professional	Oral route
	Rectal route
	Injection
Reputation of healthcare provider	Surgeon
	Radiotherapist
	Dermatologist
Degree of side effects	Good
	Neutral
	Bad
	None
	Mild
	Moderate
	Severe

## 4 Building a Descriptive Framework

At this stage, the research team must decide on the final set of attributes and how to present attributes and their associated levels. In HPR, respondents confronted with a preference elicitation task are encouraged to express their preferences for a certain object. To differentiate objects in a set of alternatives, objects are characterized by attribute levels. Behavior is stimulated by the perceived utility of the objects. There are various ways of presenting the information contained in attributes and levels. The definitions and information presented affect respondents' interpretation. The language used to describe attributes can be *misinterpreted*; how those attributes are presented (e.g., numbers, coloring, graphics) can also lead to misinterpretation. Furthermore, the framing (positive or negative) can also act as a cue to study participants and *manipulate* the attribute and level meaning [17, 18].

### 4.1 Graphical Illustrations to Inform Decisions

It is common to use graphics or text in addition to numerical information to aid respondents' understanding of levels in HPR studies [19, 20]. In addition to (or instead of) numbers, it may be useful to allow graphics to illustrate the magnitude of a risk and assist with comparisons across levels. One popular way of explaining probability is the *icon array* (Figure 3b). However, the usefulness of icon arrays becomes limited at low probabilities (< 0.1%). The literature generally supports using *blocked icons* (grouped together) rather than a scattered approach, as *scattering icons* appears to indicate randomness and uncertainty rather than risk [21].

Figure 5 illustrates different formats that can be used to communicate levels of risk attributes. For example, the levels may be presented as percentages (e.g., 10%), ratios (e.g., 0.1), or natural frequencies (e.g., 1/10). They may also be

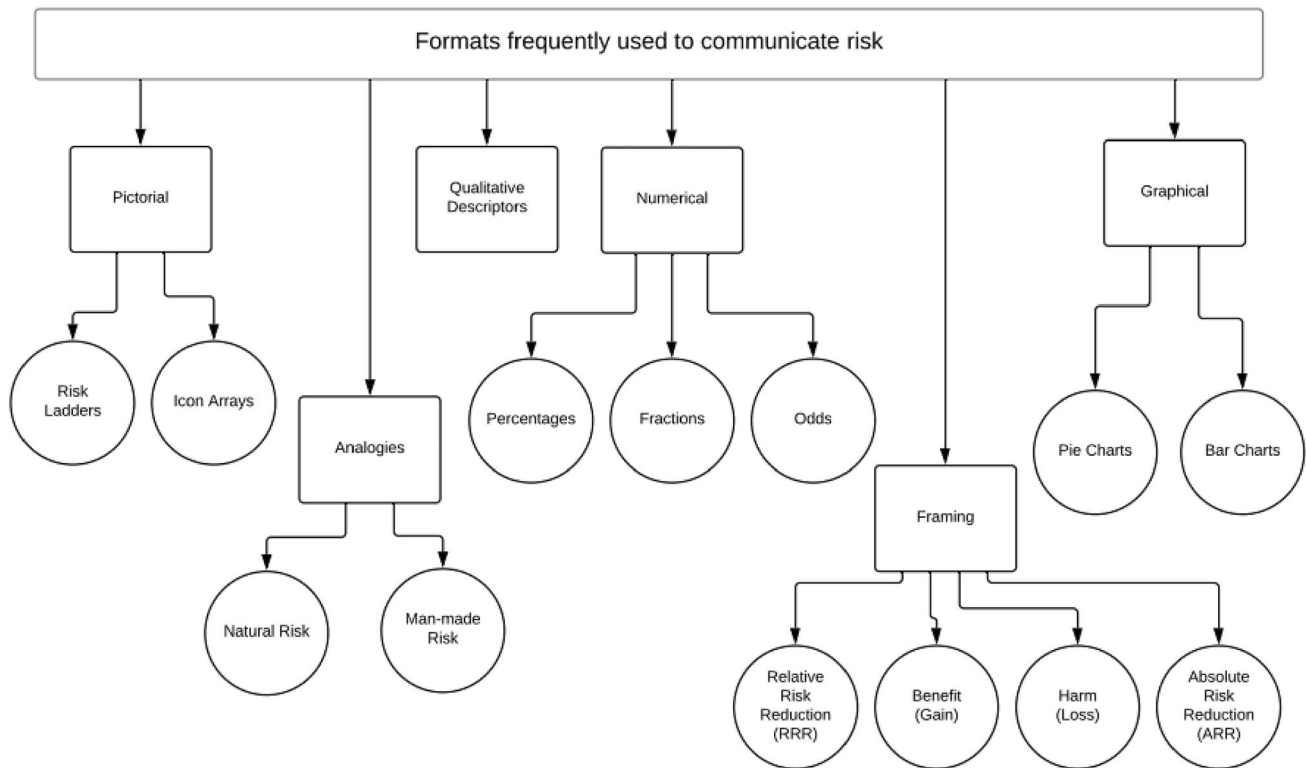


Fig. 5 Frequently used risk communication formats

presented as text with frequencies (e.g., “one out of ten”) or as a qualitative descriptor (e.g., “low probability”). Percentages are a traditional risk communication format that describe the proportion per hundred. However, for small risks, percentages can be misconstrued when decimal places are used. Natural frequencies and fractions are also susceptible to criticism (such as *numerator* and *denominator biases* that result from an inherent tendency to focus on one side of a ratio).

Adding graphics for reasons other than explaining the content of an attribute level is not advised [22]. Too much information can cause an overload for study participants [23]. Similarly, too many graphics, or graphics that are too complex, might distract study participants [24].

## 4.2 Qualitative Presentations to Inform Decisions

In some instances, researchers may find it useful to use qualitative statements rather than numerical formats to present continuous attributes, in order to relieve the cognitive burden on respondents. For example, researchers could describe pain in terms of severity (“no pain,” “mild pain,” “severe pain”) rather than using a pain scale, or describe costs as “high,” “medium,” or “low” instead of assigning values.

While such terminology can be vague, it is used regularly in some healthcare research, particularly research dealing with subjective experiences such as pain or discomfort. To describe likely benefit from chemotherapy, Marshall et al. [25] translated gene expression profiling scores into qualitative statements ranging from “low benefit” (score of 9) to “high benefit” (score of 44). Fig. 3b presents an additional example of a qualitative presentation of two heart attack levels (no heart attack, mild heart attack).

Qualitative descriptions can also be used to present risk attributes and levels. The *British National Formulary* and the United States Food and Drug Administration provide guidelines on how to report probabilities and other statistical measures [26, 27] (see Table 3).

In elicitation HPR studies, qualitative statements of quantitative levels with no numerical counterparts may encourage respondents to make their own interpretations and assign their own values. Such a result will undermine the HPR study because the research team will lose its ability to model the respondents’ preferences. Such a result may also present additional challenges in the analysis of choice data—for example, in choosing appropriate coding for the quantitative levels of a “high cost” attribute to estimate value trade-offs (such as *willingness to pay*).



**Table 3** Qualitative descriptions of risk.

Qualitative description	Risk
Very common	Greater than 1 in 10
Common	1 in 10 to 1 in 100
Less common	1 in 100 to 1 in 1000
Rare	1 in 1000 to 1 in 10,000
Very rare	Less than 1 in 10,000

### 4.3 Framing of Attributes and Levels

The framing of levels is also an important consideration for researchers. The impact of *framing effects* in HPR surveys has been investigated empirically to explore the influence of contextual cues—such as whether attributes are framed as gains or losses [28–30], and how attributes are presented generally [30, 31].

For example, Veldwijk and colleagues (2016) [32] explored how framing the mortality effects of colorectal cancer as the “probability of surviving” (positively framed) or the “probability of mortality” (negatively framed) influenced preferences for participating in screening. They found that the manner of framing the attribute had a significant influence on preferences: positive framing led to more dominant choices, while negative framing led to increased risk-taking. Similar results have been found in other studies [28], highlighting the need for researchers to consider the potential impact of attribute presentation on preferences arising from choice tasks. The presentation formats described here should be used to help convey and clarify information. Researchers should carefully pretest the HPR study to ensure that the attributes and levels are interpreted correctly and that the presentation aids respondents’ understanding of the choice concept.

Framing may create positive or negative cues, which are interpreted inconsistently across individuals [17, 18]. One potential approach, therefore, is to use what is known as a “cheap talk script” to warn participants about framing effects and other potential biases when completing an HPR [33].

## 5 Specifying Hypotheses

Hypotheses complement the decision model and the descriptive framework. Hypotheses are made at the beginning of an HPR study. Much like the hypotheses that are formally tested in randomized clinical trials, a hypothesis in an HPR study usually establishes a relationship between variables.

### 5.1 Research Question, Hypotheses, and Null Hypotheses

Hypotheses concretize the research question by stating a causal relationship between attributes of objects, characteristics of subjects, and the behavior of respondents. The hypothesis is an assumption that should be statistically falsified or confirmed during an empirical investigation.

Hypotheses are often classified as undirected or directed. Undirected hypotheses refer to a relationship without establishing a specific direction (e.g., “the risk of a side effect affects patients’ utility”). In a directed hypothesis, in contrast, there is a direction in the suspected relationship (e.g., “the risk of a side effect reduces patients’ utility”).

The statistical test of hypotheses shows whether they are true or not. This is possible with the help of a null hypothesis. The null hypothesis (also called “H0”) assumes that there is no effect between attributes of objects, characteristics of subjects, and the behavior of respondents. In contrast, an alternative hypothesis (also called “H1” or “HA”) assumes a direct relationship.

### 5.2 Hypotheses of Correlation, Difference, and Change

Hypotheses make assumptions about the strength and direction of an effect; they examine an influence of one or more attributes (independent variables) on the behavior (dependent variable) of respondents. Main effects are independent effects of a single attribute (or attribute level). The objective is to investigate the causal relationship between utility and the attribute levels.

If a hypothesis involves the relationship between multiple attributes, then interaction effects can be considered. Effects of one attribute level may be impacted by or contingent upon the level of another attribute (e.g., order effects, preferential independence, constant proportionality in time, double counting). The assumption of a linear relationship is not mandatory. Some hypotheses may test nonlinear relationships, such as tapered, exponential, and sigmoidal relationships (Figure 5).

Hypotheses can make assumptions about the heterogeneity of subjects. They can also focus on the differences between two or more study populations. The objective of such hypotheses is to investigate the causal relationship between person-specific characteristics and the utility.

Another assumption hypotheses can make regards the impact of repeating preference tasks. A hypothesis can focus on a behavior change in the same respondents, based on repeated measurements. The objective in this case is to investigate the causal relationship between utility and respondents’ observable behavior.

Two types of error can occur when testing hypotheses: type I error and type II error.

- Type I error (false positive or alpha error):  $H_0$  is rejected although it is correct, or one believes there is an effect in the population although there is none.
- Type II error (false negative or beta error):  $H_0$  is accepted although it is improbable, or one believes there is no effect in the population although there actually is one.

## 6 Limitations, Biases, and Heuristics

Any limitations, biases, or heuristics can distort the study results. The following limitations, biases, and heuristics—which are discussed below—can be triggered by attributes and attribute levels: interaction effects; multicollinearity; information bias; hypothetical bias; starting-point bias; strategic bias; and temporal, spatial, social, or hypothetical distance.

An attribute or attribute level can influence participants' perceptions of other attributes and attribute levels. *Interaction effects* occur when the effect of an attribute or level on behavior changes depending on the value(s) of one or more other attributes. Modeling interaction effects is an important aspect of understanding how different factors influence patient preferences. In HPR, this can reveal complex relationships between patient characteristics, treatment attributes, and health outcomes. Future research may investigate the modeling of interaction effects through various attributes. *Multicollinearity* occurs when attributes are correlated, compromising the model's reliability. Since it is essential for attributes in regression analysis to be independent, future research could focus on establishing standards for reporting the extent of correlation among variables within the descriptive framework. This would facilitate early-stage discussions regarding the interpretation of results, enhancing the clarity and validity of findings.

HPR studies can be subject to several sources of bias. Because these biases can mitigate and/or jeopardize the usefulness of HPR, special care must be taken to minimize bias when presenting the decision model, descriptive framework, or hypothesis. *Information bias* occurs when respondents lack sufficient information to form an accurate response to the attributes and attribute levels of an object. To avoid this type of bias, adequate information must be provided to respondents. *Hypothetical bias* is the tendency for respondents to not give a reasonable answer about a hypothetical scenario, because individuals often provide different responses to hypothetical scenarios than they would in the real world. To avoid this type of bias, studies could be conducted in a laboratory setting, so that researchers could consistently remind respondents to consider the ramifications

that their responses would produce in a real-world setting. Future research should examine how to better simulate real-world decisions, such as by employing hypothetical mitigating strategies or by using nonhypothetical objects to reduce bias. *Starting-point bias* comes into play when respondents are influenced by the initial numbers provided as examples or as part of a range in the decision model or descriptive system. Future research on the types and mode of presentation of the decision model, attributes, attribute levels, and hypothesis can reduce or solve the problem. *Strategic bias* occurs when a respondent intentionally tries to manipulate the outcome of a survey. While it is not possible to eliminate intentionally biased responses, future research should address standards for the presentation of objects to best prevent strategic behavior. The respondent's *sensitivity to scope*—that is, the extent to which a respondent's choice is influenced by the range of attributes or levels presented—is not immediately transparent to the research team. By carefully assessing respondents' sensitivity to different ranges of attribute levels during pretesting, researchers can ensure that their experiments capture true preferences and produce reliable and valid results.

Future research can analyze how study participants' perspectives impact perception and assessment of attributes and their associated levels. For example, *construal level theory* describes the relationship between *psychological distance* and the extent to which decision makers think about attributes in abstract or concrete ways [34, 35]. This theory assumes that the more distant a decision is, the more abstract (and potentially ambiguous) the attributes are within the applied decision framework (and vice versa). The dimensions of distance are *temporal*, *spatial*, *social*, or *hypothetical*.

## 7 Conclusion

HPR has emerged as an invaluable tool in the landscape of healthcare decision-making. This paper has systematically explored the intricate process of designing an HPR study, highlighting the importance of meticulously identifying, describing, and presenting attributes and attribute levels. It underscores the necessity of a well-constructed decision model, which serves as the foundation for understanding complex healthcare choices in varied contexts.

The identification and validation of attributes, a process detailed in the paper, demands a nuanced blend of both explorative and elicitation methods, informed by rigorous literature reviews and qualitative analyses. This comprehensive approach ensures that the attributes encapsulated within the study are reflective of real-world considerations and preferences.

As we discussed the realm of setting attribute levels, the paper illuminated the delicate balance required in choosing levels that are both meaningful and manageable, to accurately capture the nuances of respondent preferences without overwhelming them. The discussion on deterministic, probabilistic, nominal, ordinal, interval, and ratio attributes offered a roadmap for tailoring attribute levels to the specificities of each study, ensuring that they align with the objectives and context of the research.

The paper further highlighted the significance of constructing a descriptive framework that effectively communicates the attributes and their levels to respondents. It emphasized the role of graphical illustrations, qualitative presentations, and careful framing in aiding comprehension and mitigating biases. This aspect is crucial in ensuring that the elicited preferences are a true reflection of respondents' values and perceptions.

The paper adeptly navigated through the complexities of different types of hypotheses, providing a thorough understanding of their roles in substantiating the research. Finally, the paper acknowledged the potential limitations and biases inherent in HPR, stressing the importance of recognizing and mitigating these factors to ensure the integrity and applicability of the study findings. It offered insights into future research directions that could further refine HPR methodologies and enhance their effectiveness in capturing health preferences.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s40271-024-00673-y>.

**Acknowledgements** Editorial services were provided by Jonathan Patishall of RTI Health Solutions.

#### Declarations

**Author Contributions** All authors contributed to the conception and design of the paper. The first draft of the manuscript was written by all authors, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

**Funding** Open Access funding enabled and organized by Projekt DEAL. E.W. de Bekker-Grob was funded via the Dutch Research Council (NWO-Talent-Scheme-Vidi-Grant No. 09150171910002).

**Conflict of Interest** Dr. Axel Mühlbacher is an employee of Hochschule Neubrandenburg and GEB Gesellschaft für empirische Beratung mbH. Dr. Esther W. de Bekker-Grob is an employee of Erasmus University Rotterdam. Dr. Oliver Rivero-Arias is an employee of the University of Oxford. Dr. Caroline Vass is an employee of RTI Health Solutions. Dr. Bennett Levitan is an employee of Janssen Research and Development, LLC. He is a stockholder in Johnson & Johnson and a portfolio that at times includes other pharmaceutical and healthcare-related companies.

**Data Availability** Not applicable.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any

non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

## References

1. Lancaster KJ. A new approach to consumer theory. *J Polit Econ*. 1966;74(2):132–57.
2. Dolan P, et al. An inquiry into the different perspectives that can be used when eliciting preferences in health. *Health Econ*. 2003;12(7):545–51.
3. Coast J, Horrocks S. Developing attributes and levels for discrete choice experiments using qualitative methods. *J Health Serv Res Policy*. 2007;12(1):25–30.
4. Soekhai V, et al. Methods for exploring and eliciting patient preferences in the medical product lifecycle: a literature review. *Drug Discov Today*. 2019;24(7):1324–31.
5. Neuman W, Robson K. *Basics of social research*. Toronto: Pearson Canada; 2014.
6. Kline P. *An easy guide to factor analysis*, vol. 17. London: Routledge; 1994.
7. Goodman LA. Snowball sampling. *Ann Math Stat*. 1961;32(1):148–70.
8. Silverman D. *Interpreting qualitative data*. Sage Publications; 1993.
9. Critical Appraisal Skills Programme. *10 questions to help you make sense of qualitative research*. Public Health Resource Unit 2006 [cited 2023 26 July]; [https://casp-uk.net/images/checklist/documents/CASPQualitative-Studies-Checklist/CASP-Qualitative-Checklist-2018\\_fillable\\_form.pdf](https://casp-uk.net/images/checklist/documents/CASPQualitative-Studies-Checklist/CASP-Qualitative-Checklist-2018_fillable_form.pdf). Accessed 08 Jul 2023.
10. Pearson A, Field J. *The Joanna Briggs Institute Critical Appraisal tools for use in JBI: Systematic Reviews Checklist for Qualitative Research*. Joanna Briggs Institute 2017 [cited 2023 26 July]; [https://jbi.global/sites/default/files/2019-05/JBI\\_Critical\\_Appraisal-Checklist\\_for\\_Qualitative\\_Research2017\\_0.pdf](https://jbi.global/sites/default/files/2019-05/JBI_Critical_Appraisal-Checklist_for_Qualitative_Research2017_0.pdf). Accessed 08 Jul 2023.
11. Dixon-Woods M. The problem of appraising qualitative research. *Qual Saf Health Care*. 2004;13(3):223–5.
12. Hollin IL, et al. Reporting formative qualitative research to support the development of quantitative preference study protocols and corresponding survey instruments: guidelines for authors and reviewers. *Patient*. 2020;13:121–36.
13. Flynn TN. Valuing citizen and patient preferences in health: recent developments in three types of best-worst scaling. *Expert Rev Pharmacoecon Outcomes Res*. 2010;10:259–67.
14. Stenner P, Watts S. Doing Q methodology: theory, method and interpretation. *Qual Res Psychol*. 2005;2:67–91.
15. Schmidt K, et al. Applying the Analytic Hierarchy Process in healthcare research: A systematic literature review and evaluation of reporting. *BMC Med Inform Decis Mak*. 2015;15:112.
16. Ralley C, et al. The use of the repertory grid technique to examine staff beliefs about clients with dual diagnosis. *Clin Psychol Psychother*. 2009;16(2):148–58.
17. Tversky A, Kahneman D. The framing of decisions and the psychology of choice. *Science*. 1981;211(4481):453–8.

18. Levin IP, Schneider SL, Gaeth GJ. All frames are not created equal: a typology and critical analysis of framing effects. *Organ Behav Hum Decis Process*. 1998;76(2):149–88.
19. Harrison M, et al. Risk as an attribute in discrete choice experiments: a systematic review of the literature. *Patient*. 2014;7(2):151–70.
20. 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(2):219–28.
21. Han PK, et al. Representing randomness in the communication of individualized cancer risk estimates: effects on cancer risk perceptions, worry, and subjective uncertainty about risk. *Patient Educ Couns*. 2012;86(1):106–13.
22. Veldwijk J, et al. Words or graphics to present a Discrete Choice Experiment: does it matter? *Patient Educ Couns*. 2015;98(11):1376–84.
23. Vass C, Rigby D, Payne K. "I Was Trying to Do the Maths": exploring the impact of risk communication in discrete choice experiments. *Patient*. 2019;12(1):113–23.
24. Timmermans DR. Prenatal screening and the communication and perception of risks. *Int Congr Ser*. 2005;1279:234–43.
25. Marshall DA, et al. How do women trade-off benefits and risks in chemotherapy treatment decisions based on gene expression profiling for early-stage breast cancer? A discrete choice experiment. *BMJ Open*. 2016;6(6): e010981.
26. U.S. Food and Drug Administration. *Statistical Guidance on Reporting Results from Studies Evaluating Diagnostic Tests*. 2016 [cited 2023 15 May]; Available from: <https://www.fda.gov/media/71146/download>. Accessed 08 Jul 2023.
27. Joint Formulary Committee, *British National Formulary*. BMJ Group and Pharmaceutical Press, 2022.
28. Howard K, Salkeld G. Does attribute framing in discrete choice experiments influence willingness to pay? Results from a discrete choice experiment in screening for colorectal cancer. *Value Health*. 2009;12(2):354–63.
29. Cormier O'Connor AM, et al. Eliciting preferences for alternative cancer drug treatments: the influence of framing, medium, and rater variables. *Med Decis Making*. 1985;5(4):453–63.
30. Kragt ME, Bennett JW. Attribute framing in choice experiments: how do attribute level descriptions affect value estimates? *Environ Resour Econ*. 2012;51(1):43–59.
31. Kenny P, et al. Choosing a doctor: does presentation format affect the way consumers use health care performance information? *Patient*. 2017;10(6):739–51.
32. Veldwijk J, et al. Survival or mortality: does risk attribute framing influence decision-making behavior in a discrete choice experiment? *Value Health*. 2016;19(2):202–9.
33. Cummings RG, Taylor LO. Unbiased value estimates for environmental goods: a cheap talk design for the contingent valuation method. *Am Econ Rev*. 1999;89(3):649–65.
34. Trope Y, Liberman N. Construal-level theory of psychological distance. *Psychol Rev*. 2010;117(2):440–63.
35. Mühlbacher AC, Juhnke C. Patient preferences versus physicians' judgement: does it make a difference in healthcare decision making? *Appl Health Econ Health Policy*. 2013;11:163–80.

## Authors and Affiliations

Axel C. Mühlbacher<sup>1</sup>  · Esther W. de Bekker-Grob<sup>2</sup>  · Oliver Rivero-Arias<sup>3</sup>  · Bennett Levitan<sup>4</sup>  · Caroline Vass<sup>5</sup> 

✉ Axel C. Mühlbacher  
muehlbacher@hs-nb.de

<sup>1</sup> HS Neubrandenburg, Brodaer Straße 2,  
17033 Neubrandenburg, Germany

<sup>2</sup> Erasmus School of Health Policy & Management, Erasmus  
Choice Modelling Centre, Erasmus Centre for Health  
Economics Rotterdam, Erasmus University Rotterdam,  
Rotterdam, The Netherlands

<sup>3</sup> Nuffield Department of Population Health, University  
of Oxford, Oxford, UK

<sup>4</sup> Janssen R&D, Titusville, NJ, USA

<sup>5</sup> RTI Health Solutions, Manchester, UK