

# prefer.

PATIENT PREFERENCES

**CHMP & EUnetHTA parallel Scientific Advice:  
Qualification of a Framework and “Points to consider” for method selection along with  
five methods for performing patient preference studies to inform regulatory and HTA-  
body medical product decision-making**

Briefing document

Applicant: IMI PREFER  
Version: 2.0  
Release date: 5 March 2021

**Table of contents**

	Table of contents.....	2
	List of tables.....	5
	List of figures.....	7
	Abbreviations.....	8
	Glossary.....	10
	Overview.....	13
1	Introduction.....	13
1.1	Problem statement .....	14
1.2	PREFER objective and objectives of the qualification .....	17
1.3	Context of use .....	17
1.4	Out-of-scope of this qualification .....	17
1.5	Overview of Briefing Book content .....	18
1.6	Data supporting the qualification .....	19
1.7	Publication and transparency of patient preference study results .....	20
2	Value of patient preferences and interface with stakeholders .....	21
2.1	Situations when patient preference data are most likely to be useful to support decision-making (patient preference sensitive decisions) .....	21
2.2	Stages in the medical product life cycle where patient preference sensitive decisions may occur .....	23
2.2.1	Medical product discovery and development .....	23
2.2.2	Marketing authorisation .....	24
2.2.3	HTA & Reimbursement decision-making .....	28
2.2.4	Post-marketing .....	29
2.2.5	Additional notes .....	30
2.3	Situations when patient preference data is less likely to be useful to support decision-making .....	31
2.4	Interactions with patients as participants and partners on patient preference studies.....	32
2.4.1	Why is it important to interact with patients as research partners?.....	32
2.4.2	General principles for interactions with patients as participants in patient preference studies .....	33
2.4.3	Interactions with patients as research partners in patient preference studies.....	34
2.4.4	Profile of patients as partners .....	39
2.5	Interactions with regulators and HTA bodies on patient preference studies .....	41
2.5.1	Why is it important and when is the best time to engage regulators and HTA bodies? .....	41
2.5.2	How should regulators / HTA bodies be consulted in preference studies ...	42
2.5.3	What topics may be covered in regulatory / HTA bodies scientific advice.....	42
3	PREFER framework for patient preference studies.....	44
3.1	Objectives and overview of the PREFER framework for patient preference studies.....	44
3.1.1	PREFER framework objectives and overview .....	44
3.1.2	Key considerations relating to the PREFER framework.....	46

3.2	PREFER Framework component 1: the preference study purpose and objectives.....	50
3.2.1	Framework component 1: the preference study purpose .....	51
3.2.2	Framework component 1: preference study objectives.....	52
3.3	PREFER framework component 2: criteria for organising, designing and conducting a fit-for-purpose preference study .....	53
3.3.1	Framework component 2, organisation .....	53
3.3.2	Framework component 2, design.....	56
3.3.3	Framework component 2, conduct.....	73
3.4	PREFER Framework component 3: applying preference data to inform medical product decision-making .....	78
3.4.1	Applications of preference data to inform medical product decision making .....	78
3.4.2	Technical methods for the application of preference data .....	90
3.4.3	Incorporating preference information into industry, regulatory and HTA/payer documents .....	102
4	Methods.....	106
4.1	Background on the method selection process.....	106
4.1.1	Identifying preference methods.....	106
4.1.2	PREFER case studies and methodological research questions .....	106
4.2	Discrete Choice Experiment (DCE) .....	107
4.2.1	Introduction .....	107
4.2.2	DCE-specific aspects of framework component 2 “design: method selection and analysis planning”.....	109
4.2.3	Points to consider in choice of preference elicitation method: i.e. why DCE and not something else? .....	111
4.2.4	DCE-specific aspects of framework component 2 “design: sample definitions – justifying the sample size” .....	112
4.2.5	DCE-specific aspects of framework component 2 “design: preference question design – consideration of the appropriate number of attributes and attribute levels; patient burden issues” .....	113
4.2.6	Summary .....	114
4.3	Swing Weighting .....	114
4.3.1	Introduction .....	114
4.3.2	SW-specific aspects of framework component 2 “design: method selection and analysis planning”.....	117
4.3.3	Points to consider when setting up the preference question design.....	120
4.4	Best-Worst Scaling: Case 1.....	121
4.4.1	Introduction .....	121
4.4.2	Why Best-Worst Scaling Case 1?.....	122
4.4.3	Best-Worst Scaling Case 1-Specific aspects of framework component 2 “Design: method selection and analysis planning” .....	123
4.4.4	Best-Worst Scaling Case 1- specific aspects of framework component 2 ‘design: sample definitions—justifying the sample size’ .....	126
4.4.5	Best-Worst Scaling Case 1-Specific Aspects of Framework Component 2 “Design: Preference Question Design—Consideration of the Appropriate Number of Attributes and Attribute Levels; Patient Burden Issues” .....	127

4.5	Best-Worst Scaling: Case 2.....	128
4.5.1	Introduction .....	128
4.5.2	Why BWS case 2? .....	129
4.5.3	Best-Worst Scaling Case 2- specific aspects of framework component 2 'design: method selection and analysis planning' .....	129
4.5.4	BWS case 2-specific aspects of framework component 2 'design: sample definitions— justifying the sample size" .....	131
4.5.5	BWS case 2-specific aspects of framework component 2 'design: instrument design—consideration of the appropriate number of attributes and attribute levels; patient burden issues" .....	132
4.6	Threshold Technique .....	133
4.6.1	Introduction .....	133
4.6.2	Why TT? .....	134
4.6.3	TT-specific aspects of framework component 2 'design: method selection and analysis planning" .....	134
4.6.4	TT-specific aspects of framework component 2 'design: sample definitions—justifying the sample size" .....	137
4.6.5	TT-specific aspects of framework component 2 'design: instrument design – consideration of the appropriate number of attributes and attribute levels; patient burden issues" .....	138
5	Points to Consider for methods selection .....	140
5.1	Methodological factors.....	141
5.1.1	Does the method provide estimates at the level of the individual, the sample, or both?.....	143
5.1.2	Can the method account for and/or test for sources of preference heterogeneity?.....	143
5.1.3	What minimum sample size is required to implement the method?.....	143
5.1.4	Can the method elicit preferences for individual treatment features or attributes? If so, how many features or attributes can be evaluated?....	144
5.1.5	Can the method elicit preferences over multiple levels of each feature or attribute? .....	144
5.1.6	Does the method elicit relative importance, trade-offs, or both?.....	144
5.1.7	If the method can be used to estimate trade-offs, can the method estimate tradeoffs among multiple attributes simultaneously (or are tradeoffs estimated pairwise)? .....	144
5.1.8	Can the research method accommodate interactions between treatment characteristics? .....	145
5.2	Participant factors .....	145
5.2.1	Do patients have the cognitive capacity to complete the preference elicitation exercise successfully?.....	145
5.2.2	Is the sample large enough to provide enough information to implement the method? .....	145
5.3	Feasibility Factors.....	145
5.3.1	Is specific methodological expertise required to design the data collection materials? .....	146
5.3.2	Is specific methodological expertise required to implement the study and acquire the necessary data? .....	146

5.3.3	Is specific methodological expertise required to analyze data and interpret the results of the analysis?.....	146
5.4	PREFER case studies .....	146
6	Conclusion.....	151
6.1	PREFER framework for patient preference studies.....	151
6.2	Overview of commonly used and well understood methods.....	152
1.1	Points to consider for the methods selection .....	152
7	Questions and Applicant’s positions .....	153
	Question 1.....	153
	Question 2.....	153
	Question 3.....	154
8	Annexes.....	171
8.1	Annex I: Abstracts of all case studies.....	171
	8.1.1 Main case studies in PREFER.....	171
	8.1.2 Prospective additional industry case studies .....	192
	8.1.3 Prospective additional academic case studies.....	233
8.2	Annex II: PREFER: Historical case studies.....	257
8.3	Annex III: PREFER methodological Research Questions .....	261
8.4	Annex IV: Catalogue of identified patient preference study methods.....	267

**List of tables**

Table 2-1	Known applications of patient preference studies .....	26
Table 2-2	Examples of patient inclusion in prospective PREFER case studies .....	34
Table 2-3	Defining the study purpose.....	35
Table 2-4	PREFER case study examples of patient involvement in formulating questions .....	36
Table 2-5	PREFER case study examples of patient involvement in defining attributes and levels.....	36
Table 2-6	PREFER case study examples of patient involvement in survey presentation.....	37
Table 2-7	Example of patient involvement across the whole RA patient preference study.....	38
Table 2-8	Potential topics to be raised during regulator / HTA body consultation .....	42
Table 3-1	Link between Framework and Methods section.....	46
Table 3-2	Example of a case study that evaluated preference heterogeneity .....	48
Table 3-3	How heterogeneity in patients’ views could impact specific types of decisions.....	49
Table 3-4	How the PREFER framework can help address potential biases .....	50
Table 3-5	Example of the expertise areas of study team members supporting the PREFER case study .....	54
Table 3-6	Example of PREFER case study timing alignment with study purpose .....	55
Table 3-7	Examples of preference studies’ approach to the choice of preference study population .....	58
Table 3-8	Pros and cons of recruiting preference study participants completely within or completely outside of clinical trials .....	59

Table 3-9	Example of a preference study population from a combination of within and outside a clinical trial.....	60
Table 3-10	Example of a preference study population including care-giver population.....	61
Table 3-11	Example of case study approaches to study analysis planning.....	64
Table 3-12	Case study showing bottom-up approach to alternative description development .....	67
Table 3-13	Case study showing the PREFER framework approach for patient education and comprehension considerations in preference studies .....	71
Table 3-14	Examples of how case studies updated a questionnaire based on piloting with patients .....	74
Table 3-15	Suggested structure for communication of preference study results to study participants .....	77
Table 3-16	Example of how a PREFER case study assessed patients' views on choice of endpoints via choice share .....	80
Table 3-17	Example of a case study assessing patients' views on relative importance of attributes.....	81
Table 3-18	Example of how a preference study could inform a decision about PRO instrument scoring .....	83
Table 3-19	Examples of how preference studies informed decisions about the acceptability of benefit-risk trade-offs.....	85
Table 3-20	Example of how a preference study informed a decision about acceptability of uncertainty.....	88
Table 3-21	Example of how a preference study informed cost-effectiveness analysis .....	90
Table 3-22	Estimate of preference scores from Ho.....	92
Table 3-23	Example of preference data included in labelling information .....	104
Table 4-1	Summary of methodological research questions and methods in PREFER case studies.....	106
Table 4-2	Categories of points to consider when selecting a patient preference elicitation method .....	139
Table 5-1	Categories of points to consider when selecting a patient preference elicitation method .....	140
Table 5-2	Assessment of methods relative to methodological points to consider ....	142
Table 5-3	Learnings from the Pain Case Study on Points to Consider Regarding Methods Selection .....	147
Table 5-4	Learnings from the Multiple Myeloma (MM) Case Study on Points to Consider Regarding Methods Selection.....	148
Table 8-1	Listing of historical case studies (wave 1).....	258
Table 8-2	PREFER methodological research questions related to "Preference Question Design" and relevant case study examples.....	261
Table 8-3	PREFER methodological research questions related to "Conduct: Recruitment, and Analysis and Interpretation" with relevant PREFER case study examples .....	262
Table 8-4	PREFER methodological research questions related to methods.....	263

**List of figures**

Figure 1-1 Patient preferences as type of patient input ..... 14

Figure 2-1 Mapping preference-sensitive scenarios to stakeholder decisions..... 22

Figure 2-2 PREFER cases studies throughout all MPLC stages ..... 30

Figure 2-3 Situations where patient preference studies might be useful and less useful ..... 32

Figure 2-4 How patients can be involved at all stages of a preference study..... 35

Figure 3-1 The three broad components of the PREFER patient preference framework..... 45

Figure 3-2 Framework structure, aligned with stages and steps of patient preference studies..... 46

Figure 3-3 Mock example of using a preference study to develop an effect size for a novel endpoint: ..... 93

Figure 3-4 Example of maximum acceptable combination of risks for a given level of benefit. .... 94

Figure 3-5 Example table showing clinical data side-by-side with preference weights (mock data and weights) ..... 96

Figure 3-6 Example of a forest plot with endpoints in order of decreasing health state utility..... 97

Figure 3-7 Example figure showing location of bars at locations proportional to their preference weights ..... 98

Figure 3-8 Example parallel bar plot showing endpoints in order of decreasing weight ..... 98

**Abbreviations**

AHP	Analytic Hierarchy Process
B/R	Benefit/Risk
BRA	Benefit-Risk Assessment
BRAT	Benefit-Risk Action Team
BWS	Best Worst Scaling
CDRH	Center for Devices and Radiological Health
CIOMS	Council for International Organizations of Medical Sciences
CHMP	Committee for Medicinal Products for Human Use
DCE	Discrete Choice Experiment
DRPS	doubly randomised preference study
EMA	European Medicines Agency
EPAR	European public assessment report
EUnetHTA	European Network for Health Technology Assessment
FDA	Food and Drug Administration
FDR	First-degree relative
FXS	Fragile X syndrome
GDPR	General Data Protection Regulation
HTA	Health Technology Assessment bodies
IAHPR	International Academy of Health Preference Research
IEC	Independent Ethics Committee
IMI	Innovative Medicines Initiative
IQWiG	Institute for Quality and Efficiency in Health Care (Germany)
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
KCE	Belgian Health Care Knowledge Centre
MA	Marketing Authorisation
MAB	Minimum Acceptable Benefit
MAR	Maximum Acceptable Risk
MCDA	Multi-criteria Decision Analysis
MDIC	Medical Device Innovation Consortium
MPLC	Medical Product Lifecycle
NCB	Net Clinical Benefit



---

NGT	Nominal Group Technique
NICE	National Institute for Health and Care Excellence (UK)
PCBR	Patient Centered Benefit-Risk
PFDD	Patient Focused Drug Development
PP	Patient Preferences
PPS	Patient Preference Studies
PRO	Patient-Reported Outcome
PROTECT	Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium
QoL	Quality of Life
QALYs	Quality Adjusted Life Years
RTI-HS	RTI Health Solutions
SG	Standard Gamble
SmPC	Summary of Product Characteristics
TPP	Target Product Profile
TTO	Time Tradeoff

**Glossary**

<b>Terms</b>	<b>Definition</b>
<b>Attribute</b>	An attribute is a feature or characteristic of a medical product—such as efficacy or effectiveness, safety, means of administration, duration of effect, or duration of use, burden on patients or care givers—that may affect benefit-risk considerations.
<b>Benefits</b>	Benefits are the favourable effects of a medical product. Types of benefit include clinical benefit (see definition below). Benefits may also include important characteristics of the medical product, such as convenience (e.g., a more convenient dosing regimen or route of administration) that may lead to improved patient compliance, or benefits that affect those other than the patient
<b>Best Worst Scaling (BWS)</b>	BWS is a survey method for assessing individuals' priorities. It identifies the extremes—best and worst items, most and least important factors, biggest and smallest influences—among sets.
<b>Context of use</b>	Concise description of the method's or framework's intended use in medical product life-cycle
<b>Discrete Choice Experiment (DCE)</b>	A discrete choice experiment is a quantitative method increasingly used in healthcare to elicit preferences from participants (patients, payers, commissioners) without directly asking them to state their preferred options.
<b>Decision</b>	A judgment, conclusion or determination reached after consideration. A decision is a response in a situation that is composed of three parts: <ol style="list-style-type: none"> <li>1. There is more than one possible course of action in the choice set</li> <li>2. The decision maker can form expectations about the outcomes that follow from each course of action</li> <li>3. The consequences of the outcomes can be assessed relative to current goals and values.</li> </ol>
<b>Fit-for-purpose</b>	A conclusion that the level of validation associated with a method or tool is sufficient to support its context of use.
<b>Framework</b>	Set of principles, guidelines and tools or a process that is used when forming a decision
<b>Maximum acceptable risk (MAR)</b>	Greatest increase in probability or magnitude of a harm a patient would accept to achieve or realize a given benefit
<b>Medical Product</b>	Any product used to diagnose, treat or manage patients; includes medicinal products, devices and services
<b>Medical Product Lifecycle (MPLC)</b>	The development, authorisation and post-authorisation phase of a medical product can be divided into multiple different steps, the sum of these steps are called "lifecycle". The medical product lifecycle herein is defined as the lifecycles of drugs, biologics and medical devices.
<b>Minimum anticipated benefit (MAB)</b>	Smallest anticipated benefit needed for a patient to accept a given risk

<b>Minimum required benefit (MRB)</b>	Smallest increase in probability or magnitude of a benefit a patient would require to accept a given risk
<b>Patient Preferences</b>	Patient Preference is a type of patient experience data. Qualitative or quantitative assessments of the relative desirability or acceptability to patients of specified alternatives or choices among outcomes or other attributes that differ among alternative health interventions
<b>Patient Relevant Outcome</b>	An outcome that is meaningful to the patient, which can be health or non-health related. Examples may include outcomes related to how a patient feels or functions in daily life, quality of life or utility measures, or economic outcomes
<b>Patient Reported Outcome</b>	Any report of the status of a patient’s health condition that comes directly from the patient, without interpretation of the patient’s response by a clinician or anyone else
<b>Patient preference sensitive decisions</b>	Patient preferences are a source of evidence that may inform particular decisions that are sensitive to patient preferences, where the perspective of the patients really matter.
<b>Qualitative Preference (Exploration) Method</b>	Methods that collect descriptive data through participant or phenomenon observation, and examining the subjective experiences and decisions made by participants
<b>Quantitative Preference (Elicitation) Method</b>	Methods collecting quantifiable data that can be reported through statistical inferences or analysis
<b>Quality-adjusted life year (QALY)</b>	A measure of the state of health of a person or group in which the benefits, in terms of length of life, are adjusted to reflect the quality of life
<b>Risks</b>	Risks are adverse events and other unfavourable effects associated with a medical product. Risks include drug interactions, risks identified in the non-clinical data, risks to those other than the patient (e.g., foetus, those preparing and administering the medical product), and risks based on pharmacologic class or current knowledge of the product. Factors such as potential misuse, abuse, or diversion of the product may also be considered.
<b>Swing weighting</b>	Swing weighting is a trade-off weighting method, in which the relative importance is determined on the basis of moving from the worst to best score on a scale (full swing).
<b>Trade-off</b>	The extent to which a change in the level of one or more attributes of a medical product is perceived to be offset by a change in one or more other attributes of that product
<b>Threshold Technique (TT)</b>	The threshold technique is a stated preference method that quantifies both preference and the strength of preference between two healthcare options from the perspective of a decision-maker. The process is implemented by asking a decision-maker to choose between two alternatives and systematically varying the level of a key attribute of one

---

	alternative until the decision-maker indicates the level at which he or she is indifferent between the two alternatives. The initial choice indicates the preference for one alternative over another. The amount of change in the level of the key attribute required to induce a change in the initial choice is a measure of the strength of preference.
<b>Unmet healthcare need</b>	Condition for which there exists no satisfactory method of diagnosis, prevention, or intervention to improve patient care available in the Community or, even if such methods exist, in relation to which the alternative method concerned will be of major therapeutic advantage to those affected

## Overview

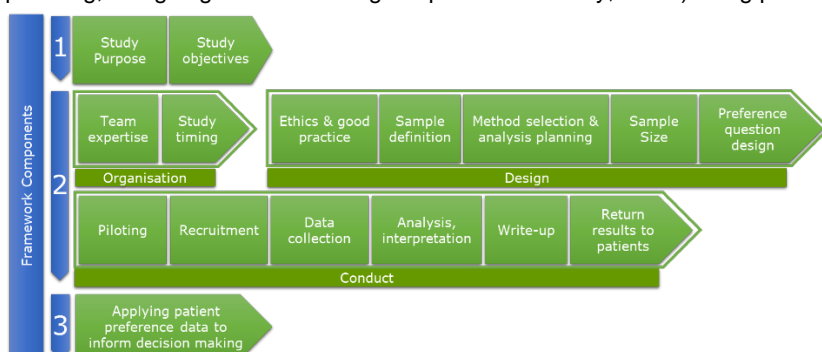
Patients' preferences are a growing topic of interest, and stakeholders (e.g. regulators, payers, industry, and patient organizations) have called for greater involvement of patients. However, to date, there are no generally accepted recommendations to help guide applicants in the design, conduct and analysis of patient preference studies that would be suitable to inform regulatory and reimbursement decision-making throughout the treatment development life cycle. This document aims to provide an initial set of tools to address this gap, based on the systematic review of available evidence and the evidence generated by IMI PREFER. Hence, PREFER aims to seek qualification for a framework and 'Points to Consider' for methods selection as a way of providing insight and suggestions on how patients' perspectives could be measured through patient preference studies and then incorporated into the regulatory approval process and inform HTA and payer decision making. These perspectives include what matters to patients, how much it matters, the acceptability of trade-offs between benefits and harms, and the acceptability of uncertainty

PREFER took a structured approach to this qualification that goes beyond what is presented in this document. The foundation was built on systematic literature searches and comprehensive stakeholder interviews. Key results of that foundation are summarized as background and motivation on the value of patient preferences (Section 2), and on interfaces with stakeholders for the main three sections of this document (Sections 3 to Section 5). The foundational work also informed the research and operational plans of PREFER, which led to a series of case studies to address high priority research questions, specifically for those methods that were assessed as most promising (Annex I).

The objectives of the PREFER framework for patient preference studies is to serve as a tool that:

- 1 Informs a preference study research team on key considerations when designing, conducting and applying the results of a fit-for-purpose preference study
- 2 Guides decision-makers when assessing and using preference study results to inform decision-making
- 3 Supports the discussion between industry, regulators, HTA bodies and payers about preference studies intended to inform medical product decision-making.

The PREFER framework comprises three components: 1) defining the preference study purpose and objectives, 2) planning, designing and conducting the preference study, and 3) using preference study results.



This qualification package presents five methods (discrete choice experiment [DCE], best-worst scaling [BWS] case 1, BWS case 2, threshold technique [TT], and swing weighting [SW]) (section 4), which can systematically quantify patients' preferences and reveal insights into patients' perspectives to generate data that could then be used to inform regulatory or HTA and payer decision-making. This list represents a set of suitable methods that have been used to inform regulatory or HTA and payer decision making, contains different types of methods and comes with a relevant body of evidence.

The 'Points to consider' on method selection complements a related chapter in the PREFER framework (section 3) and describes the methodological, participant and feasibility factors that are relevant for the selection of a suitable method. These factors are then used to evaluate the five methods described above to consider to the extent possible. These factors can also be used to evaluate additional methods if the original five are considered unsuitable for use in a particular preference study.

The full description of the framework for preference studies, the overview of the five commonly used methods and the 'Points to consider' for method selection are not only intended to provide a clear path forward for PPS using one of these methods, but may help to establish baseline expectations and support discussions between industry, regulators and HTA bodies/payers about preference studies, and hence support the broader aim of increasing patients' input to medical product decision-making.

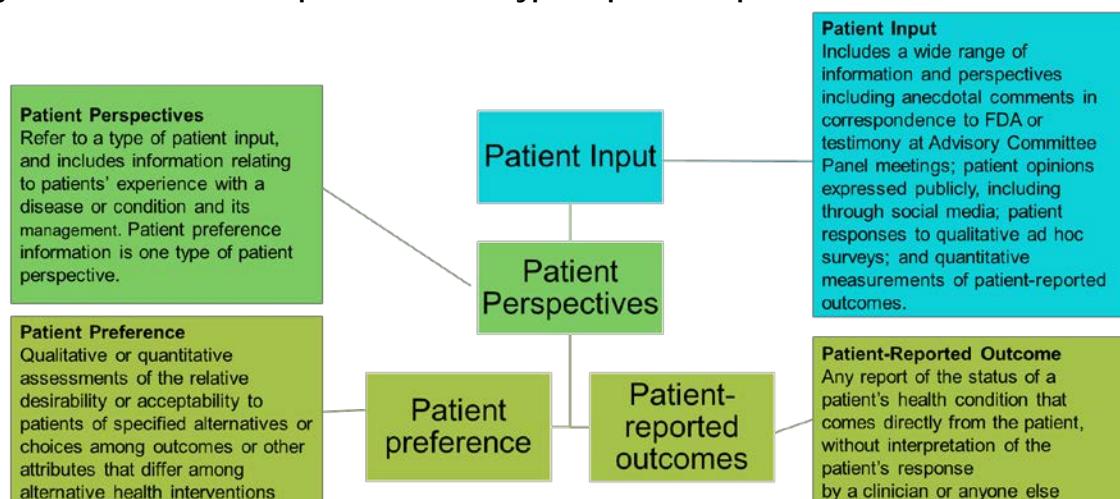
# 1 Introduction

## 1.1 Problem statement

Stakeholders involved in healthcare increasingly view patients' perspectives as important in decision-making throughout the medical product life cycle, specifically to inform research and development of drugs and devices (Van Overbeeke E 2019; Parsons S 2016; Lowe MM 2016; Crocker JC 2016) as well as benefit-risk assessments (BRA) (Christiaens W, 2012) and health technology assessments (HTA) (Bilvick Tai BW 2016; Abelson J 2016). One of the underlying reasons for this is that patients are not only the people benefiting from potential new medicines, but are also those exposed to potential risks and burdens, especially upon use after approval. In addition, patients might have different preferences compared to decision makers upon which they make benefit-risk trade-offs (MDIC, 2015; Ho MP 2015; Marsh K 2018; Mühlbacher AC and Kaczynski A. 2016; Marsh K 2017; Mol PG, 2015). Finally, patients' preferences for health benefits may also depend on the unmet need within the target indication, which may influence the assessment of the added value of a new treatment in an HTA.

The European Medicines Agency (EMA) and the United States Food and Drug Administration (FDA), HTA bodies and reimbursement agencies, as well as industry, have all expressed an interest in incorporating patients' voices in their decision-making (EMA 2013, Hockley K 2014; Christiaens W, 2012; Mülbacher AC & Marsh 2019; CDRH 2015; EMA Regulatory science strategy reflection 2020; EMA, ICH guideline M4E(R2), 1999 ). The patients' voice can be considered in many ways to inform decision-making (Figure 1-1) and patient input may cover different types of information captured in divergent ways. Of particular interest in medical decision-making is the patient's experience with a specific disease and its treatment, one specific type of this patient perspective is called patient preference information.

**Figure 1-1 Patient preferences as type of patient input**



Patient Perspectives in the CDRH Guidance is analogous to Patient Experience Data in 21<sup>st</sup> Century Cures

Source: FDA Final Patient Preference Guidance Document. August 24, 2016. <https://www.fda.gov/media/92593/download> and <https://www.fda.gov/media/139088/download>

### Fundamental differences between PROs and patient preferences

**Patient preferences** could be elicited to provide the relative importance of different PROs for patients, though the preferences are not the PROs themselves. Patient preference studies reflect the values of patients in respect to treatment alternatives or their attributes. Patient preference studies, however, can utilize information from PROs and help interpret and weigh changes in PROs as well as help collect information for PRO development on which attributes are meaningful to patients.

In short, patient preferences can tell you what to measure that matters to the patient, and PROs are one means by which to measure them in clinical studies (the how)

Source: 1) [Mohamed A, 2010](#) ; 2) [Patalano F 2020](#)

**Patient Reported Outcomes (PROs)** collect data on health status directly from the patient, without comparison to alternatives or inclusion of preferences; that is, PROs are realized outcomes. PROs are complementary to preferences, but are not the same. They are both assessed via interviews or survey instruments completed by patients, but they measure distinct concepts.<sup>1,2</sup>

Regulators, HTA bodies and industry wishing to incorporate the patient's perspective in decision-making need commonly agreed definitions as well as approaches for capturing these perspectives during development and assessments of medical treatments. "Patient preference information" as defined by both the FDA and this document means "the information resulting from qualitative or quantitative assessments of the relative desirability or acceptability to patients of specified alternatives or choices among outcomes or other attributes that differ among alternative health interventions." (CDRH). The term "patient preferences" refers to preferences not only from patients, but also from caregivers, including parents.

On a regulatory level, the FDA has issued guidance on patient preference information for medical devices, generated via patient preference studies (CDRH). The Council for International Organizations of Medical Sciences (CIOMS) is additionally developing guidance in this context ([CIOMS](#); [EMA, ICH guideline M4E\(R2\)](#) ). At a national level, certain HTA bodies such as the German Institute for Quality and Efficiency in Health Care (IQWiG) or the Belgian Health Care Knowledge Centre (KCE) have provided recommendations on how to incorporate the views of patients and the public more consistently in reimbursement and resource allocation decisions ([Christiaens W, 2012](#); [KCE, 2019](#)). To date however, no clear framework exists on using patient preference studies to inform medical product decision-making.

Thus far, there are only a few published examples of patient preferences in regulatory decisions; for instance to inform regulatory BRA of a medical device for obese patients ([Ho MP, 2015](#)) or to evaluate patient preferences in the context of particular administration forms of medicines ([Rummel M 2015](#)) (see Table 1-1). Examples of using patient preferences at the reimbursement level include the conduct of two preference studies by IQWiG ([Hummel MJM, 2012](#); [IQWiG, 2014](#); [Postmus D, 2016](#); [Postmus D, 2018](#)) and some other initiatives ([Mühlbacher AC, 2017](#); [Menon D, 2014](#)).

In summary, while there is clear consensus from all stakeholders for the need to consider the patient perspective, so far these efforts have focused on adding individuals or groups of patient representatives to the discussion either in pre-decision conversations or during decision-making meetings. Increasing both the level of evidence provided by patients and their representativeness is the key value proposition of patient preference studies.

Patient preference studies generate patient preference data that have the potential to inform medical product decision-making in addition to existing evidence sources along the medicinal product life cycle. This is particularly the case in decision-making contexts that are sensitive to the preference of the patient (see [Section 2.1](#)), called patient preference-sensitive decisions. Building on the concept of patient preference-sensitive situations from the [MDIC report \(2015\)](#) and [CDRH guidance \(2015\)](#), there was a need to better understand this concept, which the PREFER team further elaborate in this document. In essence, in certain situations decision-makers may feel the need to better understand what matters to patients (e.g. what their primary needs are or what clinical endpoints are important to them), how much these matter to patients (e.g. what the relative importance of clinical endpoints is to them), what acceptable trade-offs are for patients (e.g. how benefits and risks are weighed) and how much uncertainty patients accept (patients' level of tolerance for uncertainty). Not all patients have the same preferences in this respect, so assessing and addressing heterogeneity requires dedicated consideration as explained further in this document.

The lack of guidance for designing, conducting, organising and using patient preference studies as well as communicating results was one of the main concerns revealed by stakeholders during PREFER's initial research ([De Bekker-Grob EW, 2018](#); [Janssens R, Russo S, 2019](#)). Expectations of stakeholders (industry, regulators, HTA bodies and payers, physicians, academics and patients) about the feasibility of generating and using patient preference information, as well as expectations regarding the impact of patient preferences in decision-making, is likely to depend on how, by whom and when the patient preference data are generated. Stakeholders' main concerns were the:

- lack of a clear, practical framework for defining: the research question for a patient preference study; the organisation (team, timing); and the design, conduct and application of patient preference information to inform decision-making
- Unknown level of patient knowledge and how to ensure comprehension among participants
- lack of methodological understanding and familiarity by all stakeholders of the selection and use of fit-for-purpose preference methods
- lack of clarity on how stakeholder interaction could occur during the design and conduct of a patient preference study, as well as in the adoption of patient preference study results (i.e. communication concerns)

A clear, dedicated framework for patient preference studies is therefore key to implement patient preference information in medical decision-making. Such a framework should include points to consider when selecting high quality and fit-for-purpose methodologies to use in a patient preference study, as well as guidance on stakeholder interaction.



## 1.2 PREFER objective and objectives of the qualification

The objective of PREFER is to guide industry, regulatory authorities and HTA bodies and reimbursement agencies on how patient preferences can be assessed and used to inform medical product decision making by developing expert and evidence-based recommendations.

PREFER would like this qualification opinion to provide confirmation that patient preference studies can provide an acceptable form of evidence to inform HTA bodies and Regulatory Agencies decisions in patient preference-sensitive situations.

## 1.3 Context of use

The proposed framework supports the development and specification of study purpose (objectives) and the design, conduct, analysis and reporting of patient preference studies to inform decision-making by industry, regulators, HTA bodies and payers.

The 'Points to Consider' for method selection, together with additional details of five key quantitative methods (while recognizing that further methods are available), provide an appropriate approach for selecting preference method that can be applied in conjunction with the framework.

See [section 2 figure 2-1](#) for an overview of how PREFER consider that preference data can be used to inform medical product decision-making.

## 1.4 Out-of-scope of this qualification

### Patient perspective data other than patient preference data

The collection and use of patient perspective data other than qualitative and quantitative patient preference data is out of scope.

Patient preferences are a particular type of patient perspectives as illustrated in [Figure 1-1](#).

### Healthcare professional preferences

The description of the PREFER framework focuses on how it could be used to collect patients' (and potentially caregivers') views. The PREFER framework could be used to design, conduct and analyse a study of healthcare professionals' (HCPs) preferences. While this information could add value and provide further evidence to inform regulatory decision-making, considering the specific issues related to HCP preferences is beyond the scope of the proposed framework.

The importance of the patient's voice in drug development, decision-making and drug use is clearly acknowledged, and the main objective of the PREFER framework is to offer a tool to support the inclusion of patients' views so they can be better included as part of the evidence supporting regulatory decision-making. While HCP preferences can certainly add an important and potentially different perspective, PREFER does not advocate that patient preference studies should always be accompanied by parallel HCP preference studies.

### Types of decisions not covered

- **Shared decision-making 'at the bedside'** between an individual patient and physician
  - Exploring or eliciting preferences from a single individual (e.g. in an outpatient setting during shared decision-making between a prescriber and patient) is out-of-scope of

the Briefing Book. The framework focuses on patient preference studies with the aim of obtaining preferences from a representative sample of a specific population of patients.

- Although the framework is not focused on the use of preference data in shared decision-making, the preference data obtained within such framework can help inform shared decision-making.
- Additionally, many insights gained from patient preferences demonstrate that using the framework may apply to shared decision-making (e.g. correlations between preferences and easily measured psychological, demographic or other variables may enable physicians to quickly estimate a patient's preference at point of care and then make treatment decisions accordingly). This will be facilitated by the development of a lay version description of the framework useful for communication to patients and other relevant stakeholders.
- **HTA and payers' decisions based on cost-per-QALY calculations**, obtained using public preferences for generic health outcomes.
  - Although quality-adjusted life years (QALYs) can in principle be calculated based on patient preferences (using a utility measurement method), they are usually based on public preferences, in order to take societal preferences into account in resource allocation decisions in healthcare as a whole. The rationale for using societal preferences for such decisions is that the general public finances healthcare as a taxpayer and hence its preferences should matter in the decision-making process.
  - Moreover, for system-wide decisions, generic health outcome measures enable comparisons between different diseases.
  - PREFER recognizes the importance health economic evaluations performed from the societal perspective using generic outcome measures to many HTA agencies and payers; however, PREFER does not see these as topics that PREFER would be able to usefully cover within the qualification.
- Use of preference data for developing **clinical practice guidelines**
  - Although PREFER will not develop recommendations regarding the use of preference data in developing or revising clinical practice guidelines, the preference data obtained within the scope of PREFER can be used to inform clinical practice guidelines.

## 1.5 Overview of Briefing Book content

The content of the briefing book is based on existing research in the field of patient preference studies as well as research and preference study experience generated during the PREFER project. The document consists of 6 sections:

- **Section 1** serves as introduction and details the **objectives** of the qualification as well as the **context of use**.
- **Section 2** of the briefing book covers two aspects. The first section covers the **value of patient preference studies**. An overview of the value of patient preference studies,

situations when patient preference data are most likely to be useful to support decision-making (patient preference sensitive situations) or less likely to be useful, MPLC phases when patient preference data are most likely to be useful to support decision-making, and an explanation about what fit-for-purpose patient preferences are. In the second section, **stakeholder involvement in patient preference studies** is discussed. Advice is given for working with key stakeholders (patient as partners in patient preference studies, regulators, HTA bodies and payers) on patient preference studies, including why is it important to work with stakeholders, as well as how, when and whom to involve.

- **Section 3** covers a suggested **framework** for patient preference studies that:
  - guides a preference study research team through key issues when designing, conducting and applying the results of a preference study
  - guides decision-makers when assessing and using preference study results to inform decision-making
  - supports the discussion between industry, regulators, HTA bodies and payers about preference studies intended to inform medical product decision-making
- **Section 4** outlines **patient preference study methods**, including elicitation (quantitative) preference methods.
- **Section 5** details **Points to Consider for method selection**, which are outlining selection criteria, to allow users to make an informed choice for method selection.
- **Section 6** concludes this briefing book with a discussion and a call to action for further research in the field of patient preference studies.
- **Annexes:**
  - Case studies and results abstracts are described in further detail in [section 1.6](#) below.

## 1.6 Data supporting the qualification

To support this qualification procedure, PREFER established evidence from literature reviews, interviews, focus groups and a spectrum of patient preference studies (referred to as 'case studies'), the latter of which support different aspects of the qualification procedure. The first aspect considered in the supporting evidence relates to the feasibility and robustness of the framework (e.g. as it relates to the planning, conduct, analysis and reporting of preference studies and their results). The second and equally important aspect is to answer methodological questions that are perceived as important gaps that may limit the utility of preference studies to support decision making.

Different types of studies support this qualification, namely:

- **PREFER core preference studies** in rheumatoid arthritis (RA), neuromuscular disorders (NMD) and lung cancer (LC) which respectively are designed to specifically address both clinical and methodological issues. (see Annex I, [section 8.1.1](#)).
- **Additional case studies** conducted by academic and industry partners which are designed in a way to address at least one important methodological question identified in the PREFER project and beyond (see Annex I, [section 8.1.2](#) and [section 8.1.3](#)).

- The inventory of **published and unpublished historical case studies** built by PREFER, a subset of which will serve as supporting information for the qualification procedure (see [Annex II](#)).
- Evidence from **literature reviews, interviews, focus group discussions, expert opinions**

## 1.7 Publication and transparency of patient preference study results

Since the start of the project, PREFER has recognized the need for data transparency of patient preference studies, as this has been widely discussed and agreed for other study types (e.g. clinical trials, post-authorisation safety studies). All results generated by PREFER from literature reviews, interviews or focus group discussions are published in open-access journals, presented at conferences, made available on the PREFER public website and presented to different audiences via webinars or lectures. PREFER has posted all prospective preference studies within PREFER (n=11) in the open access [Health Preference Study and Technology Registry \(HPSTR\)](#) , which is maintained by the International Academy of Health Preferences Research (IAHPR) for reasons of transparency. The core case studies are already registered in this registry – other registries have been considered but in general are less suitable than HPSTR for patient preference studies. When patient preference studies are conducted alongside clinical trials, registration in EudraCT or clinicaltrials.gov or similar registries is recommended.

In conclusion, PREFER believes that the concepts laid out in the framework section 3 below and its “points to consider” for method selection will promote the use of patient preference studies as a tool by which patients’ preferences can be used to inform decision-making.

## 2 Value of patient preferences and interface with stakeholders

### 2.1 Situations when patient preference data are most likely to be useful to support decision-making (patient preference sensitive decisions)

Patient preferences are a source of evidence that may inform particular decisions, where the perspective of the patients is particularly important and is expected to have an influence on the decision.

Patient preference information adds the most value when a complicated decision is required and depends on the patient's perspective about the most important benefits to them and what benefit-risk trade-offs are acceptable (Postmus D , 2018). This contrasts with straightforward, clear-cut decisions (as explained in Section 2.3); for example, when treatment is better on all relevant outcomes and entails lower risk, or treatment is better and cheaper than the comparator.

According to the FDA (CDRH, 2016), decisions are preference sensitive when:

1. multiple treatment options exist and there is no option that is clearly superior for all patients;
2. the evidence supporting one option over others is considerably uncertain or variable; and/or
3. patients' views about the most important benefits and acceptable risks of a technology vary considerably within a population or differ from those of healthcare professionals.

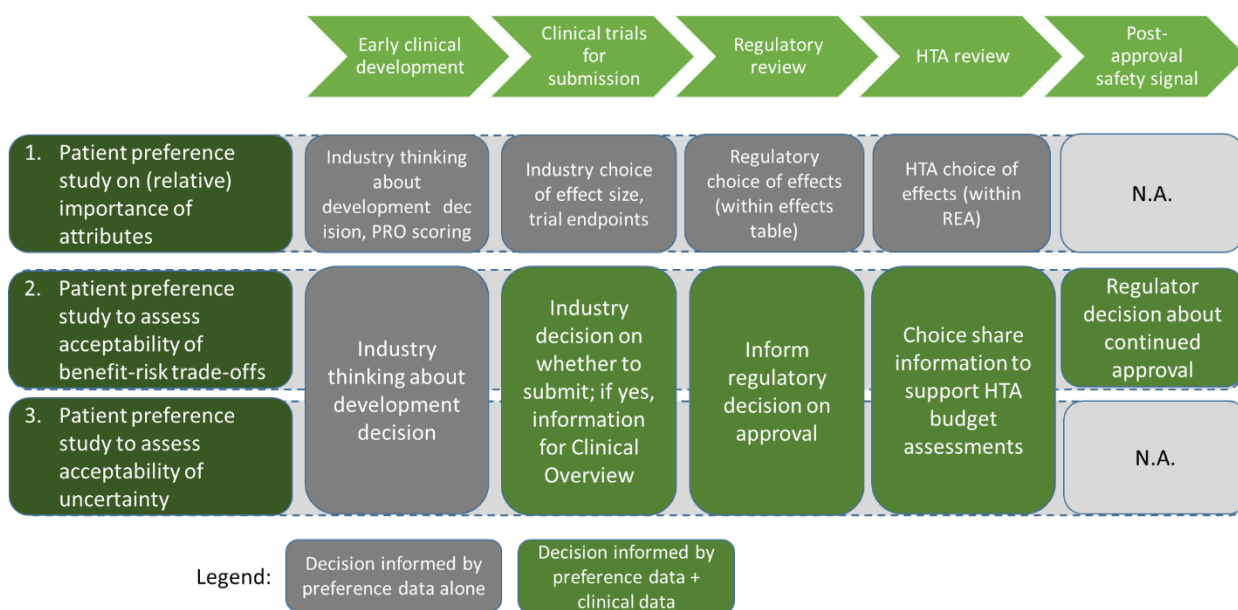
Building on the FDA's definition, PREFER has proposed the following categories of patient preference sensitive decisions to provide additional clarity.

What situations are patient preference sensitive?

1. When a preference study provides patients' views (i.e. patient relevant trial outcome parameter(s)) about the most important attributes (e.g. benefits, risks, mode of administration) of a specific disease/medical product.
2. When multiple treatment options exist (including the scenario of an active treatment vs. placebo or vs. standard of care) and there is no option that is clearly superior for all patients.
3. When an important aspect of the evidence supporting one option over other(s) is considerably uncertain or variable, and there is a need to understand patients' tolerance for uncertainty.
4. There is a need to investigate the heterogeneity in views among patients, or differences in views between patients and other stakeholders (e.g. HCPs), for any of the scenarios 1-3.

Figure 2-1 provides a mapping of preference-sensitive categories to stakeholder decisions throughout the medical product life cycle, that will be discussed below. Figure 2-2 shows an example of how a patient preference study can help the decision-making process relating to the acceptability of benefit-risk trade-offs.

**Figure 2-1 Mapping preference-sensitive scenarios to stakeholder decisions**



(Section 3.4.1 includes further information about applying preference data to inform the decisions described in Figure 2-1 above, and section 3.4.3 includes information on incorporating preference information in industry, regulatory and HTA documentation.)

In these preference-sensitive decisions, it is important to know more about the following scenarios:

- **What matters to patients** - which decision criteria/endpoints are important to patients? This usually results from qualitative research.
- **How much it matters to patients** – what is the relative importance of decision criteria/endpoints to patients? This results from quantitative preference studies.
- **The acceptability of trade-offs** – how do patients weigh benefits versus risk/harm and burdens. This might include the scenario where:
  - a choice must be made between different benefits (different hypothetical health states)
  - a choice must be made between one available treatment vs. no treatment (e.g. a treatment is available but has rare, serious, side-effects – some patients could reasonably choose to decline such a treatment, whereas others could reasonably choose to accept the treatment despite the side-effects)
  - a treatment is available that offers very moderate efficacy and has a very benign safety profile, and where it would be helpful for regulators to understand if the very moderate efficacy is something that patients think has value
  - the choice is between two very different treatment options (e.g. surgery vs. chronic treatment), or the choice is between treatment options with very different

profiles (e.g. drug A offers benefit X and side-effect Y; drug B offers benefit M and side-effect Z).

- **The acceptability of uncertainty** – what is the patients' tolerance for uncertainty; such uncertainty may relate to unknown long-term safety as well as efficacy.

## 2.2 Stages in the medical product life cycle where patient preference sensitive decisions may occur

Patient preference data may be useful in decision-making at different phases of the medicinal product life cycle, from drug discovery and development up to market authorisation, HTA and reimbursement and beyond e.g., post-marketing (see [Section 2.2.4](#)) ([Van Overbeeke E, 2019](#)). A more in-depth discussion is provided in [Section 3](#) (Framework).

### 2.2.1 Medical product discovery and development



#### 2.2.1.1 Understanding patients' views to inform decisions on which disease characteristics or endpoints to study

At the early drug development level, patient preferences may be useful in 'what matter scenarios' such as **identifying areas of unmet medical need** ([Selig W. K. D. 2016](#); [Johnson FR, 2016](#); [Cook, 2019](#)) (as exemplified in the PREFER COPD case study). Patient preferences can also inform the **content of a target product profile** ([Preference information guidance 2016](#); [Marsh K 2016](#); [Stewart K, 2016](#); [Mühlbacher AC 2016](#)).

Understanding patients' views on which disease characteristics matter to them (e.g. which organs are affected and how), can also inform decisions about which medicinal products to develop, based on whether or not a particular product is having an effect on those disease characteristics that patients consider to be important ([Selig W. K. D. , 2016](#)).

Qualitative patient research is often conducted to identify what is (most) important to patients, with PPS conducted to determine how much (relative importance) it matters. At the stage of clinical trial design, patient preferences may be useful in 'how much does it matter' scenarios, such as indicating which clinical endpoints are of highest importance to patients ([Preference information guidance 2016](#)), and thus helping to ensure their inclusion. PPS can therefore be instrumental in deciding what is (most) important to patients to incorporate in clinical trial design, and PROs are one of the ways in which evidence is then captured in addition to clinical endpoints and even digital evidence captured in trials. Patient preferences may also provide support in these 'how much does it matter' scenarios by helping to quantify the relative importance of individual clinical outcomes ([Preference information guidance 2016](#); [Smith M 2016](#); [Evers P, 2016](#); [IQWiG 2014](#); [Stamuli E, 2017](#)). They help ensure drug development activities or plans for a particular treatment include more patient-relevant outcomes and hence may enhance compliance ([Janssens 2019](#)). At earlier stages of clinical

development when preference studies are conducted to inform patient relevant endpoints in the trials, the risks of a new product are typically not known. Thus, such preference studies may focus on the benefits only (and trade-offs among different benefit profiles).

### 2.2.1.2 Understanding patients' views on non-health benefits

Patient preference studies may reveal important information related to the context wherein treatments are administered or information is provided. Non-health benefits may include convenience and mode of administration. One example of a benefit-risk trade-off situation might be deciding between a medical treatment given once a week with moderate efficacy vs. a drug given twice daily but with good efficacy. Another example might be if a situation with two drugs with similar benefit-risk profiles, but one drug is given as a daily pill and the other is given as a monthly injection.

### 2.2.1.3 Understanding patients' views on the acceptability of benefit-risk trade-offs and acceptability of uncertainty

At the drug discovery and development level, patient preferences about acceptability of benefit-risk trade-offs can inform go/no go decisions ([Ho MP 2016](#), [Van Til, 2014](#)) i.e. internal prioritization portfolio decisions. Of note, the benefits considered within a benefit-risk trade-off could non-health benefits (e.g. ease of use).

Patient preference information may be useful in 'acceptability of trade-off' scenarios and for understanding patients' views on acceptable levels of uncertainty (see description in [section 3.4.1.5](#) of potential sources of uncertainty) or defining subgroups with different benefit-risk trade-offs ([MDIC 2015](#); [Ho 2016](#)) (both scenarios are exemplified by the PAVING study).

### 2.2.1.4 Using patient preferences to inform PRO scoring

Patient preferences to inform PRO scoring will be detailed in Chapter 3 [section 3.4.1.3](#).

## 2.2.2 Marketing authorisation



In a marketing authorisation context, considering patient preferences may lead to more patient-centric decisions and enhanced transparency ([Postmus D, 2016](#)).

Inclusion of patient preference information in a Clinical Overview is supported by the [EMA, ICH guideline M4E\(R2\)](#) guideline. This states that "Information about the patient perspective may be considered when describing the therapeutic context, benefits, risks, and the benefit-risk assessment" and explains that "Patient perspective information describes the attitudes and preferences of patients with respect to the therapeutic context, benefits, and risks".



Product labelling may also be informed by patient preferences ([Patient preference information guidance 2016](#); [MDIC 2015](#); [Ho MP, 2016](#); [MDIC 2019](#)) as exemplified by the FDA labelling of rituximab.

### **How choice of patient-relevant endpoints can inform regulatory decision-making**

In the situation where patient preferences have been used to inform decisions about which endpoints to study, these (presumably patient-relevant) endpoints can in turn inform regulatory decision-making. As outlined in the [section 3.4.3](#) about selecting key favourable effects within the EMA Day 80 assessment report template, the key favourable effects will often be the primary efficacy endpoint and key secondary endpoint. Hence, when clinical study endpoints have been selected based on patient preference data, the use of primary and key secondary endpoints as the 'key favourable effects' would represent the indirect influence of this data.

### **Understanding patients' views on acceptability of benefit-risk trade-offs, acceptability of uncertainty**

The EMA advice to reviewers within the Day 80 assessment report template states that assessment of the acceptability of benefit-risk trade-off is subjective and that "patient input should be taken into account and explained, if available". Such input could be provided by the results of a patient preference study assessing patients' views on the acceptability of a benefit-risk trade-off, with the benefits including non-health benefits (as ICH M4E(R2) states: "Benefits may also include important characteristics of the medicinal product, such as convenience".)

### **Examples of preference studies that informed regulatory decisions**

In addition to the preference study conducted by [Postmus D \(2016\)](#), there have been numerous patient preference studies that have supported regulatory decisions at different stages of product development. Studies for which most information is available are summarised in **Table 2-1 Table 5-1** and further described below; the text below also mentions additional ongoing preference studies informing regulators' decisions (e.g. the FDA).

**Table 2-1 Known applications of patient preference studies**

Example application of patient preference study	Methodology used	Additional details
The <b>Tubes Under Local Anesthesia</b> (Tula) System study, to determine performance threshold for use as the primary endpoint.	Threshold technique	FDA recently approved a tympanostomy delivery system in which a patient preference study (FDA News Release , 2019; FDA Summary of Safety and Effectiveness data (SSED) Tula System, Tula System 2019), conducted using the threshold technique (TT), was used to determine the performance threshold for use as the primary endpoint in the pivotal clinical study for the procedure.
<b>Endpoint identification in COPD</b> , to understand patient-relevant symptoms.	Qualitative patient insights and quantitative discrete choice experiment	Novartis used a multi-phase approach (literature search, social media listening study, online bulletin board followed by quantitative patient preference study using Discrete Choice Experiment methodology; Patalano F 2020) to understand the symptoms that patients with chronic obstructive pulmonary disease (COPD) think are most important to treat. The National Institute for Health and Care Excellence in the UK provided scientific advice on the design of the study which may provide support for including additional patient-centered endpoints in clinical studies of COPD treatments (NICE, 2019).
<b>Maestro Rechargeable System for Obesity</b> , where the device’s approval was supported by the results of an FDA-sponsored patient preference study examining patients’ views on the relative importance of effectiveness, safety and other attributes of weight-loss devices.	Discrete Choice Experiment	FDA used the results of an FDA-sponsored patient preference study (Ho MP, 2015), conducted using a discrete-choice experiment (DCE), to demonstrate that there is a subset of the obese population in the US who would regard the benefits of a device to treat obesity as outweighing its risks even though the device did not achieve its primary endpoint in a pivotal study. The patient preference information was cited as instrumental in the approval of the device (FDA News Release 2015).
<b>Spavato Nasal Spray for Treatment Resistant Depression</b> , preference study to assess patients’ trade-off preferences for key benefits and harms associated with treatments for treatment-resistant depression.	Discrete choice experiment	Janssen presented the results of a patient-preference study, conducted using a discrete-choice experiment (DCE), to an FDA Advisory Committee as part of the FDA approval process (Janssen Research & Development 2019). Physicians on the committee indicated that the patient preference information helped them understand the patient voice (FDA, Joint Meeting of the Psychopharmacologic Advisory Committee (PDAC) and the Drug Safety and Risk Management Advisory Committee , 2019).
<b>NxStage Solo for Home Haemodialysis</b> , study to understand patients’ views on the acceptability of benefit-risk trade-offs.	Threshold technique	NxStage and a patient advocacy group conducted a patient preference study using the threshold technique (TT) to determine the maximum acceptable risk of death resulting from needle dislodgement that patients would accept to have home hemodialysis (Medical Device Innovation Consortium, MDICx Series, NxStage). The results of the study were used to support an expansion in the indications for use to allow patients to use home hemodialysis without having a care partner present (FDA, Approval letter for NXStage System One, 2017).

<p><b>Dexcom G5 Continuous Glucose Monitoring</b>, survey to understand the concerns of patients and parents about the safety of an insulin pump.</p>	<p>Qualitative study</p>	<p>The US FDA conducted qualitative interviews with patients and parents to understand their perspectives on the safety of using an insulin pump (FDA CDRH Patient Engagement). As a result of this research, the FDA and the company developed risk mitigation measures designed to prevent unintended insulin boluses.</p>
<p><b>Rituxan HYCELA Labeling</b>, to understand patient views on intravenous compared to subcutaneous formulation. The same approach was taken for Herceptin: <a href="https://pubmed.ncbi.nlm.nih.gov/28648618/">https://pubmed.ncbi.nlm.nih.gov/28648618/</a></p>	<p>Patient preference questionnaire in an interventional study</p>	<p>Genentech commissioned an interventional patient preference study using a crossover trial design to compare subcutaneous to intravenous administration of rituximab in blood cancers (Rummel M, 2017). The primary endpoint was the proportion of patients who preferred the subcutaneous formulation to the intravenous formulation. The results of this patient preference trial were included in the Patient Experiment section of the product label (Rituxan Hycela prescribing information2017).</p>

In addition, there are other ongoing known applications of preference studies informing regulators' decision (e.g. FDA). The Center for Devices and Radiological Health (CDRH) mentions collaborations in the area of medical devices for obesity (mentioned in the above list), medical devices for Parkinson's disease, medical devices for amputation and minimally invasive glaucoma surgical devices (FDA CDRH Patient Engagement; FDA Advancing Regulatory Science, Comparing Qualitative and Quantitative Approaches to Eliciting Patient Preference: A Case Study on Innovative Upper Limb Prosthesis; FDA Advancing Regulatory Science, Patient and Provider Views on Clinical Endpoints: A Qualitative Preference Study of Minimally Invasive Glaucoma Surgical (MIGS) Devices).

**Known applications of preference studies by CBER:** The 2018 FDA director's report ([CBER FY 2018 Report](#)) mentions that the Center for Biologics Evaluation and Research (CBER) launched three patient preference studies for prospective CBER products in the disease areas of osteoarthritis, sickle cell disease and brittle diabetes.

### 2.2.3 HTA & Reimbursement decision-making



**Known applications of preference studies by Health technology assessments (HTAs)** support reimbursement decision-making by collecting the best available scientific evidence on a disease and its treatments. This scientific evidence-base is the cornerstone of HTA. Society views patients' perspectives as important for decision-making, as the ultimate aim of decision-making in healthcare is to improve patients' health. HTA agencies are exploring different ways of taking patients' perspectives into account, with some examples described in [Facey K \(2017\)](#).

Patients' preferences may provide insights for health economic evaluations (e.g. cost-utility analyses, where utility is based on data elicited directly from patients) ([Mühlbacher 2017](#), [Craig 2017](#), [Mott 2016](#)) contribute to prioritization of topics for HTA ([Danner 2011](#)), inform heterogeneity and segments of the patient population ([Mühlbacher 2015](#), [Craig 2017](#)) or tailor reimbursement decisions based upon patient heterogeneity ([Mühlbacher 2015](#)).

#### **How choice of patient-relevant endpoints can inform HTA and reimbursement decision-making**

HTA bodies use country-specific criteria to assess the value of new health technologies, including new medicinal treatments. One criterion that is relevant in any jurisdiction is the relative effectiveness of the new technology, compared to the current standard of care or other treatments used for a specific indication. For the relative effectiveness assessment (REA) of a new treatment, it is important to know patients' perspectives on the aspects of their health or current treatment strategy they want to see improved. Therefore, a first step in the assessment process is to identify the therapeutic needs of patients, which are indication-specific but not product-specific. EUnetHTA has established recommendations – Patient Input in Relative Effectiveness Assessments" – although these currently only focus on qualitative approaches to patient engagement during the scoping phase

of the HTA, which can be done by means of a qualitative patient preference study. Such a study can address the ‘what matters’ question and help to identify the endpoints to be assessed in the REA part of HTA. PREFER’s claim is that quantitative patient preferences could complement qualitative patient input.

Selecting, prioritizing or weighing endpoints and criteria is another value offered by patient preferences (Mühlbacher 2016, Mühlbacher 2017, Marsh 2017). Patient preference data are hence complementary to clinical evidence and cost-per-QALY estimates based on public preferences (Bouvy 2020). A major practical limitation is that traditional preference approaches do not have a way to be mapped to the cost-per-QALY approach used in HTA (Huls, 2019).

**Understanding patients’ views on non-health benefits**

At the HTA and reimbursement level, patient preferences are valuable in ‘how much does it matter’ scenarios. Besides the relative importance of ‘hard’ clinical outcomes, non-health benefits (e.g. mode of administration) could also be weighted in a relative effectiveness assessment based on patient preferences (Mühlbacher 2015, Gutknecht 2016).

**Understanding patients’ views on acceptability of benefit-risk trade-offs, acceptability of uncertainty:**

Patient preferences may be valuable when examining relative benefit-risk trade-offs (Mühlbacher 2016, Mott 2016), predicting uptake rates (Mott 2016) and indicating the general acceptability of a technology to patients (MDIC 2015, Gutknecht 2016, Brooker 2013).

Effectiveness data on primary outcomes, such as overall survival, are not always available, sufficient or convincing at the time of a HTA. For instance, for a disease like prostate cancer, with good overall survival rates and a relative favourable life expectancy, follow-up time to show relevant primary outcomes will be longer than the average clinical trial duration. In the absence of such long-term follow-up data, it may be important to know how patients value and weigh quality of life benefits in the short term compared to long-term survival benefits. In this context, patient preferences can help to bridge the uncertainty that is inherent in the long-term extrapolation of clinical trial outcomes and might help make temporary reimbursement decisions until long-term data are available.

Besides the aforementioned benefits of patient preferences for HTAs, patient preference data can also help to predict uptake rates of a new treatment (i.e. choice share), which might be important for estimating the expected budget impact of using a new treatment. Suppose that, based on a patient preference study, it is shown that about 70% of patients would prefer the new treatment, then the budget impact should be calculated on 70% of the eligible population.

**2.2.4 Post-marketing**



At the level of post-marketing, patient preference studies may inform ‘**acceptability of trade-off scenario’s** and risk assessments underlying product recalls. Planning and evaluating BRAs and risk

management may also be inspired by patient preferences (Smith M 2016). Products' benefit-risk profile is assessed at regular intervals in a PSUR and in label updates. If new safety information becomes available in the post-marketing phase, then it could be relevant to assess patients' views on the product's benefit-risk profile (in light of this new safety information).

### 2.2.5 Additional notes

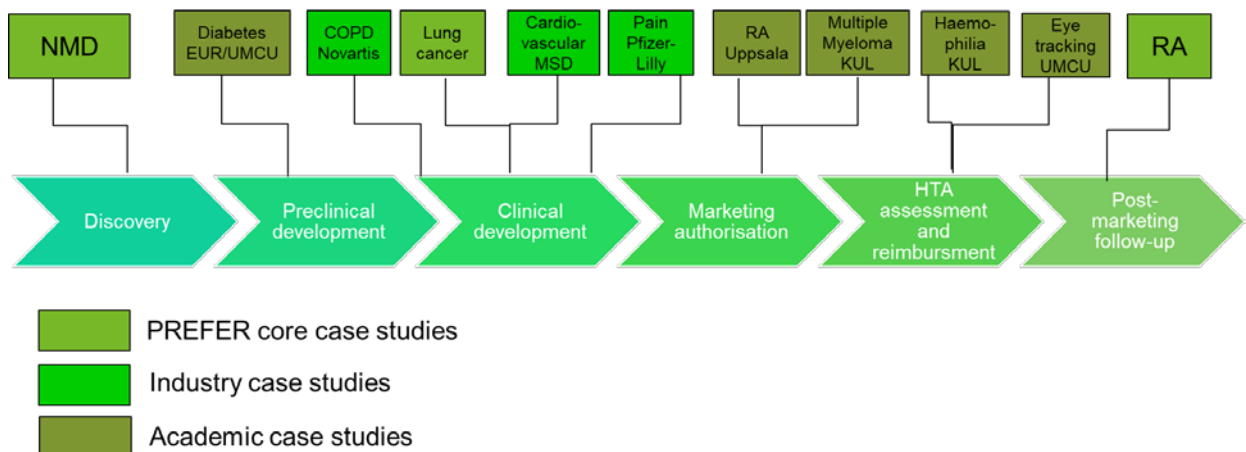
With the proposed framework, PREFER takes the first steps in trying to create scientific approaches to patient preference studies that can be assessed in terms of their validity and reliability. The aim is to provide patient-based evidence that is useful to clinical trial sponsors for their development considerations, regulators for their benefit-risk decision making, and HTA bodies and payers for their assessments and appraisals. The current situation is that HTA bodies and payers want to consider patient preferences but do not do it systematically because:

1. they are not familiar with patient preference studies they do not know when a patient preference study is likely to be valuable for decision-making, and do not know how to critically assess the quality of a patient preference study (e.g. identifying potential bias or methodological errors)
2. consideration of patient preference studies is not systematically embedded in the assessment processes.

In addition, no legal framework currently exists that allows HTA bodies to incorporate patient preferences in decision-making. This can only change if the assessment criteria for a patient preference study are made explicit, which is what PREFER's final recommendations and its framework aim to do. Gaps need to be addressed in order to advance the measurement and use of patient preferences in all these decisions. More specifically, there needs to be a focus on providing guidance on when and how to measure patient preferences to inform these stakeholder decisions and increase stakeholder familiarity with performing and evaluating patient preference studies.

PREFER conducted preferences studies throughout all stages of the medical product lifecycle (MPLC) as shown in Figure 2-2 below.

**Figure 2-2 PREFER cases studies throughout all MPLC stages**



**Equity:**

PREFER acknowledges the potential impact of using patient preferences in decision-making on health equity, due to the fact that patient preferences data will be available for some but not all patient populations. The intention of the qualification is to promote the broader use of patient preference studies in the research, development and evaluation of novel treatments (in situations where such data is expected to add value to the decision-making process) and to improve the quality of information about patient preferences available to decision-makers. This should – over time – increase the availability of preference data in value dossiers, hence leading longer-term to a situation where there is greater equity and more patient-relevant information incorporated into the decision-making process.

This type of disparity is a problem inherent in patient engagement as there are differences between patient groups in the level of advocacy for different diseases. However, in contrast to other forms of patient engagements, where the existence of patient organisations is often important, patient preference studies can be conducted in diseases that are not represented by a patient association. Furthermore, patient preference studies are a way of collecting input from a representative sample of the patient population.

With the proposed framework, PREFER takes the first steps in trying to create scientific approaches to patient preference studies that can be assessed in terms of their validity and reliability in order to provide patient-based evidence that is useful to clinical trial sponsors for their development considerations, regulators for their benefit/risk decision making, HTA bodies and payers for their assessments and appraisals.

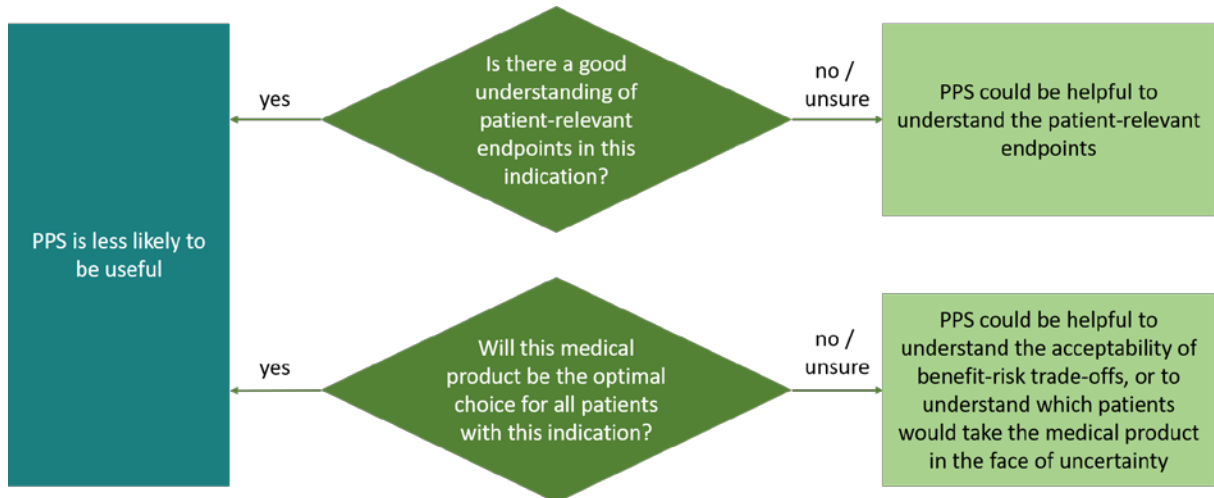
### **2.3 Situations when patient preference data is less likely to be useful to support decision-making**

Specific situations were identified in the literature ([including the MDIC 2015 report](#)) and in interviews conducted by PREFER in which patient preference data might not (always) be useful for decision-making. [Janssens et al \(2019\)](#) states that “some academics and regulators stated that patient preference studies only need to be conducted when they are relevant for decision making.”

Patient preferences are less likely to add value when the disease state, technology (medicine), study design, and clinical inputs and acceptable benefit-risk trade-offs are generally understood by both sponsors and regulatory staff, or when there is significant regulatory precedent for approval. Other examples include when a new treatment is clearly superior to existing therapies (more benefits, fewer risks and, from the HTA/payer perspective, less costly) or when the treatment addresses an unmet medical need in a disease with poor outcomes such that the risks of the treatment will not be greater than the risks of the untreated disease. Furthermore, certain decisions might not be sensitive to patient preferences when other (clinical) evidence is sufficient to support decision-making, or when preferences of other stakeholders (general public, physicians) or other evidence, are considered to outweigh the importance of the preferences of patients. An example might be some HTA decisions where the preferences of the general public should inform the decision ([Mott 2018](#), [Diksen, 2014](#), [Janssen 2016](#)).

[Figure 2-3](#) outlines the questions that can assist when determining if patient preference studies might be useful and consequently when these could be less useful.

**Figure 2-3 Situations where patient preference studies might be useful and less useful**



Patient preference data are less likely to be useful in situations when the patient is not a major decision maker or stakeholder. This may be the case in situations when it is primarily the preferences of others, particularly providers that has most benefit in determining the use of a particular technology (e.g. a surgeon deciding to use one particular surgical tool over another. This type of decision-making is, in any case, out of scope of PREFER.

## 2.4 Interactions with patients as participants and partners on patient preference studies

### 2.4.1 Why is it important to interact with patients as research partners?

There are substantive and fundamental ethical and content-related rationales for involving patients as partners in patient preference studies:

1. patients have the moral right to be involved in research that concerns them directly or indirectly
2. researchers do not necessarily know what treatment features matter most to the patients
3. involving patients as partners increases the relevance, appropriateness, feasibility and acceptability of a patient preference study
4. patients can offer a unique perspective that differs from those of researchers, healthcare professionals or other experts, as they have experiential knowledge of the disease.

In addition, patient involvement in the patient preference study itself is complementary to scientific evidence on patient perspectives and experiences. It has a different purpose and hence the role of patient partners is different from that of patient participants



## 2.4.2 General principles for interactions with patients as participants in patient preference studies

Interactions with stakeholders during a patient preference study should take into account certain fundamental ethical, legal and social principles. Those principles come from established European, national and institutional (legal/ethical) frameworks, and were refined in other IMI projects. Below we list, adapt and define the most important stakeholder interaction principles that are applicable to patients as participants in preference studies, to interactions with patients as partners in patient preference studies (see [Section 2.4.4](#)) and with other stakeholders ([Section 2.5](#)) (<https://imi-paradigm.eu/petoolbox/>).

### Ten principles for interaction with patients

1. **Patient centricity:** During all activities performed as part of a patient preference study, it is essential to consider how patients can be put at the center of these activities, hence to assess whether, how, when and which patients can/should be involved.
2. **Clear communication and transparency:** Information should be provided to people so that participation can occur in a meaningful and trustful manner. People may express conflicting opinions, in which case options to address these will be discussed, and any constructive dialogue will be encouraged. All materials and communication activities need to be clear and tailored to the needs of each stakeholder group to ensure proper understanding.
3. **Inclusiveness:** The diversity of the specific patient population (e.g. vulnerable persons and those not easy to reach) should be represented at all levels of a patient preference study
4. **Responsive and reciprocal:** We aim to make each interaction meaningful and interesting for the researchers as well as the participants and end-users, by listening to input and providing feedback.
5. **Respectful:** All input received from participants or partners (e.g. medical knowledge, policy information, health outcomes) will be treated with respect. Individual rights, privacy and confidentiality will be respected.
6. **Well-prepared:** Interactions will be purposeful, well-organized and clearly indicate how the input of people will be used at the start of every stakeholder interaction.
7. **Objective:** All activities and exchange of information must be done in a neutral, balanced and objective manner, free of conflicting interests.
8. **Proportionate:** All stakeholder interaction efforts (time, burden, etc.) should be proportionate and reasonable, and estimated as much as possible in advance of the interaction.
9. **Non-interference with current health care:** There should be no impact on the relationship between the participant and his/her health care professional as a result of the patient's interaction in the preference study.
10. **Impactful and sustainable:** Each stakeholder interaction should be as beneficial and impactful as much as possible e.g. for stakeholders and society as a whole. This will also help achieve an ongoing, sustainable level of stakeholder engagement throughout preference research.

### 2.4.3 Interactions with patients as research partners in patient preference studies

Patients should be involved as active research partners in a patient preference study, not just as participants. Prefer defines the term ‘patient’ to also include proxies for patients, such as parents and care-givers, if patients are unable, to express their own preferences (e.g. due to age, cognitive capabilities)

In addition in some cases, the ‘patient’ may in fact be a healthy member of the public (for example in the context of a vaccine programme) or an ‘at risk individual’ (for example in the context of another type of preventive medicine programme). In such cases, the recipients of the proposed therapy may not view themselves “patients”; sensitivity regarding terminology is important.

#### 2.4.3.1 How to involve patients as partners in a patient preferences study

Patient involvement in a patient preference study means establishing a partnership with researchers by becoming a study team member or by taking an advisory role (e.g. on an advisory board). Within this partnership, patients can contribute to decision-making about the study before, during and after the research process. Examples of patient inclusion in PREFER case studies is show in [Table 2-2](#).

**Table 2-2 Examples of patient inclusion in prospective PREFER case studies**

Study	Role of patient	Details/Comments
<b>PREFER Multiple Myeloma (MM)</b>	Study team member	<ul style="list-style-type: none"> <li>involved in study protocol development</li> <li>co-investigators in the ethics committee protocol</li> </ul>
<b>PREFER PAVING</b>	Advisory Board member	<ul style="list-style-type: none"> <li>advisor for protocol development</li> </ul>
<b>COPD study</b>	Advisors/consultants	<ul style="list-style-type: none"> <li>advisor on study design, discussion guides and survey questionnaires</li> <li>involved study results interpretation</li> </ul>
<b>PREFER RA</b>	Study Team member	<ul style="list-style-type: none"> <li>decision making about the clinical objectives</li> <li>design of recruitment procedures, content of focus group guide and survey instrument</li> <li>prioritization of attributes and survey pre-testing</li> <li>involved in study results interpretation and dissemination</li> </ul>
<b>NMD study</b>	Transitioned from Advisory Board member to study team member	<ul style="list-style-type: none"> <li>involved in study protocol development</li> <li>involved study results interpretation</li> </ul>
<b>Lung Cancer</b>	Consultants	<ul style="list-style-type: none"> <li>involved in qualitative and quantitative study protocol development</li> </ul>

The intensity and level of the patient partnership can vary from targeted or embedded consultation to co-producing the patient preference study. In a targeted consultation, patients are consulted on specific questions or study aspects on an ad hoc basis. Embedded consultation means that patients are regularly consulted throughout the entire research process. If patients are involved as co-

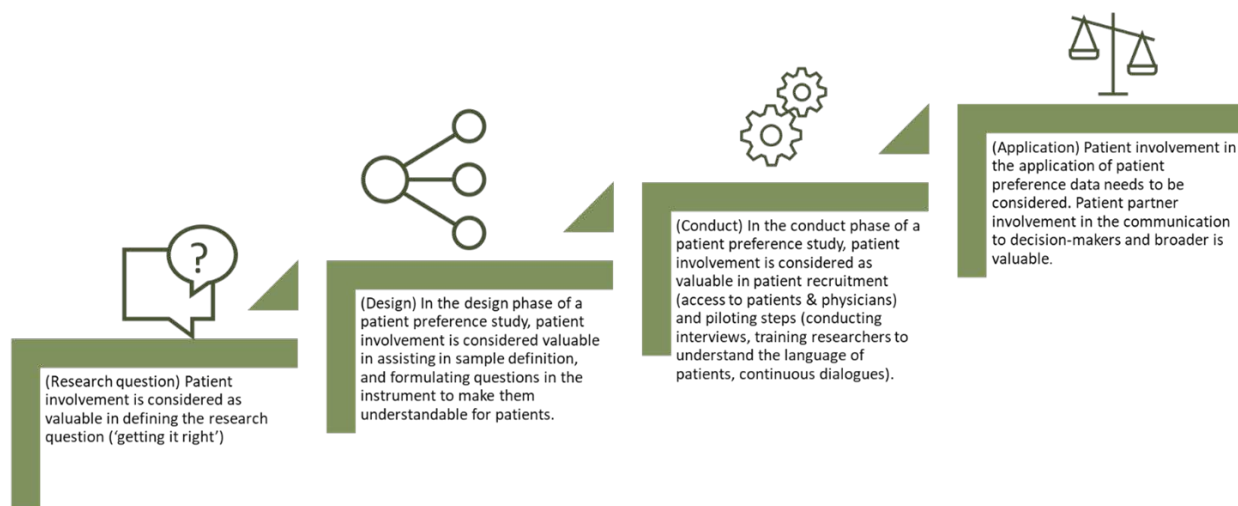
producers of research, they make decisions in collaboration with the research team and carry the same responsibility for the decisions made.

**In the PREFER RA case** the level of involvement in study design varied according to the stage of the study, and was most intense during the development of the focus group schedule, analysis of focus group findings, attribute selection and development of the survey instrument, and survey pre-testing.

### 2.4.3.2 When to involve patients: Preference study activities with patients

Patients can be involved as partners throughout the different steps of a patient preference study, as illustrated in the framework (see Section 3). These steps are described in Figure 2-4 below. Table 2-3 lists examples of how patients were involved at all stages of the RA PREFER case study.

**Figure 2-4 How patients can be involved at all stages of a preference study**



**(Research question):** Defining the research question is the initial step in a patient preference study. Close collaboration between patients and researchers during this phase of a patient preference study can help ensure that the research question is clearly phrased for it to be understandable or comprehensible to patients and addresses an issue that is not only of scientific and commercial interest, but of relevance to the targeted patient community itself. Pointing out ambiguities or difficult language that possibly could lead to future misunderstandings is another valuable part where patients can help. (Table 2-3).

**Table 2-3 Defining the study purpose**

Study	Problem	Context	Contributions
<b>PREFER MM study</b> <b>PREFER lung cancer case study</b>	Definition of the research question	Initially based on academic partners, later involved patients and clinicians	Patients helped to the improve the research questions
<b>PREFER RA case study</b>	Identification and development of the clinical research objectives	Patient research partners collaborated with the clinical research team	Patients co-developed the research objectives

**(Design):** When designing the patient preference study, patient involvement is considered valuable in helping to:

- define sample inclusion and exclusion criteria
- develop the data collection instruments (e.g. formulation of the questions, selecting and defining attributes and levels) to ensure they can be easily understood by patients and avoid any ambiguity and misunderstandings
- ensure that questions and answers are plausible, relevant and meaningful based on the patients' experience, all of which help improve the validity of the preference study results

Patient involvement is absolutely necessary at the level of survey presentation, namely in assessing the format/layout of the survey, mode of presentation (digital or paper-based), length of survey, consistency in wording and inclusion of general elements contributing to trust between patient and researcher. Tables 2-4, 2-5 and 2-6 show examples of how patients were involved in study design of PREFER case studies.

**Table 2-4 PREFER case study examples of patient involvement in formulating questions**

Study	Problem	Context	Contributions
<b>PREFER multiple myeloma study</b>	Initial proposed version of the swing weighting (SW) questions was too complex for patients	The questions and answer options were revised to make them more precise, clear and concise	Improvement of accuracy and understandability of the finalized focus group discussion questions and survey
<b>PREFER RA case study</b>	Public perceptions of RA are often inaccurate; patient research partners advised it was particularly important to convey the nature of the condition in detail to enable participants to make informed choices	The context of the choice task in this study was altered as a result of patient partner validation of the focus group study results	Inclusion of specific examples of ways in which the early symptoms of RA might impact patient's daily activities
<b>PREFER COPD study</b>	Suitability and patient-friendliness of the language and phraseology	The survey was reviewed by a patient network	Improvement of survey

**Table 2-5 PREFER case study examples of patient involvement in defining attributes and levels**

Study	Problem	Context	Contributions
<b>PREFER multiple myeloma study</b>	<ul style="list-style-type: none"> <li>• Patients highlighted the importance of considering the timeframe in which a certain negative treatment effect would take place</li> <li>• What levels should be included for the</li> </ul>	Levels were thoroughly discussed among the methodological, and clinical experts and patient partners	Patients, academic partners and physicians decided to include the levels 3 and 7 years, as these were both plausible and realistic

	attribute 'expected additional life years'		
<b>PREFER RA case study</b>	<ul style="list-style-type: none"> <li></li> </ul>	Patient research partners ranked the attributes identified in the qualitative focus groups in order of importance	Patients advised on the final selection and presentation of attributes used in the quantitative study

**Table 2-6 PREFER case study examples of patient involvement in survey presentation**

<b>Study</b>	<b>Problem</b>	<b>Context</b>	<b>Contributions</b>
<b>PREFER multiple myeloma study</b>	First version of this guide contained too many phases, which made the focus group discussion too long and burdensome for the patients	Patients highlighted that it was important to build patients' trust in the survey, to increase the likelihood for full completion of the survey	<ul style="list-style-type: none"> <li>The survey was shortened and simplified</li> <li>Visuals were combined with text, and wording was made consistent</li> <li>Patient research partners proposed specific text to include in the invitations and communication to patients, clarifying the study's purpose and benefits</li> </ul>

**(Conduct):** At the conduct phase of a patient preference study, patient involvement is considered valuable, e.g. at the level of patient recruitment (access to patients and physicians), conducting interviews, and training researchers to understand the language of patients, interpreting the results of a patient preference study and communicating results to patients. At the pilot, it is highly recommended that patient partners are involved via a think-aloud methodology and that those who were not involved in the study design should be included as participants, pilot interviewees, assistants to patients explaining questions, or as partners ensuring that suggestions are implemented.

For patient recruitment, patient partners should be supported with suitable study material. For the interpretation of results (after analysis), patient involvement is highly recommended for the clarification of nuances, to indicate whether the results make sense, or provide explanations on results. Such interpretations help to contextualise the data from the point of view of the patient and need to be combined with the viewpoint of other researchers. Interpretation of data is needed both at the qualitative and quantitative phase of the patient preference study.

In addition to the communication of results to stakeholders for use in decision-making, results should also be communicated back to patients themselves -- both those who participated in the study as well as non-participants who were part of the target patient population. Patient partners should be involved whenever there are plans to communicate to patients about the patient preference study. Dissemination of results should only begin once results have been reviewed and researchers are confident of their accuracy.

**(Application):** At the application phase of patient preference results, patient partner involvement is valuable in communicating study results to decision-makers and other audiences (e.g., the general public, journalists, etc.), but especially to patient participants.

Any effort to communicate with study participants should be guided by a communication plan, developed in collaboration with patient partners. The communication plan should detail who should

communicate the results, and when, how and what will be communicated. Communication plans should adhere to the following nine recommended steps:

**Recommendations for communication plans**

1. Give the participant general information before the start of the study
2. At the start of the study, repeat the most important information and gauge participants' expectations
3. Integrate patients' suggestions and opinions while drafting preliminary lay summary templates
4. Keep the participants up-to-date during the study but do not disclose unvalidated results
5. Communicate the due date of the closure of the study and the due date of the disclosure of the study results
6. Before the disclosure of the patient preference study results, give an overview of the process which has led to the final outcomes
7. Employ user-testing to evaluate the final version of the plain language summary in a group of laypeople or patients
8. Share the results of the patient preference study after the completion of the study report in a comprehensible and clear way or state where a lay summary of results will be made available to the public if study participants contact details are unknown.
9. Provide contact details to the participants in case they have further questions.

**Table 2-7 Example of patient involvement across the whole RA patient preference study**

<b>Study phase where patient partner were involved</b>	<b>Specific activity where patient partner input was needed</b>	<b>Relevant experience/expertise of patient partners</b>
<b>Definition of research question and target population</b>	Determine research objective and define research questions based on prior experience with RA project related to biomarker development.	<ul style="list-style-type: none"> <li>• Insight into patient priorities for rheumatology research</li> <li>• Involvement in a previous research study in a related area, in particular qualitative data analysis</li> </ul>
<b>Selection of methods and instruments</b>	<ul style="list-style-type: none"> <li>• Development of study protocol</li> <li>• Development of focus group schedule</li> <li>• Development of survey instrument</li> <li>• Development of study documents for participants</li> <li>• Development of disease background information and communication to study participants – this step is critical to inform treatment preferences, as participants do not have direct</li> </ul>	<ul style="list-style-type: none"> <li>• Experience of living with RA</li> <li>• Experience of RA treatment</li> <li>• Insight into public perceptions of RA</li> <li>• Experience of taking part in surveys</li> <li>• Representation from different European countries (UK, Germany, Netherlands, Sweden)</li> <li>• Experience of involvement in related cross-European</li> </ul>

Study phase where patient partner were involved	Specific activity where patient partner input was needed	Relevant experience/expertise of patient partners
<b>Patient recruitment</b>	experience of having the disease, and public perceptions of RA are often inaccurate (refs) <ul style="list-style-type: none"> <li>• Ensure study procedures and survey instrument are appropriate for use in different European countries, and appropriate for members of the public to complete</li> </ul> Ensure recruitment strategy takes into consideration the needs of the target population – first degree relatives were indirectly recruited via patients with RA, therefore recruitment procedures needed to appeal to, and be sensitive to, the needs of patients with RA.	projects (Birch et al 2020; De Souza et al 2020; Pollock et al 2017) <ul style="list-style-type: none"> <li>• Experience of living with RA</li> <li>• Experience of RA treatments</li> </ul>
<b>Piloting</b>	Participate in virtual survey pre-test session with researchers to share	Experience of completing surveys (not essential)
<b>Analyzing, interpreting and disseminating data</b>	Coding and interpretation of qualitative data.	<ul style="list-style-type: none"> <li>• Analysis of qualitative data</li> <li>• Experience of living with RA</li> <li>• Experience of RA treatments</li> </ul>

#### 2.4.4 Profile of patients as partners

Patients (or a care partner/caregiver of a patient living with disease) who are identified as patient partners in patient preference studies should collectively represent the diversity of perspectives and experiences of the target population. Key considerations may include (but are not limited to): diversity in experience with disease and treatment e.g. range in disease severity), demographics (e.g. age, sex/gender, race/ethnicity), sociocultural background, health literacy, experience with clinical trials for a similar disease, etc.

The study objectives and activities should determine the expertise and experiences needed from a patient partner, which may differ based on the type of input that is needed at different stages of the study. Such expertise and experience may range from being able to review documents, translating protocols, liaising with other patients (organisations), interviewing. This may require e.g. community outreach, using easy to understand language, an awareness of accessibility/inclusivity issues, and the provision of training for both patients and researchers where appropriate. In some cases, it may require patient partners with a combination of experiences.

**Example** (experience with research)

In the RA case study, some supporting patient research partners had previously worked with the study team on a systematic review of qualitative evidence and had received training in qualitative analysis and meta-synthesis. This enabled the patient partners to actively contribute to the coding and interpretation of focus group and interview data collected in the PREFER case study. The findings of this previous project also shaped the definition of the clinical objectives of the RA case study, in collaboration with patient partners.

### 2.4.4.1 Other important points to consider when working with patients

A patient preference study should be conducted by a multidisciplinary team, that has knowledge of the disease and patient preference study conduct and/or methodologies, possibly in partnership with industry. Additional information is provided in [Section 3](#).

#### **Example** (arrangements)

In the PREFER RA case study, patient research partners were reimbursed for their time in the form of shopping vouchers appropriate for their country (UK, Germany, Sweden, Netherlands), and were given information about implications of this reimbursement for tax or benefits status.

Researchers should allow flexible involvement whereby different patients individually take on different tasks. The location of appointments should also be taken into account for patients to be practically able to participate. To ensure researchers are able to identify and recruit patient partners with necessary experience and expertise, appropriate strategies should be adopted to enhance inclusivity.

To ensure researchers can identify and recruit patient partners with necessary experience and expertise, appropriate strategies should be adopted to enhance inclusivity. This includes:

1. use of easy to understand, non-technical language, but including glossaries of technical terms where required
2. clear and concise descriptions of what the patient partners' role will be
3. flexibility around meeting times, including out-of-office hours
4. outreach work to involve patient partners in community settings
5. opportunities to contribute remotely (e.g. via email, by attending teleconferences or video meetings)
6. use of easily accessible meeting venues (e.g. lifts/ramps, locations that are easy to travel to)
7. ensuring meetings are structured to accommodate patient partners' needs (e.g. frequent breaks, refreshments, lay summaries of presentations/documents, care givers can attend)
8. reimbursement of any expenses and payments for patient partner time
9. recaps of study background and objectives, progress, and impact of patient partner activities at regular intervals
10. allowing sufficient time for completion of involvement activities (no short deadlines)
11. no requirement for patient partners to sign or review lengthy, complex documents or legal agreements.
12. provision of training for patient partners if specific skills or knowledge are needed or desired to support meaningful involvement (e.g. to contribute to aspects of data analysis or study conduct, assertiveness skills to support participation in management meetings)



13. provision of training for researchers on public involvement, and where specific skills or knowledge are needed to enable them to effectively involve members of the public (e.g. communication skills, needs awareness, outreach training).
14. Provision of training for researchers on public involvement, and where specific skills or knowledge are needed to support them to involve members of the public effectively (communication skills, needs awareness, outreach training)

## **2.5 Interactions with regulators and HTA bodies on patient preference studies**

### **2.5.1 Why is it important and when is the best time to engage regulators and HTA bodies?**

If patient preference evidence is expected to improve the quality of decision-making at both the regulatory and HTA levels, it is important to consult with these decision makers at the inception of a patient preference study to ensure that it will deliver the results they need. Such early engagement will ensure patient preference increases the relevance, quality and legitimacy of the subsequent decisions.

Patient preference information can inform regulators to make difficult decisions, e.g., patient preference sensitive decision. The value and weight of patient preference information in decision-making may depend on several factors (e.g. the stage of the product lifecycle, the decision-making context, type of medical product, disease area, and unmet medical need). These factors are important to identify together with decision makers at the outset.

After mutual identification of the need for a patient preference study, its value and role in the decision-making, sponsors should seek advice on a potential inclusion of patient preference studies completing the overall evaluation dossier. Consultation with a decision-making body at the stage of designing a patient preference study can ensure mutual understanding of the study's design appropriateness and type of output, taking the criteria of 'fit for purpose' into account (see [section 4](#) and [section 5](#)).

In an HTA context there is a specific need to agree on the way patient preferences can fit within or alongside established assessment methodologies. Depending on the HTA rules in the country, patient preference data can or cannot be integrated in economic evaluations. In many jurisdictions, general public preferences are preferred instead of PP information, or a combination thereof, for the purposes of resource allocation decisions ([Van Overbeeke E 2019](#), [Huls 2019](#), [Marsh 2020](#)). Patient preferences go beyond preferences included in QALY framework or most other traditional frameworks (e.g., budget impact, added clinical benefit etc.). It is important to discuss beforehand with an HTA agency and/or a payer which requirements prevail and whether patient preference data can be used as a complementary source of evidence for the assessment of a new technology.

Regulators and HTA bodies should be consulted early in the PPS planning (e.g., scientific advice).

### 2.5.2 How should regulators / HTA bodies be consulted in preference studies

Most HTA and Regulators in key regulated environments have structured involvement processes that are exemplified here for EMA and EUnetHTA scientific advice processes. RA / HTA bodies include at times patients into the Scientific Advice process. Patient preference study methodology differs from that of clinical trials or observational studies and are more comparable to those of utility studies. Regulatory and HTA bodies should ensure appropriate expertise for interpreting patient preference study protocols or preference results for an effective collaboration on the study design.

PPS sponsors are encouraged to consult in scientific advice processes, as well as using convergence mechanisms such as FDA or EMA Scientific Advice procedures, EUnetHTA scientific advice (or a follow-up procedure thereof) for joint advice from EMA / HTA bodies. Most of these organisations have structured involvement processes that are exemplified here for EMA and EUnetHTA scientific advice processes.

### 2.5.3 What topics may be covered in regulatory / HTA bodies scientific advice

Based on its preference study experience, IMI PREFER proposes the topics in [Table 2-8](#) should be addressed in regulator / HTA body Scientific Advice.

Regulators and HTA bodies act within regional and national frameworks and have clear objectives, beliefs and experiences that cannot be easily aligned. Based on the experience from IMI PREFER studies, the following points should be considered in the scientific advice provided by regulators and HTA bodies.

**Table 2-8 Potential topics to be raised during regulator / HTA body consultation**

Framework section	Topics to be covered in Scientific Advice
<b>Framework component 1 "primary study purpose"</b>	<p>Whether the regulator/HTA body agree with the 'whose preferences' aspect of the research question in the context of the:</p> <ul style="list-style-type: none"> <li>planned use of the results (study endpoint selection, benefit-risk trade-offs)</li> <li>heterogeneity of the target group</li> <li>representativeness of the target study group for the patients who will be treated with a drug in a given therapeutic environment once approved (across regions, countries).</li> </ul> <p>Early in the development of new medicinal products, this will be built on target patient profiles or confirmatory study concepts.</p>
<b>Framework component 1 "preference study objectives"</b>	<p>If applicable, for example if any aspects of the research sub-questions will be of special relevance to the decision-maker.</p>
<b>Framework component 2: "Design – method selection and analysis planning"</b>	<p>Whether the regulator and/or HTA body agree that, the selected method is appropriate to the research question and the proposed sample population.</p> <p>Whether the regulator/HTA body agree that, the proposed analytic approach (i.e. what approach will be taken to framework component 3 'Application of preference data') is acceptable to support the intended regulatory and/or HTA decision(s).</p>

<p><b>Framework component 2: “Design – sample definition”</b></p>	<p>Whether the regulator and/or HTA body agree with the proposed study population inclusion &amp; exclusion criteria and agree with the approach to sample sizing. For example, a regulator may require study sample having physician-confirmed diagnosis of a specific condition because mis-diagnosis or mixing up with another condition is known to be quite common.</p> <p>A discussion with regulator/HTA body about the preference population is especially helpful when proposing that the population is made up of care-givers.</p>
<p><b>Framework component 2: sample size</b></p>	<p>Whether the regulator and/or HTA body agree with the approach to sample sizing.</p>
<p><b>Framework component 2: “preference question design”</b></p>	<p>Whether the regulator and/or HTA body agree with the instrument design, including selected attributes.</p>

Examples:

1. Important evidence for the usefulness of scientific advice has been gathered by IMI PREFER from work with the UK’s National Institute for Health and Care Excellence (NICE) on the COPD case study. The scientific advice clarified the stakeholder’s expectations and most critical questions, and enabled the research question to be refined. This example showed that the Scientific Advice helped to clarify expectations on critical topics such as appropriate patient population, intra-country differences, extrapolation criteria.
2. Critical input was provided by HTA bodies for the PAVING study. This was highly relevant as the objective was to understand the trade-offs that patients make when they are asked to choose between gene therapy and the current standard of care.
3. In the rheumatoid arthritis study by Uppsala University, regulatory input was sought, which was of particular importance as the study aimed to estimate the minimum acceptable benefit and to explain preference heterogeneity.

All these learnings from stakeholder engagement have been included in the framework presented by IMI PREFER. As each case of patient preference study, context and use case is different, scientific advice remains critical until patient preference methodologies are fully implemented into drug development and approval.

In summary, the context of the use of the results is critical. The use cases identified in [section 3.4.3](#) inform on:

- what uncertainty is acceptable to patients
- the selection of endpoints for clinical trials
- identify views on benefit-risk trade-off decisions

Views on trade-offs in the label will require different level of reassurance by regulators and HTA bodies on the validity of the results.

## 3 PREFER framework for patient preference studies

### 3.1 Objectives and overview of the PREFER framework for patient preference studies

This section describes the objectives and structure of the PREFER framework for patient preference studies ([section 3.1.1](#)), a discussion of some key considerations about methodological and operational issues relating to the PREFER framework, plus a discussion of how the framework addresses issues relating to heterogeneity among patient preferences and potential concerns about bias ([section 3.1.2](#)).

The aspects of the framework, which can be especially helpful to discuss during a Scientific Advice process, are described in [section 2.5](#).

#### 3.1.1 PREFER framework objectives and overview

Frameworks for specific topics have promoted a common understanding across regulators, HTA bodies and payers, and industry on best practices to follow in implementing new scientific paradigms and have enabled a more structured dialogue around that topic. See, for example, recent work on frameworks for structured benefit-risk analysis ([Pignatti 2015](#); [FDA 2018](#)), the HTA Core Model and methodological guidelines for relative effectiveness assessment of EUnetHTA, and the recently updated [ICH E9 2017](#) guideline about estimands.

The objective of the PREFER framework for patient preference studies is to:

- Inform a preference study research team on key considerations when designing, conducting and applying the results of a fit-for-purpose preference study
- Guide decision-makers when assessing and using preference study results to inform decision-making
- Support the discussion between industry, regulators, HTA bodies and payers about preference studies intended to inform medical product decision-making.

This PREFER framework for patient preference studies aims to facilitate discussions between industry, regulators and HTA bodies and payers about why, when, and how to do patient preference studies as well as how study results can be reviewed and the results used to inform decision making. The PREFER framework also stresses the role of patients and patient representatives in patient preference studies. The PREFER framework is intended to be applicable to all patient populations, with the caveat that sometimes it may be appropriate to ask for preferences from care-givers on behalf of patients rather than asking for preferences from patients directly (as discussed further in [section 3.3.2.2](#)). Very occasionally it will be inappropriate to ask patients for their preferences as this might cause psychological harm (e.g. [Hollin 2016](#) discuss adjustments to inclusion criteria for a stated preference study to avoid psychological harm to participants). In addition, the PREFER framework is intended to be applicable when asking about judgements (what would be best for someone else), when appropriate, e.g. in cases such as schizophrenia, when patients may have difficult assessing hypothetical situations applicable to themselves.

The framework is intended to be applicable to both qualitative and quantitative patient preference studies. Qualitative and quantitative studies are expected to have different applications e.g.

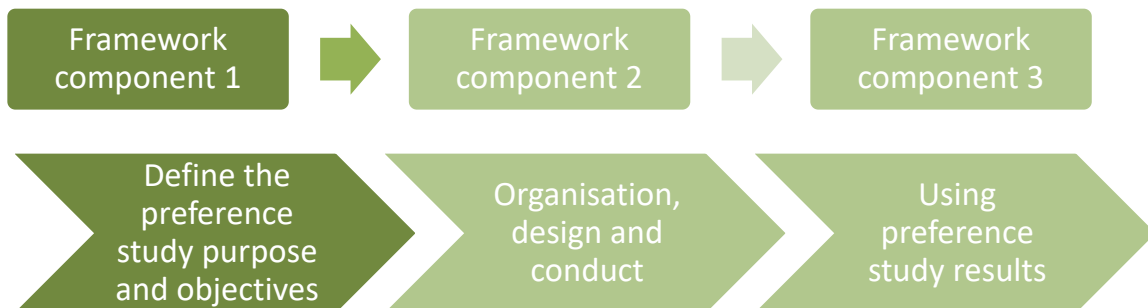
- When using preference study results, regulatory, HTA and/or payer decision-making would typically be informed by results from quantitative preference studies.
- Quantitative preference studies will often - but not always - be preceded by a qualitative study, since a qualitative study can help provide information on the attributes to be included in a quantitative study. (However, as described in section 3.3.2.5, the choice of attributes can also be done without the use of a qualitative study e.g. by using a top-down approach).
- A qualitative study could be done in isolation (without a follow-on quantitative study) e.g. to characterise the unmet need.

This PREFER framework builds on much existing work in this area, such as the overview of patient preferences in the MDIC (2015) report; the ISPOR report on good research practices for conjoint analysis (Bridges 2011); the FDA CDRH 2015 guideline about submission of preference data; , and further recent work such as the “state of the practice” article on choice experiments to quantify preferences for health and healthcare by Mühlbacher and Johnson (2016) and the overview of preference study stages and steps from Van Overbeeke E(2019). It synthesizes many ideas and concepts from IMI PREFER’s Work Package 2 evidence and expert reviews, and Works Package 3’s case studies and pre-existing work into a unified structure. Annex 1 presents the design of and observations from these case studies in the form of abstracts. Throughout the PREFER framework sections, reference to these observations will be made for illustrative purposes.

This PREFER patient preference study framework has three broad components (Figure 3-1):

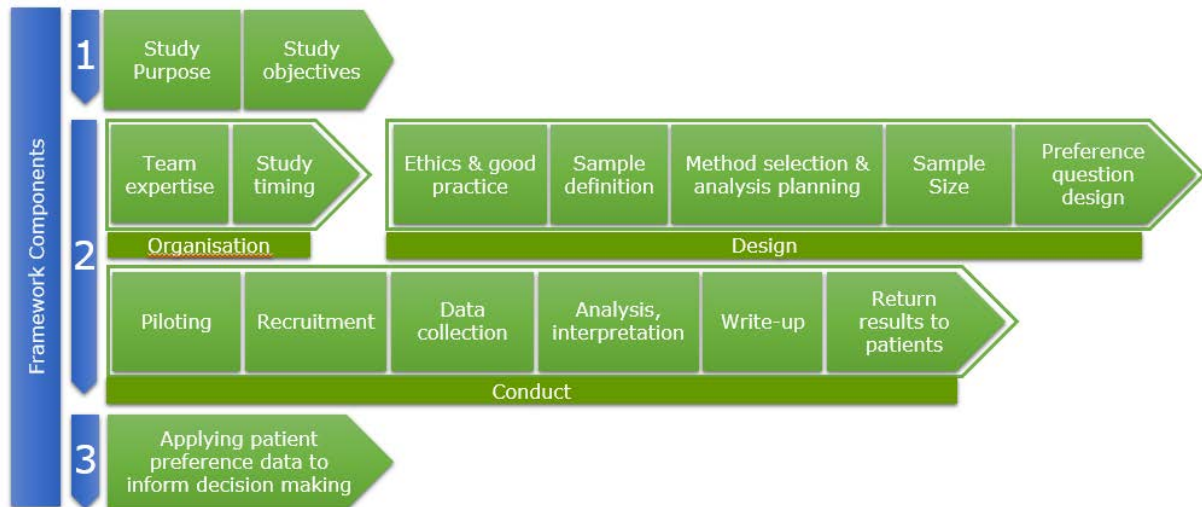
- Definition of the preference study purpose, i.e. the research question to be addressed by the preference study (section 3.2)
- Principles for organising, designing and conducting a fit-for-purpose preference study (section 3.3)
- Application of preference data to inform medical product decision-making (section 3.4).

**Figure 3-1 The three broad components of the PREFER patient preference framework**



The structure of the framework described in this document is based on an adapted version of the stages of a patient preference study described by Van Overbeeke E (2019) (Figure 3-2).

**Figure 3-2 Framework structure, aligned with stages and steps of patient preference studies**



Source: adapted from the diagram included in [Van Overbeeke E\(2019\)](#)

The framework is intended to be applicable to patient preference studies across a broad range of:

- classes of medical products, including medicinal products and medical devices,
- therapeutic classes of medicinal products.

### 3.1.2 Key considerations relating to the PREFER framework

This section describes how the framework links to the methods section of the briefing book ([section 3.1.2.1](#)), how the framework addresses operational issues ([section 3.1.2.2](#)), an overview of how the framework addresses preference heterogeneity ([section 3.1.2.3](#)) and how the framework addresses potential concerns about bias ([section 3.1.2.4](#)).

#### 3.1.2.1 Link between PREFER framework and methodology

All concepts in the PREFER framework are independent of the choice of preference method. However, when applying the framework, some elements of the framework will require method-specific details. [Table 3-1](#) below describes which elements of the framework require method-specific considerations and the relevant methods section of the briefing book.

**Table 3-1 Link between Framework and Methods section**

Framework component	Whether there will be method-specific considerations when applying the framework
<b>Component 1 – preference study purpose &amp; objectives</b> (covered in <a href="#">section 3.2</a> )	PREFER framework considerations for component 1 are method-independent
<b>Component 2 – organisation, design and conduct</b> (covered in <a href="#">section 3.3</a> )	Application of some aspects of PREFER framework component 2 considerations are method-independent, specifically: <ul style="list-style-type: none"> <li>• Organisation                             <ul style="list-style-type: none"> <li>○ Team expertise</li> </ul> </li> </ul>

	<ul style="list-style-type: none"> <li>○ Preference study timing</li> <li>• Design:             <ul style="list-style-type: none"> <li>○ Ethics and good practice</li> <li>○ Sample definition</li> </ul> </li> <li>• Conduct:             <ul style="list-style-type: none"> <li>○ Piloting</li> <li>○ Participant recruitment</li> <li>○ Data collection</li> <li>○ Write-up</li> <li>○ returning results to patient participants</li> </ul> </li> </ul> <p>Application of the remaining aspects of PREFER framework component 2 will require method-specific considerations. This applies specifically to:</p> <ul style="list-style-type: none"> <li>• Design:             <ul style="list-style-type: none"> <li>○ Method selection and analysis planning</li> <li>○ Sample sizing</li> <li>○ Preference question design</li> </ul> </li> <li>• Conduct:             <ul style="list-style-type: none"> <li>○ Analysis, interpretation</li> </ul> </li> </ul> <p>See <a href="#">Section 4</a> for further information the five methods covered in this briefing book and <a href="#">Section 5</a> for about points to consider related to these methods.</p>
<p><b>Component 3 - applying preference data to inform medical product decision-making</b></p> <p>(covered in <a href="#">section 3.4</a>)</p>	<p>PREFER framework considerations for component 3 are method-independent.</p>

### 3.1.2.2 How the PREFER framework addresses operational issues

The PREFER framework provides high-level references to operational issues that may be relevant when planning, conducting or reporting a preference study. The aim is to:

- i. highlight operational issues which may need further consideration when working on a preference study and
- ii. provide reviewers of preference study results with relevant context for such operational decisions (e.g. the choice of whether to recruit preference study participants, within or outside a clinical trial may be influenced by operational issues, relating to embedding a preference study within a clinical trial; the choice of method for a preference study may be influenced by operational issues relating to how the choice of method influences the duration of the study).

Details on operational aspects of preference studies (e.g. how to recruit patients, typical budget) are out of scope of the framework.

### 3.1.2.3 How the PREFER framework addresses preference heterogeneity

Preference heterogeneity refers to the degree to which preferences at an individual level differ from preferences expressed at a collective level ([Huls et al, 2019](#)). Since preference studies measure individuals' preferences - which are by nature subjective - these preferences may differ between

individuals; for example, some patients might be more willing to accept a higher level of risk for a specific level of benefit than other patients (Van Overbeeke E 2019, FDA Guidance on Patient Preference Information, 2016). Some of these differences can be explained by variations in observable characteristics of individuals (e.g., age, severity of conditions, co-morbidities) while other differences may be attributed to unobservable factors (e.g., personal taste, family circumstances). The observable characteristics may include clinically meaningful subgroups specific to the purposes of individual patient preference studies. Understanding the degree of heterogeneity in patients' views within a given sample can be an important aspect of a patient preference study depending on what type of decision that study is intended to address (see Table 3-2 below). Note that preference heterogeneity is distinct from variability of preferences measured for any given individual. An individual's preferences may change over time, or a particular preference assessment instrument may do a poor job of assessing preferences for a particular class of individuals. Hence, the variability in results of a preference study reflects sample preference heterogeneity (between sample variation), individual sample variability (within sample variation) as well as noise introduced by assessment instrumentation.

The PREFER framework covers issues of population preference heterogeneity in several manners:

- In section 3.2.2, preference study objectives: preference study teams should consider the need for study objective(s) that investigate preference heterogeneity across the patient sample.
- In section 3.3.2.3, method selection and analysis planning: one point to consider here is the planned approach a priori to any analyses assessing patient heterogeneity (i.e. analyses linked to the study objectives relating to heterogeneity)
- In section 3.3.2.4, sample size: if a study objective relates to a specific sub-group of patients, one aspect of sample sizing is the potential need for the sample to include a sufficient number of patients in that sub-group.
- In section 3.3.3.3, data collection: data on baseline characteristics, disease characteristics or any other characteristics of the anticipated patient population that may influence their response to in a patient preference study should be collected during the study.
- In section 3.3.3.4, analysis, interpretation: one point to consider within the interpretation is if/how patient heterogeneity influences in the interpretation of results.

**Table 3-2 Example of a case study that evaluated preference heterogeneity**

Study	Way in which the case study addressed heterogeneity	Details
<b>PREFER industry study: patients with a history of myocardial infarction (MI)</b>	Specific study objectives relating to heterogeneity	The primary study objective was to compare preferences for patients at 2 different stages of disease: acute ( $\leq 1$ year of hospitalization) and chronic ( $> 1$ year after hospitalization)  A secondary objective was to assess preference heterogeneity in other relevant subgroups, e.g. by age



	group, gender, medical history, risk of future events, etc.
The preference study report covered heterogeneity within both the results and discussion section.	Preferences for antithrombotic treatment attributes were similar for patients in the acute and chronic stages of disease. Overall heterogeneity of response was observed within specific subgroups, e.g. patients who are 65 years old and above valued reduction in risk of heart attack more than patients who are below 65 years old. Meanwhile, patients without any bleeding risk factors valued reduction in risk of cardiovascular death and heart attack more than patients who have at least one bleeding risk factor.

Of note, the impact of heterogeneity in patients’ views differs depending on the type of decision to be made. See [Table 3-3](#) below for more details.

**Table 3-3 How heterogeneity in patients’ views could impact specific types of decisions**

Type of preference-sensitive decision	Impact of heterogeneity in patients’ views
<b>Understanding patients’ views on the relative importance of issues relevant to their disease or treatment</b>	Heterogeneity in patients’ views could affect the degree to which a choice of patient-relevant endpoints is applicable across a population.
<b>Acceptability of trade-offs</b>	Potentially not all patients may want the new medicine (because not all patients may be comfortable with the trade-off / comfortable with the uncertainty). Hence one aim of heterogeneity analysis is to understand the proportion of the sample who would choose a new medical product and which type of patients would want the new medicine. Another common aim is to assess whether it is possible to identify subgroups who would or would not accept the new medicine under specific levels of uncertainty.
<b>Acceptability of uncertainty</b>	

### 3.1.2.4 How the PREFER framework addresses potential concerns about issues of bias

One concern of stakeholders is whether results of patient preference studies are unbiased ([Janssens R, Russo S, 2019](#)). As with clinical trials (ICH E6), the scientific integrity and credibility of preference study results are closely linked to the study design. Aspects of preference study conduct are also relevant to the overall integrity and credibility of the results.

The PREFER framework aims to emphasize the importance of various study design and conduct issues which are necessary for results that can be used to inform decision-making. Specific aspects of the

framework, which are particularly relevant to addressing concerns about potential for biased results, are described in [Table 3-4](#).

Furthermore, the use of scientific advice options is encouraged so that study sponsors can discuss preference study proposals with regulators and/or HTA bodies; such discussion is expected to assist all parties in developing more experience with and expertise in acceptable approaches to patient preference studies. Such discussion can also ensure that information from preference studies will meet the needs of decision-makers. Details of topics that can be particularly helpful to cover in a Scientific Advice process are described in [section 2.5](#).

**Table 3-4 How the PREFER framework can help address potential biases**

Potential concern about issue of bias	High-level recommendations to address this type of potential bias	Section of the framework in which has further discussion of this recommendation
<b>Lack of pre-specification</b>	As for clinical trials, an analysis plan should be written prior to the results becoming available.	<a href="#">Section 3.3.2.3</a> , method selection and analysis planning
<b>Inappropriate choice of method</b>	The validity and reliability of a method should be considered when selecting a method, or ways to establish validity and reliability should be examined.	<a href="#">Section 3.3.2.3</a> , method selection and analysis planning
<b>Lack of alignment between patient sample and study purpose; potential selection bias when recruiting patients into the preference study</b>	The patient sample should be aligned with the research question, and the representativeness of the preference study population for the target population should be considered.	<a href="#">Section 3.3.2.2</a> , sample definition
<b>Inappropriate choice of attributes</b>	The choice of attributes or scenarios should take into consideration the information that is relevant to the associated decision.	<a href="#">Section 3.3.2.5</a> , preference question design
<b>Issues caused by missing data</b>	Approaches to missing data should be described at the analysis planning stage and at the stage of writing up preference study results.	<a href="#">Section 3.3.2.3</a> , method selection and analysis planning <a href="#">Section 3.3.3.5</a> , write-up

### 3.2 PREFER Framework component 1: the preference study purpose and objectives



This section describes the suggested structure to characterise the preference study purpose ([section 3.2.1](#)), and the consideration of preference study objectives ([section 3.2.2](#)).

### 3.2.1 Framework component 1: the preference study purpose

As in all research, it is critical to identify and characterise the purpose of the preference study. A clearly defined study purpose assists with the consideration of the study objectives (discussed further in section 3.2.2), and hence with the choice of preference assessment method, sample definition, sample size, complexity of the survey instrument, and many other aspects of the study design.

As described in section 2, preference data can be helpful in some scenarios, but is less helpful in other scenarios. A patient preference study should only be performed if it is clear from the study purpose that the preference study can be informative for decision-making, and the study purpose cannot be addressed by existing information.

Aligned with the advice on the research question for conjoint analysis from Bridges (2011), the study purpose should include information about:

- a) **The decision and decision-maker(s)** i.e. which decision(s), by whom (industry and/or regulatory and/or HTA body) will be informed by the results from the patient preference study, together with information about the relevant **decision context(s)**
- b) **Description of how this decision is preference-sensitive**
- c) **Whose preferences** are of interest

These topics are discussed in more detail below.

**a) The decision and decision-maker(s)** i.e. which decision, by whom (industry and/or regulatory and/or HTA body), will be informed by the results from the patient preference study, together with information about the relevant decision context(s).

The design and conduct of the preference study are driven by the decision that needs to be informed by the preference study results, along with the end-user(s) of the preference study results. For example, the study purpose may determine the level of rigor acceptable to the decision-maker – a preference study to inform a regulatory decision about approval of a new drug might require more rigor than a preference study to inform an industry decision about go/no-go of an early-phase drug.

A study purpose could also include decisions by several decision-makers – for example, a decision about the choice of endpoints for a submission study could be relevant to both a regulator and a HTA body.

A description of the decision should include information about the decision context. For decisions relating to patients' views on the relative importance of issues relevant to their disease or treatment, the decision context will typically describe what is currently known about the topic. For decisions relating to patients' views on trade-off or uncertainty situations, the decision context will typically describe the other treatment options available for this indication that will be taken into account by the decision-maker. Different decision-makers can, of course, have different views on which other treatment options are relevant to their decision. For example, the treatment options considered by the decision-maker could vary according to different standards of care in different countries and/or they could vary between regulatory and HTA decision-makers.

Examples of how preference studies can inform industry, regulatory, HTA and payer decisions are described in section 2.2.

**b) Description of how this decision is preference-sensitive**

As discussed in [section 2.1](#), and consistent with the reasons for using patient preferences described in [Janssens & Huys \(2019\)](#), preference-sensitive decisions include situations where there is a need to understand patients' views on:

- i. the relative importance of issues relevant to their disease or treatment (e.g. to inform the choice of endpoints)
- ii. trade-off situations
- iii. acceptability of uncertainty.

**c) Describing whose preferences are of interest**

The study purpose should describe whose preferences are of interest. The selection of these preferences should be appropriate to the decision identified in part a) of the study purpose. The description of whose preferences are of interest would typically include, at a minimum, the indication and the associated population relevant to the decision. For example, in a study informing the regulatory approval of a drug for type 2 diabetes, the description of whose preferences are of interest could be type 2 diabetes patients aged  $\geq 18$ '; for a study informing an HTA decision on a new asthma drug intended as add-on therapy, the description of whose preferences are of interest could be 'asthma patients aged  $\geq 18$  requiring add-on therapy'. Additional considerations could include the degree of experience of the patients with the disease and/or (current or new) treatment (e.g. at risk, newly-diagnosed, newly-treated, experienced with treatment), the degree of disease activity, and confounding factors (e.g. family members with related illnesses, experience with particular treatments). See further discussion of the study population in [section 3.3.2.2](#).

It may not be possible to collect patients' preferences directly; for instance, if the population in question cannot reasonably be expected to reliably self-report (e.g. young children, individuals with cognitive problems such as Alzheimer's disease). The study purpose should clearly describe whose preferences are of interest, even if this group's preference cannot be collected directly. [Section 3.3.2.2](#) has a further discussion of when it may be appropriate to survey care-givers instead of patients.

**3.2.2 Framework component 1: preference study objectives**

The study team should identify appropriate study objectives that describe how the preference study results will inform the relevant decision (as identified within the study purpose).

The primary objectives of a preference study should describe how the preference information will be used to inform the decision(s) identified in the study purpose. As discussed in [section 3.4.2](#), the application of preference data to inform decisions can be done by preferences in isolation, preferences in parallel with clinical data, and/or preference data mathematically combined with clinical data, and a primary study objective would typically be based on one of these approaches. For example, if the study purpose is to inform a decision about the choice of endpoints, the study primary objective could describe how this will be done (e.g. based on importance weighting from the patients). If the study purpose is to inform a decision about the acceptability of benefit-risk trade-offs, the study primary objective could describe how this will be done (e.g. assessment of maximum acceptable risk of safety

issue X for a specific benefit Y / assessment of choice share of new medical product A vs. existing standard of care B for a benefit-risk profile based on benefit 1, benefit 2, benefit 3, risk 1 and risk 2).

The secondary/exploratory objectives of a patient preference study could relate to issues of preference heterogeneity. For example, secondary objectives could include:

- assessing whether preferences are consistent across subgroups (e.g. whether preferences are consistent across patients with varying time since diagnosis of the disease, whether preferences are consistent across patients with differing severity of disease, whether preferences are consistent across sub-groups of patients who do or don't have experience of a specific adverse event).
- investigation of whether patients' preferences are associated with specific characteristics (e.g., patient socio-demographic characteristics, psychological constructs, disease state characteristics).

As when assessing subgroups in a clinical trial setting (ICH E9), investigation of preferences by subgroup should include consideration of issues such as pre-specification and multiplicity adjustments.

### 3.3 PREFER framework component 2: criteria for organising, designing and conducting a fit-for-purpose preference study

#### 3.3.1 Framework component 2, organisation



##### 3.3.1.1 Framework component 2, organisation – team expertise

This section describes recommendations about the expected areas of expertise for the team that will plan, conduct and report a preference study.

###### Expected areas of expertise

The team responsible for a preference study should include members with expertise in:

- medical aspects of the disease and its treatment(s)
- patient preference methodology, conduct and analysis of patient preference studies.

As applicable to the study purpose and objectives, the team responsible for the preference study could also include members with expertise in the following areas:

- regulatory affairs (for a preference study intended to be used to support regulatory decision-making)

- HTA/reimbursement activities (for a preference study intended to be used to support HTA decision-making)
- patient-reported outcomes (PRO expertise can be particularly valuable for a qualitative preference study, since concept elicitation for PROs is very similar to components of qualitative preference studies)
- medical product development
- statistics
- psychosocial scientists e.g. (if planning to include psychosocial instruments in the survey)
- education tools (if planning to include educational tools in the survey)
- patient engagement.

This team may be internal to the organisation running the preference study and/or may include external experts, e.g. as members of a steering committee. Examples of the expertise of study members from PREFER cases studies are shown in [Table 3-5](#). A preference study may be done by or in collaboration with an external partner (e.g. an external consultant, a patient association) and/or a preference study may be conducted as a consortium (e.g. if several industry partners have a common interest in understanding which are the most patient-relevant endpoints in a particular disease area).

Patient involvement in patient preference studies is critical as stated in [FDA Patient Preference Information 2016](#) (and also by [Van Overbeeke E \(2019\)](#), the patient should be ‘the central focus of the study’. Patients are, of course, critical to patient preference studies as participants, but also as partners in the design and conduct of a patient preference study. See [section 2.4](#) for further discussion of this topic.

**Table 3-5 Example of the expertise areas of study team members supporting the PREFER case study (PAVING, Novartis COPD study)**

Study	Expertise areas of the study team members	Details
<p>PREFER additional academic case study: <b>PAVING</b></p>	<p>The study team include included members with expertise in the following areas:</p> <ul style="list-style-type: none"> <li>• medical aspects of the disease and its treatment(s)</li> <li>• patient preference methodology, conduct and analysis</li> <li>• regulatory affairs</li> <li>• HEOR/HTA/reimbursement activities</li> <li>• PRO</li> <li>• medical product development</li> <li>• statistics</li> <li>• patient engagement</li> </ul> <p>The study team also included two patients.</p>	<p>In addition, the study team was supported by an advisory board with expertise in:</p> <ul style="list-style-type: none"> <li>• patient preference methodology, conduct &amp; analysis</li> <li>• HEOR/HTA/reimbursement</li> <li>• PRO</li> <li>• regulatory affairs</li> <li>• medical aspects of the disease and its treatments</li> <li>• patient experience/needs educational tools.</li> </ul>
<p>PREFER additional industry case study: <b>Novartis COPD</b></p>	<p>The study team include included members (from within Novartis and from the vendor) with expertise in the following areas:</p>	<p>The five patient groups involved were The British Lung Foundation, UK; The COPD Foundation, USA; La Fondation du Souffle,</p>

	<ul style="list-style-type: none"> <li>• medical aspects of the disease and its treatment(s)</li> <li>• patient preference methodology, conduct and analysis</li> <li>• HEOR/HTA/reimbursement activities</li> <li>• medical product development expertise</li> <li>• statistics</li> <li>• patient engagement</li> </ul> <p>The study team also worked with five patient groups.</p>	France; Lung Foundation, Australia; and J-Breath, Japan.
--	---	--

### 3.3.1.2 Framework component 2, organisation – preference study timing

The timing of the preference study should be such that it allows the results to be available in a timely manner to support the decision described in the study purpose. An example from a PREFER case study is shown in [Table 3-6](#). For a preference study related to a decision about a specific medical product, the timing should also be aligned with an appropriate level of knowledge about the associated medical product. For example:

- the timing of a preference study to inform the choice of key endpoints would typically be such that the results are available prior to the design of the submission study that would incorporate the chosen endpoints
- the timing of an industry-sponsored preference study to provide information about the acceptability of a benefit-risk trade-off scenario for regulatory decision-making would typically be such that information about which key benefits and key risks are expected to contribute to the benefit-risk trade-off scenario is known prior to setting up the patient preference study, and the preference study results are available for incorporation into the Clinical Overview.

As mentioned earlier, a preference study could inform decisions by more than one decision-maker, such as a regulator and an HTA body. In this scenario, the timing of the study should ensure the results are available prior to the earliest decision.

Additionally, a development program could include more than one preference study. For example, an initial preference study might be conducted in early development to inform the choice of endpoints in a pivotal study; a later preference study might be run to provide information on the acceptability of benefit-risk trade-offs.

**Table 3-6 Example of PREFER case study timing alignment with study purpose**

Study	Alignment of study timing with study purpose	Details
Additional industry study: Novartis COPD	The study was performed so that the results were available ahead of the planning of the submission study (i.e. ahead of the decision identified in the study purpose).	Development of the preference study protocol included scientific advice with NICE.  The preference study was completed in parallel to the clinical phase 2 study, ahead of submission of the clinical study plan for phase 3 and

		subsequent discussions.	stakeholder
--	--	-------------------------	-------------

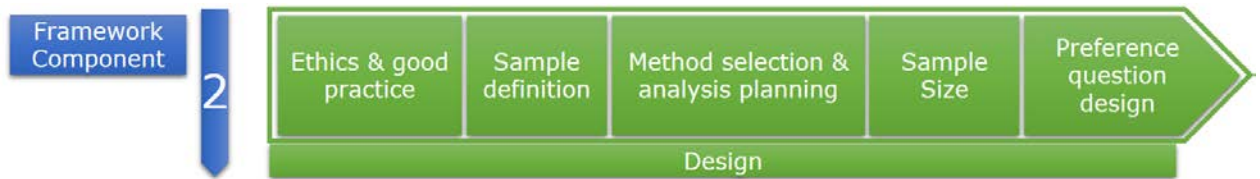
**Operational considerations**

Operational issues will also influence the timing and planning of a preference study, and include the following topics:

- the time needed for consultation with regulatory and/or HTA bodies, if applicable (see further discussion in [section 2.5](#))
- financial resources
- study duration, including the time required for arranging contracts with outside vendors/sub-contractors (if applicable) and adequate pilot testing of survey
- working with patients and patient representatives.

It is important to note that, as mentioned in [section 3.1.2.2](#), operational issues are not addressed in detail within the framework; the intention is rather to flag operational issues that may need further consideration when working on a preference study and to provide reviewers of a preference study results with relevant context for operational decisions.

**3.3.2 Framework component 2, design**



The planned approach to study design, conduct and analysis should be described in a study protocol. Revisions to the protocol after study initiation may occur and can be captured using similar approaches to clinical trial protocol amendments.

One further option to consider is the creation of a distress protocol that describes how to assist a patient in the event that the patient becomes emotionally or psychologically distressed.

**3.3.2.1 Design: ethics and good practice**

Just like clinical trials, patient preference studies should adhere to ethical principles. Ethical considerations include:

- ensuring the study was reviewed by or determined to be exempt by an ethics review board
- ensuring patients provided informed consent
- ensuring the study sponsor provides a lay summary of the preference study results to participants (this aligns with current guidelines about informing trial participants about clinical trial results)
- the trial sponsor should considering registering the preference study at the protocol stage and publishing the study results



- some preference studies will be conducted within clinical trials and would thus follow registration and reporting requirements for clinical trials

Good practice considerations include:

- If preference study results inform a regulatory or HTA body decision, then the corresponding data could be included in the drug label, the regulatory public assessment report or in the HTA body assessment (as applicable).

Operational considerations include:

- the need for a second discussion between the study sponsor and the ethics review board if pre-testing / pilot-testing results in changes to the survey.

### 3.3.2.2 Design: sample definition

This section describes points to consider when defining the study population, inclusion and exclusion criteria and when it may be appropriate to collect preferences from care-givers.

As mentioned in [section 2.5](#), the definition of the study population may be of particular relevance during a scientific advice discussion (this is especially relevant when considering the collection of preferences from care-givers).

**Points to consider when defining the study population:**

- **Alignment of the preference study population with the study preference study purpose**

As described in [section 3.2.1](#), the preference study purpose should specify whose preferences are of interest, which should be appropriate to the decision identified in part a) of the study purpose. Whose preferences are of interest will typically be described by specifying the indication and the associated population (e.g. type 2 diabetes patients aged  $\geq 18$ ; asthma patients aged  $\geq 18$  requiring add-on therapy). The preference study protocol should describe further population inclusion and exclusion criteria as relevant. If needed, e.g. to address study objectives about differences in preferences between specific types of patients, the protocol may also need to describe plans to stratify the sample or otherwise ensure sufficient diversity in the sample. For a preference study supporting a specific medical product, the preference study population should typically be aligned with the population for whom the product is intended.

- **Consideration of the representativeness of the preference study sample for the target population of the medical product**

As for clinical trials (see [ICH E9](#)), no preference study can be totally representative of the target population because of factors such as cultural influences that might vary over time or by country, or other issues. Moreover, it is impossible to know whether a sample is representative of the diversity of preferences that different patients within a patient population might have. Therefore, the objective should be to reduce the influence of known factors where feasible and discuss them when interpreting the preference study results. Examples of choice of study population from PREFER case studies are shown in **Table 3-7**.

- To the extent possible, a preference study should measure the preferences of a representative sample – i.e. a sample sufficiently similar to the intended population – so that the results of the study can be generalized to the population of interest (CDRH 2015; FDA 2018). In many cases, the population of interest will be the planned or actual indication of the medical product that prompted the preference study, and efforts should be made to ensure the preference study population is similar (e.g. in terms of clinical and demographic characteristics) to the intended population for the medical product. Of note, recruitment of patients via a single patient association may result in sample bias (Van Overbeeke E, 2019) and it may be helpful to consider recruitment through a variety of routes.

The level of knowledge of a patient (e.g. about a specific product, or about a specific indication, or about specific endpoints) would typically not be relevant to the choice of preference study population. The design of a preference study includes a step to confirm that participating patients have understood the material relevant to the preference study activities (see discussion of ‘assessment of study materials’ in section 3.3.2.5). However, separately to a patient’s level of knowledge, a patient’s severity of disease or experience with a specific treatment could be relevant to the choice of study population (for example, if a new medical product is intended for patients with severe disease, then a preference study to inform decision-making related to that product might also target patients with severe disease; if a new medical product is intended as add-on to a specific background therapy, then a preference study to inform decision-making related to that product might target patients already taking the specific background therapy).

Consideration should be given as to whether a self-reported diagnosis of disease is acceptable, or whether a physician-confirmed diagnosis is required.

Of note, a representative sample may imply inclusion of patients from multiple countries in order to support the generalisability of the results. Including patients from multiple countries may have operational implications such as submission of the protocol to multiple ethics review boards and potentially a need for translations. See further details in section 3.3.2.5 about considerations relating to translations and section 3.3.3.4 about considerations when analysing data from multiple countries.

**Table 3-7 Examples of preference studies’ approach to the choice of preference study population**

Study	Choice of preference study population	Details
<p>PREFER study: prospective RA case study</p>	<p>As it is not currently possible to cure RA, there is interest in the idea of treating at-risk individuals. The RA preference study aims to assess the preferences of at-risk individuals from two groups:</p> <ul style="list-style-type: none"> <li>- first-degree relatives of patients with RA</li> <li>- general population (without a diagnosis of RA)</li> </ul>	<p>Both the general population and first-degree relatives can be considered to be at risk of developing RA. The risk is, however, estimated to be four times higher in the first-degree relatives.</p> <p>Participants were told to assume an increased risk of RA.</p>
<p>Rituximab preference study</p>	<p>The preference study population consisted of patients for whom</p>	<p>From Rummel et al, 2015 :</p>

(non-PREFER case study)	<p>the product (rituximab) was indicated.</p> <p>Patients' preferences for the type of administration (subcutaneous or intravenous) was assessed in a cross-over study design that meant patients experienced both modes of administration.</p>	<p>"Rituximab in combination with chemotherapy is the standard of care for patients with diffuse large B-cell lymphoma (DLBCL) and follicular lymphoma (FL)"; the study population included patients with DLBCL.</p>
-------------------------	---	--

- **Defining study inclusion/exclusion criteria with the aim of avoiding psychological harm to the study participants**

Very occasionally, asking patients or care-givers for preferences might cause psychological harm because this might, for example, make them aware of aspects of their disease of which were previously unaware (Hollin 2016). This should be considered when defining the study population.

- **The preference study population can be recruited from a clinical trial (and hence defined in the same way for the clinical trial population)**

If a preference study is being run to support development of a specific medical product, then there may be interest in embedding the preference study within a clinical trial for the medical product in question. Preference study participants can be recruited either completely within or completely outside of a clinical trial, or from a combination of within and outside a clinical trial, as in the preference study described in the [Janssen R&D esketamine submission \(2019\)](#). Some pros and cons of each approach are described in [Table 3-8](#), with examples from PREFER case studies in [Table 3-9](#).

**Table 3-8 Pros and cons of recruiting preference study participants completely within or completely outside of clinical trials**

Approach to recruitment	Pros	Cons
<b>Within clinical trial</b>	<ul style="list-style-type: none"> <li>• Straightforward to ensure alignment between preference study population and clinical trial population.</li> <li>• High degree of confidence that the preference population has the disease in question.</li> </ul> <p>A large array of clinical information is available, facilitating the interpretation of preference data.</p>	<ul style="list-style-type: none"> <li>• Patients within clinical trials are not likely to be representative of the broader patient population.</li> <li>• Patients within a clinical trial may be less risk adverse compared to the overall patient population as they have already shown they accept certain risks included in being treated with the medical product.</li> </ul> <p><b>Operational issues</b></p> <ul style="list-style-type: none"> <li>• Requires careful consideration of the timing of the preference study design, since pre-testing the components of a preference study has to be done ahead of writing the clinical study protocol if the preference study must be fully described within the protocol</li> </ul> <p>It can be more expensive to do the preference study within the clinical trial.</p>

<b>Outside clinical trial</b>	Potential to recruit a broader range of patients, rather than only the patients in a clinical trial.	<ul style="list-style-type: none"> <li>• Potentially more difficult to establish with confidence that study participants have the disease in question, particularly if the diagnosis or other clinical characteristics cannot be verified.</li> <li>• Potentially limited information on the clinical characteristics of the preference study participants</li> </ul> <p>More focus needed on how to align the preference study and clinical study populations (only applicable if such alignment will be relevant to the decision to be supported by the preference study results).</p>
-------------------------------	--	--

**Table 3-9 Example of a preference study population from a combination of within and outside a clinical trial**

Study	Choice of preference study population	Details
Preference studies to support <b>esketamine</b> submission (non-PREFER case study)	A preference study population can include patients recruited from both within and outside a clinical trial.	From the <a href="#">Janssen R&amp;D esketamine submission (2019)</a> "The preference survey was administered to 2 patient samples: <ul style="list-style-type: none"> <li>• a <b>clinical trial sample</b> of subjects participating in SUSTAIN-2 and SUSTAIN-3 at sites in the US, UK, Canada and Australia who had direct experience with esketamine nasal spray treatment</li> <li>• a sample of <b>patients from an online panel</b> selected via a detailed screening survey to identify those with a medical history consistent with TRD"</li> </ul> (TRD = treatment resistant depression)

**Care-givers as preference study participants**

The preference study population could be caregivers providing preferences on behalf of patients; this would typically apply when patients cannot report for themselves ([FDA 2018](#)), such as e.g. young children and patients with severe cognitive impairment. An example from a PREFER case study is shown in [Table 3-10](#). If the collection of care-giver preferences instead of patients is considered, it is helpful to discuss this approach upfront with the relevant decision-maker.

Preferences of caregivers may be useful if the caregiver has the legal right to speak on behalf of, and make treatment decisions for, the patient, or when the caregiver has this role in clinical practice (whether or not there is an explicit legal right). In such cases, the caregiver will be asked to at least provide input into the treatment decision, if not make the treatment decision outright on behalf of this patient. As such, caregivers' preferences can be useful evidence to inform regulatory and/or HTA decision-making.

Caregivers can potentially provide preference information from three perspectives:

- Caregiver preferences for patient outcomes: caregivers respond for themselves regarding the effect of treatment on patient outcomes (which effects on the patient matter to the caregiver and in what way?)
- Caregiver preferences for caregiver outcomes: caregivers respond to the effect of treatment on their own daily lives (how does the use of the treatment affect the caregiver?)
- Caregiver beliefs about patient preferences for patient outcomes: caregivers respond as proxies for patients (what does the caregiver believe the preferences of the patient are?).

Most existing preference studies involving caregiver respondents ask the caregiver to report on his or her own preferences for the effect of a treatment on the patient (e.g. [Nunley et al., 2019](#); [Oremus et al., 2015](#); [Hauber et al., 2014](#)).

**Table 3-10 Example of a preference study population including care-giver population**

Study	Choice of preference study population	Details
PREFER prospective case study: NMD	The NMD case study included both patients and caregivers. The caregivers were included to provide on their beliefs about patient preferences for patient outcomes.	“Additionally, for purposes of assessing these topics in this particular population, caregivers will be included in this study as they may be a potential source of additional attributes or attributes levels otherwise not verbalised by the patients themselves and may provide an insight into disease aspects that patients might not fully understand. Caregivers will be asked to contribute based on their judgements regarding their patient needs and not about their own personal preferences. Caregivers representing the voice of patients that otherwise could not participate themselves will also be invited.”

### 3.3.2.3 Design: method selection and analysis planning

This section describes points to consider when selecting a method and planning analyses for the preference study. Operational issues are briefly noted as well.

#### Points to consider for method selection

A key point to consider when selecting a method is its alignment with the study purpose and objective, in addition to considerations relating to participant and feasibility factors. See [Section 4](#) for details on five common quantitative elicitation methods and [section 5](#) for a more in-depth review of points to consider for method selection.

#### Points to consider when planning analyses:

As noted above, the PREFER framework is aligned with key principles of [ICH E9 \(R1\)](#), the addendum on estimands and sensitivity analysis in clinical trials. Specifically, concepts include use of detailed study objective(s), the choice of method that aligns with these study objectives, and selection of study metrics that align with estimand requirements appropriate to the sample and considering how

the preference data will inform the decision (for example: choice share, minimum required benefit, maximum acceptable risk).

1. For both **qualitative and quantitative methods**
  - a. Analyses should be pre-specified, describe the analytical approach (e.g. descriptive statistics, thematic analysis, modeling if relevant), include any planned statistical testing and explain its relevance to the study objective(s).
  - b. The basic approach for the planned pre-specified analyses or assessments could be summarized in the protocol.
    - i. The level of detail in the protocol describing the planned analysis would depend on whether a separate analysis plan document exists for the study. In the absence of such a plan, the analysis description in the protocol should be sufficient for a reader to conduct a similar analysis if the data were available. If there is a separate analysis plan, the analysis description in the protocol should be sufficient for the reader to understand the approach to the analysis and how this approach will lead to results that address the research question.
    - ii. The protocol and/or analysis plan could include an outline of standard analyses for the chosen method, how data would be collected, how data for the research question(s) would be analysed, whether analyses of subgroup or covariates are anticipated (including how these will be identified/defined) including variables that may be used to explain preference heterogeneity. Examples of variables include demographic characteristics, disease characteristics, treatment experience, risk perception or comprehension, and results from any psychological instruments (described in [section 3.3.2.5](#)).
    - iii. Any revisions, well documented as version changes to the pre-specified analysis plans in the protocol and/or analysis plan document, could be made in parallel to or after finalizing the study materials (e.g. survey instrument, discussion guide) to allow inclusion of any changes necessary (assessment of study materials; see 'Preference Question Design' below).
  - c. If patient partners are included in the study team, describe whether and how these patient partners will contribute to the analysis of data and interpretation of results of the patient preference study (see Section 2 Value of patient preferences).
  - d. Describe how the preference data can be used to support the agreed decision (i.e. as described in the study purpose): will the preference data be used in isolation / in parallel with clinical data / mathematically combined with clinical data.
  - e. As part of planning analyses a data management plan should be established and followed. All data should be handled consistently with that plan.
2. **For qualitative methods**- pre-specified approaches may include ([Pope C 2000](#); [Lacey A. and Luff 2007](#)):

- a. approaches for developing descriptions of alternatives and survey materials, collecting participant responses; the process by which transcripts will be transcribed; what approaches will be used for developing and refining codes for and during analysis of transcripts and indexing or charting; the type of supporting software to be used for the analysis of the transcripts, as applicable.
- b. A description of which analytical approach will be used (e.g. thematic analysis, discourse analysis, grounded theory) (Howitt D 2016, Gale 2013; Bradley 2007) in the context of the current study.
- c. how data will be compiled and patterns identified, and results interpreted, such as who will be involved in the interpretation and how specific quotes will be selected from among all other quotes. Describe if and how results will inform alternatives selection and associated descriptions for quantitative study (See Design: Preference question Design section). Patient involvement can be considered, as noted in [Section 2 Value of patient preferences](#).

**For quantitative methods** – pre-specified approaches may include

- a. how variables will be defined and created.
- b. Choosing the correct statistical test(s) and output(s).
- c. Approach to Assessment of survey material(s) and (if applicable) piloting, noting how necessary changes will be implemented and documented.
- d. Assessing for questionable response patterns or evidence of invalid responses that may impact study results; for example, analysis of responses to comprehension questions and if results indicate a respondent's inability to comprehend the material or complete the preference survey in the intended manner. This may include formal assessments of internal validity and reliability, depending on the method selected if such assessments are possible (see [section 3.3.2.3](#))
- e. approach for handling missing data and identifying data from those completing the survey in an unexpected manner (e.g. 'speeders' and partial completion), as well as approach on how to use such data.
- f. There may be a need to conduct additional analyses after completing pre-specified analyses to better understand potentially unexpected results from the preference study (see [section 3.3.3.4](#)). The approach to conducting such analyses, particularly when involving adjustments to models used and assumptions made within those models, should be described as much as possible before beginning analysis.

[Table 3-11](#) provides an example of how PREFER studies pre-specified analyses within a case study. While all case studies could be examples, only one is presented for brevity.

**Table 3-11 Example of case study approaches to study analysis planning**

Study	Approach to analysis planning	Details
<p><b>PREFER study:</b> prospective RA</p>	<p>Statistical analysis plans were created for both the qualitative and quantitative phases of the study, prior to study initiation.</p>	<p><b>Qualitative phase:</b> Nominal group technique was used to conduct the pre-specified number and size of focus groups. Results were collected and transcripts from the groups were coded and analysed using the framework method (software identified where appropriate). The plan for avoiding bias was pre-specified and described. A description of how the results would be ranked and ordered to inform development of attributes into the quantitative survey was provided.</p> <p><b>Quantitative phase</b> Final survey content, including attributes from focus groups, will be developed with patient partners and pre-tested. Endpoints and methods with associated attributes and levels were clearly identified. Questions to assess data quality were identified and included in survey. DCE design and question patterns were pre-specified. Detailed modelling parameters, heterogeneity analyses, and data quality checks were included.</p>

### 3.3.2.4 Design: sample size

#### Importance of justifying the sample size

A preference study protocol should include a justification of the proposed study sample size, based on the primary objective and – if applicable – based on secondary objectives: As in a clinical trial (see the [ICH E9 \(1998\)](#) recommendations on the sample sizing), the number of participants in a preference study should always be large enough to provide a reliable answer to the study objective using the chosen method.

If a primary or secondary objective relate to a specific sub-group of patients, the sample may need to include a sufficient number of patients from this sub-group ([CDRH, 2015](#)).

The approach taken to justify the sample-size will be method-specific. See further discussion of sample sizing within each of the five methods described in [section 4](#). Additionally, a summary of considerations relating to sample size across methods is summarized in [Table 5.2, section 5, Points to Consider](#).

### 3.3.2.5 Design: Preference question design

Some aspects of preference question design depend upon the method to be used for answering the study objective. For method-specific considerations about preference question design, see the corresponding component in [section 4, Methods](#).

Nevertheless, there are core components and general considerations to inform patient preference question design which can involve development of an interview or a discussion guide (for qualitative studies) or a survey instrument (quantitative study, for example) used when eliciting preferences from participants. Within these, the PREFER framework focuses on the following common components and considerations related to preference question design, with each briefly discussed below:



- Context description
- Description of alternatives
- Other design elements related to quantitative studies
- Question development
- Patient education and comprehension (including example exercises)
- Psychological instruments
- Considerations on cognitive burden and capacity
- Translations
- Assessment of the study material(s)

### **Context description**

As noted in Component 1 ([section 3.2.2](#)), clearly defining the study purpose and objective(s) is a critical task that needs completing before beginning preference question design development.

Once the study purpose, objective(s) and planned application is defined, the preference question design development can begin ([Bridges 2011](#)). A critical part of preference question design is the study introduction, as it orients the participant to the entire study. The introductory section includes the context description (sometimes referred to as a scenario or vignette) and may include a collection of baseline participant or disease characteristics (potentially relevant to heterogeneity assessments; see [section 3.1.2.3](#)). The context description has multiple purposes:

- It should clearly describe the reason why the participant is being asked to take part in the study and what topics the preference questions will cover.
- It should inform the participant about the role they should assume for the study. For example, a patient taking part in the study will be asked to complete activities or evaluation tasks from their perspective. For a caregiver, the role description could ask them to respond with what they think is best for the patient or what they would choose if they were making the decision on behalf of the patient.
- It should fully describe the scenario in which the preference questions will be asked. This scenario should provide a realistic decision context (i.e. reflect a reasonable reality of the described situation) from the perspective of those participating in the study. It should include disease description, the situation the participant is asked to consider for the study (e.g. existence of a new treatment option, no treatment options), and, as appropriate, relevant characteristics of current treatment strategies (standard of care) and health care setting-related characteristics (e.g. access to treatment or out of pocket costs) that could influence how participants respond to, but are not included within, the preference questions being posed to patients.

### **Description of alternatives**

Preference questions are used to understand the relative desirability or acceptability to patients of specified alternatives or choices among outcomes or other attributes. The type of features included depend upon the method selected, the study objectives, and those considered important to the

patient. Therefore, the alternatives over which such assessments are to be made must be described to patients. Such alternatives can include a desired state or scenario (e.g. improved quality of life) or multiple characteristics of a treatment (e.g. improved pain control and frequency of dosing). The information used to describe alternatives should reflect the information that are most important to patients.

For preference studies designed to inform a decision-maker about patient preferences related to a specific medical product, the information used to describe alternatives should take into consideration the information about the treatment that is relevant to the preference-sensitive decision – this is generally, differences among the alternatives (see [section 2](#)). Additionally, the information included about the alternatives would take into consideration the decision-maker, bearing in mind that different decision-makers might have different views on which features are important.

Two main approaches are used to identify information about alternatives addressed in the preference questions, often referred to as ‘top-down’ and ‘bottom-up’. The approach taken can be a mix of both and is dependent upon the study purpose and objectives, and the amount of background knowledge within the research team. Regardless of the approach used, all alternative descriptions should be assessed with study participants prior to full study implementation (as described below, Assessment of Study materials).

- **Top-down:** this approach uses existing knowledge or medical product development expertise, including literature review of studies previously conducted in similar populations of interest (qualitative or quantitative).
  - An example of top-down is selecting the primary objective/key benefit from a clinical trial and (anticipated) key risks from the investigational product. Published core outcome sets (COMET) may also inform feature development. Such approaches could be particularly relevant when preference data will be used with clinical trial data to support a medical product decision ([section 3.4.2](#)).
  - **“Anchor”** attribute(s): If a preference study is conducted in a context where the side effects are not yet known, it can be helpful to include “anchor” attributes, which are side-effects with differing levels of impact on the patient (e.g. attributes corresponding to a mild severity or reversible adverse event, a moderate severity adverse event, and a serious adverse event, as in the study described by [Johnson 2019](#)). Once the actual side effects are known, preference information for the actual side effects can be approximated by comparing them to the anchor attributes with a similar level of impact on the patient
- **Bottom-up:** this approach uses direct conversation with the patient/caregiver to understand what matters most to them in the management of the disease. Such data can be gathered through focus groups, semi-structured or open-question interviews, social media listening, and online bulletin boards. Inevitably, such conversations lead to more attributes than can be included in a quantitative survey so ranking exercises are helpful to identify attributes of greater importance than others to inform future research.

Additional considerations when developing alternatives and associated descriptions include

- Alternatives should be defined with as much precision (as little ambiguity) as possible given the research question and decision context, including providing details about the impact and timing of any events.
- The final alternative descriptions included in study materials would be ‘preferentially independent’ – that is, the preference for one alternative would not be influenced by the decision around another alternative.
- Alternatives can include non-health related items such as mode, frequency, or route of administration, location of receiving treatment (e.g. in-patient vs. at home)

As applicable based on study objectives and application of the preference study (sections 3.2.1 and section 3.4.2), appropriate links to clinical trial endpoints would be considered during preference question design. Table 3-12 presents a single example among the many PREFER case studies that utilized the bottom-up approach to developing alternative descriptions.

**Table 3-12 Case study showing bottom-up approach to alternative description development**

Case Study	Plan	Details
<p>PREFER study: prospective RA study</p>	<p>A bottom-up approach was planned: <b>Phase 1</b>, literature search plus focus groups in three countries, separate focus groups for the two sample populations used.</p> <p><b>Phase 2</b>, final content of the quantitative survey was developed in collaboration with patient research partners</p>	<p><b>Phase 1: Literature</b> - A total of 3954 abstracts were reviewed. Twenty studies met the criteria for inclusion, including 15 studies in patients with RA, two studies in the general population, and three studies in first-degree relatives who are at an increased risk of RA.</p> <p><b>Focus groups</b> - Due to the Covid pandemic restrictions, focus groups could only occur in one country; therefore, adjustments were made to have ranking exercises occur in two countries based on focus group results.</p> <p><b>Phase 2:</b> Patient partners informed survey instrument revisions and usability testing of the final draft version. They also contributed to the development of the choice task scenario, and the selection and presentation of survey attributes.</p> <p>The alternative descriptions were preferentially independent. These also included non-health related items (how the medicine is taken and how often).</p>

**Other design considerations related to quantitative studies**

**Development of levels** - For some attribute-based quantitative methods (DCE, Swing weighting, threshold technique) the selection of levels for each attribute is required. The process for developing levels for each attribute can be informed through the alternative description approaches described above. Selection of the type, number and way of presenting associated level(s) for each attribute depends upon the research objective/question and method selected, the clinical relevance and the

clarity to patients. When selecting the type, number, range and presentation of levels, the following are some general considerations:

- Can be categorical or continuous
- Absolute level vs. improvement from current status (incremental)
- Describe change in absolute value instead of relative risk or percentage
- Should be generally realistic or perceived as such; this could be based upon what is anticipated from the medical product in development and/or published data for the treatment choice(s) relevant to the study objectives
- If applicable to the study objectives levels should be interpretable and large enough to induce a decision for participants, such as ranking or trade-off
- Often the same number of levels would be provided for each attribute to minimize unintended bias towards those attributes with more levels ([Mühlbacher and Johnson, 2016](#))

**Internal validity assessments(s)** – Such assessments are a reflection of the degree to which results from a preference study are trustworthy and meaningful; many methods (see [Methods, Section 4](#)) incorporate specific internal validity assessments, the range of which is well documented ([Janssen 2017; Johnson 2019](#)). Addressing issues of internal validity in preference studies should begin with the protocol and statistical analysis plan, where internal validity evaluation should be pre-specified ([Section 3.3.2.3](#)). The chosen internal validity assessment will depend on the study objectives, the preference method, the sample size, and the length and cognitive burden of the study. Additionally, not all internal validity assessments are appropriate for every study since:

- not all measures apply to all preference methods,
- the number of potential validity assessments included in a given study may be limited by the available sample size
- some internal validity assessments may impact the length or burden of the study, reduce the statistical precision of preference estimates or increase the cost of the study overall ([Janssen et al 2017](#)).

Interpreting findings related to internal validity assessments is complex and research continues into how best to do this. Apparent violations of internal validity can be the result of behavioural considerations such as learning, fatigue, or simplifying heuristics used by patients to minimize cognitive effort. These findings could indicate problems related to the study materials or may reflect rational decision-making by respondents. What appear to be 'irrational' preference results do not necessarily mean that a patient's responses are invalid. For example:

- Subjects may infer additional information beyond what is presented in the study materials or use heuristics such as assuming higher costs reflects higher quality ([Mühlbacher and Johnson 2016; Ryan et. al. 2009](#)).
- In choice-based studies, such as those based on DCE or best-worst scaling Case 2 or 3, a patient's response pattern may indicate a lack of attention to the task; that same patterns may also indicate the particular options shown in the relevant choice tasks have similar

utilities for that respondent. In the latter case the selection in those choice tasks appropriately reflects the inherent uncertainty in preferences for these options.

- In a dominance test, one option is defined to be unambiguously better than another; however, the difference in the utilities associated with the two alternatives could be very small. Thus, subjects could reasonably pick the dominated option.

Notably, if a participant is not paying attention to the tasks, there will often be other evidence of this in the data, such as short time to complete the survey, always picking the option in the same position in a series of choice questions, an unusual dominance pattern, or unusually high variance in a participant's responses. Dissecting the drivers of these potential responses is critical and should be a component of planned analyses, including, as appropriate, testing the impact of including these respondents in the dataset, modelling such responses in the analysis, or removing these respondents from the analysis. The analysis plan may also indicate when certain analyses should not be conducted based on the outcome of the validity assessment.

### **Question development**

Development of the questions or exercises that will be presented to a participant to collect stated preference data requires multiple considerations, so inclusion of experts in such design supports appropriate design consideration ([section 3.3.1.1](#)). The approach taken should be documented and rationale should be provided on how the design answers the research question(s).

Depending on the quantitative method used, questions or exercises may require participants to weight, rank, or choose among alternatives. When developing such questions, considerations include ([Bridges 2011](#); [Mühlbacher and Johnson 2016](#)):

- Ensuring a realistic decision context is presented for the perspective of those participating in the study
- Avoiding including irrational or nonsensical questions or exercises that may lead participants to stop paying attention or use simplifying heuristics (e.g. focusing on only one attribute and make a decision based on the level of that single attribute only).
- Utilizing the appropriate number of tasks in the context of the chosen method (for example, methods focused on ranking may present only one task with many alternatives, whereas methods requiring a choice between alternatives will include multiple tasks with two or three alternatives each) (see [section 4](#)).

Additionally, there may be questions or exercises that are qualitative in nature, appropriately open-ended and unbiased ([Lacey and Luff 2007](#)). These are often found within qualitative studies in early medical product development, for example, when ascertaining what is important to patients when treating their disease. These also can be included in quantitative studies to allow a deeper understanding of the patient, their condition and their responses to the quantitative sections of the survey. For instance, patients may be asked why they ranked certain attributes in the way they did.

### **Patient education and comprehension**

Once alternatives are described and questions are developed, principles of clear communication should be applied when preparing all study materials. Care should be taken to frame the questions or exercises in an unbiased, clear, and succinct manner. Key in this context is the involvement of

patients as partners in the attribute and level development phase, as well as in the question development phase (see [section 2](#)). Considerations to achieve this goal are as follows:

- Explanations of alternatives and their descriptions should be tailored to the anticipated health literacy of the planned sample population, including use of plain language. Such descriptions or definitions would be provided when questions are first presented and throughout the remainder of the study. Researchers should make a proposal for tasks covering the alternative descriptions, to be discussed with physicians and patients, and subsequently be assessed/pre-tested with patients.
- As appropriate, the use and descriptions of numeric and probabilistic variables for the features would follow best practices ([Bridges et al 2011](#)). Various publications and accepted practices for communicating levels of risks and associated uncertainties have been documented (e.g. as graphical representations, risk grids, and use of descriptive text with number format). Specifically, a systematic literature review and summary was conducted within PREFER, Task 2.5 [PREFER 2.5 Final Report].
- Educational tools (e.g. visuals, scenarios with choice tasks) could be incorporated to further enhance understanding and motivation to complete the survey [PREFER 2.5 Final Report; [Vass 2019](#)]. Attention should be paid to the length and complexity of the educational tool.
- Inclusion of comprehension questions in the study materials to assess understanding of alternatives and their descriptions, along with the exercises themselves or any associated numeric or probabilistic concepts helps inform the researchers if the text conveys the intended message. Such comprehension questions also may inform the interpretation of study results ([Section 3.3.3.4](#)).

All approaches to clearly communicating and assessing how well-informed participants are through the descriptions should be ascertained by an assessment of the survey prior to full study implementation (see 'Assessment of the Survey' below).

Importantly, there may be instances where the sample population presents with a disease leading to cognitive impairment, such as some dementias and neuromuscular diseases. In such cases, particular care should be taken that the descriptions of the alternatives, context (i.e. vignettes), and questions, along with other study materials are appropriately worded and thoroughly assessed prior to full study implementation. It may also be possible that such populations are incapable of understanding the preference questions format. Finally, consideration should be given to include caregivers of such patients to provide additional information on preferences for such a population.

[Table 3-13](#) presents two examples among PREFER case studies related to including patient education and comprehension considerations into the studies.

**Table 3-13 Case study showing the PREFER framework approach for patient education and comprehension considerations in preference studies**

PREFER Case Study	Plan	Details
<p>Prospective <b>PAVING case study</b></p>	<p>Survey development included a qualitative phase and a quantitative phase.</p> <p>Patient partners and advisory board consultation were used to determine the final quantitative survey alternative descriptions and study material language.</p> <p>Assessment of the survey planned to determine clarity of language.</p> <p>An educational tool was included</p> <p>Comprehension questions were included in the quantitative survey.</p>	<p>Qualitative ranking by focus groups and assignment of points through pre-specified methodology resulted in identification of the top five attributes for consideration in the quantitative phase of the study.</p> <p>Advisory board consensus was reached to include attributes in the survey that were most important to patients, with emphasis on including a quality-of-life-related attribute.</p> <p>The phrasing of another attribute (probability that prophylaxis can be stopped) was rephrased to ‘chance to stop prophylaxis’ as the latter phrasing was more comprehensible to patients.</p> <p>Both text and figures were used to present attributes and levels of the alternatives to patients. One attribute (factor level) was excluded as it was dependent upon another attribute (i.e. it was not preferentially independent).</p> <p>During ‘assessment of the survey’ (referred to as ‘pilot’ of the full survey in four patients), only minor text changes were made.</p> <p>9 out of 117 participants incorrectly answered the comprehension question.</p>
<p>Historic case study: <b>Framing Risk Attribute</b>, Veldwijk et al, 2016</p>	<p><b>Main objective:</b> To test how attribute framing in a DCE affects respondents' decision-making behavior and their preferences.</p>	<p><b>Methods:</b> Using two versions of a DCE questionnaire, a representative sample of a Dutch population was asked to consider genetic screening for colorectal cancer (CRC). The risk attribute included was framed either positively, as the probability of surviving CRC, and negatively, as the probability of dying from CRC.</p> <p><b>Results:</b> The majority (56%) of the respondents ranked survival as the most important attribute in the positively framed DCE, whereas only a minority (8%) of the respondents ranked mortality as the most important attribute in the negatively framed DCE. Respondents made dominant choices based on survival significantly more often than based on mortality.</p> <p><b>Conclusion:</b> The framing of the risk attribute significantly influenced all attribute-level estimates and resulted in different preference structures among respondents in the positively- and negatively-framed data set.</p>

### **Psychological instruments to potentially explain differences in preferences**

Currently, preference studies provide little information on why patients might differ in their preferences, although it is understood that patients' preferences can be influenced by a wide range of situational and dispositional variables. There is general consensus regarding characteristics of the decision process and the situational factors involved in the construction of preferences, but less is known about the influence of individual differences on preference formation ([Appelt 2011](#)). It is therefore important to identify the psychosocial constructs that affect patients' preferences in order to gain an understanding of the primary influences on their health-related decision making.

Evaluating the psychological profile of patients, including personality traits and emotional and cognitive functions, may reveal critical determinants of the decisional processing of patients and may detect crucial factors to explain and predict patient preferences and health-related decisions ([Russo 2019](#)). Some examples identified by Russo and colleagues as potentially informative for inclusion in preference studies include:

- health locus of control – a generalized expectation about whether one's health is controlled by one's own behaviour or forces
- patient activation – the degree to which an individual possesses knowledge, motivation, skills and confidence to make effective health-related decisions
- health literacy and numeracy – the patient's ability to read, understand and use healthcare information appropriately and ability to apply and manipulate numerical concepts, respectively.

### **Considerations of cognitive burden and capacity**

Throughout the design process, a consideration on the cognitive burden and/or capacity of the participant should be kept at the forefront. Such considerations include the patient population (e.g. age, presence of cognitive impairment, educational level), using best practices in communicating the attributes (and levels, as applicable), and length of the interaction/survey. Particular attention should be placed on the medium of conducting the study: online or in person. Assessment of the study materials (see below) in an iterative fashion through face-to-face interviews is critical for verifying the questions are understood in the intended manner.

Additional consideration should be made for those who may encounter potential emotional or psychological harm when presented with outcomes/harms of the medical product or disease. At the preference instrument design stage, the risk of this harm can be minimized by involving medical experts and patients (or patient representatives) in the review of study materials, as well as by the close observation and collection of feedback during pre-testing (assessment of study material, described below).

### **Translation of study material(s)**

For patient preference study materials that require translation, it is recommended that the ISPOR Principles of Good Practice for translation of patient-facing material ([Wild 2005](#)) is followed. Initial translations could be made by official translation companies. However, having patient review of these



translations would be helpful to affirm the translated versions are understandable to patients in their local language.

**Assessment of study material(s)**

It is recommended that study materials are assessed by interaction with patients prior to full study initiation. Commonly referred to as pre-testing, study materials can often be assessed through one-on-one interviews or talk-aloud exercises with a convenient sample of patients. These interviews can be used to assess whether the content is understood in the intended manner, whether questions and exercises are clearly understood, realistic, adequate in terms of length, and whether levels are sufficient to induce trade-offs. This assessment also provides insight into the amount of cognitive burden to participants. Depending on the complexity of the research question or context, a subsequent pilot in a small portion of the full sample may be desirable (see [section 3.3.3.1](#)).

At this point in the study design, the study team may identify content within the study materials that may be causing unexpected harm to participants. Such information can be used to guide changes to the study materials or to inform the development of distress protocols or other mitigation efforts.

**3.3.3 Framework component 2, conduct**



During the 'conduct' stage of a preference study, teams should continue to apply the principles of ethics and good practice as outlined in [section 3.3.2.1](#).

**3.3.3.1 Conduct: piloting**

For the PREFER framework, piloting is typically only completed for quantitative studies. It is defined as a soft launch of a survey with a small subset of the full sample to check, for example, that the survey and data collection work as expected, and for possible excessive cognitive burden for patients. This differs from 'Assessing study materials (described in [section 3.3.2.5](#)), which is good practice across both qualitative and quantitative studies.

Prior to launching the quantitative survey into the full sample, piloting and associated assessments can be performed to test if programming and encoding of the survey instrument was correct (results as expected), to assess the validity and reliability of the preference elicitation method, and to better understand if the range of levels is sufficient to induce trade-off behaviour. Errors can then be corrected as needed and clearly documented prior to launching the survey into the full sample population. Examples from PREFER case studies are shown in [Table 3-14](#).

The planned approach for piloting should be described in the protocol, as applicable, including planned approaches for documenting necessary changes.

**Table 3-14 Examples of how case studies updated a questionnaire based on piloting with patients**

Study	Information about the pre-testing and impact on the final questionnaire	Details
<p>PREFER study: prospective study of preferences for patients with history of myocardial infarction (MI)</p>	<p><b>Objective of the quantitative pilot interview</b></p> <p><b>Number of patients included:</b> 40 chronic MI patients.</p>	<p>The primary goal of the quantitative pilot was to update the information priors used in the DCE experimental design. The initial priors employed were based on existing literature and/or expert opinion regarding magnitudes of the preference parameters. If the new information on patients' preferences obtained from the quantitative pilot differed significantly compared to those derived from the literature, the DCE experimental design was to be updated.</p>
	<p><b>Outcome of the quantitative pilot interviews:</b> confirmation that the design was inducing trade-off behaviour; update to the main survey to address unexpected responses on one parameter.</p>	<p>Majority of patients were trading across the choice questions (n=39, 98%) with only a small number of patients who always choose based on one attribute (n=8, 20%).                      "All parameter estimates were moving in the expected direction (except for risk of MI) where patients valued risk reduction in CVD, IS and ICH positively. Findings from the quantitative pilot were used to update the DCE experimental design for the main survey. Given the unexpected results observed in patients' preferences for the risk of MI, the DCE experimental design was updated"</p>
<p>PREFER study: prospective RA study</p>	<p><b>Objective of the quantitative survey development and pre-testing:</b> To review the final quantitative attributes and survey instrument prior to finalization</p> <p><b>Number of patients included:</b> ~10</p> <p><b>Outcome of the quantitative pilot interviews:</b> Adjustment to quantitative attributes and questionnaire as detailed in next column.</p>	<p>Patient partners have contributed to the development of the survey instrument in a number of ways, including revising the draft survey and usability testing of the final draft version. They have also contributed to the development/wording of the choice task scenario, and the selection and presentation of survey attributes.</p> <p>As a result of patient partner input we have:</p> <ul style="list-style-type: none"> <li>• Changed the way that RA is described in the background information for survey participants</li> <li>• Removed an unhelpful diagram from the background information</li> <li>• Removed a measure of Health Locus of Control to reduce the length of the survey</li> <li>• Changed the level of baseline risk status we ask participant to imagine in the choice task</li> <li>• De-prioritized attributes highlighted in the qualitative phase that were unlikely to be informative due to dominance (e.g., approval of healthcare professional)</li> </ul>

		<ul style="list-style-type: none"> <li>Revised the wording of the survey instrument, and participant information/consent to improve readability.</li> </ul>
--	--	---

**3.3.3.2 Conduct: participant recruitment**

Since participant recruitment is an operational aspect of a preference study, this activity is not described in any detail within the framework. (The target patient population, and the importance of including a representative sample, has been described in [section 3.3.2.2](#)).

Participation in a preference study has the potential to cause a participant emotional or psychological harm. One step to prevent, inform and help patients in case of emotional or psychological harm is to ensure that during patient recruitment, patients are provided with an information sheet prior to enrolling in the study. Such an information sheet should describe any potential harms of the study and provide contact information in case a patient has any questions or concerns. This is similar to the approach taken in a clinical trial, where an information sheet and informed consent would inform a patient of any potential harm within a clinical trial.

**Operational considerations:**

- Time taken to recruit patients (which can often be longer than anticipated), in particular when recruiting patients with a rare disease and/or patients meeting complex inclusion and exclusion criteria
- Costs associated with recruitment activities and/or reimbursement of patient expenses

**3.3.3.3 Conduct: data collection**

Since data collection is an operational aspect of a preference study, only a few considerations are notable for inclusion within the framework to support conduct in a quality manner:

- data should be collected as described in the final protocol
- for quantitative studies, careful consideration should be given to developing coding for data collection in on-line based surveys; for example, when using randomization of research questions or multiple blocks/sets of questions.

Data collection will be required for:

- the patient preference survey
- Relevant patient characteristics, such as information about demographic,; psychological and disease characteristics (disease duration, severity, etc.) –this data should be aligned with the study objectives (e.g. if a secondary objective relates to whether preferences vary according to disease severity, data on disease severity will need to be collected)

**3.3.3.4 Conduct: analysis, interpretation**

This section describes analytical considerations applicable across multiple methods, although numerous types of analysis can be method-specific, see corresponding sections in [Section 4, Methods](#).

The analyses conducted in a preference study are generally those specified within the statistical analysis plan and/or protocol. However, there is often a need for additional analyses to better understand potentially unexpected results. A careful explanation and implications of these additional analyses or deviations from the planned analyses should be described and provided within the interpretation and write-up of the preference study (ICH E9).

### Considerations with respect to interpreting study results

For qualitative studies, such considerations may include:

- Understanding for which features data saturation was reached, and what level of importance patients gave to features discussed and why.

For **both qualitative and quantitative** studies, such considerations may include:

- Consideration of the majority but also minority groups within results. As appropriate, topics of heterogeneity identified within study objectives and planned analyses would be described (section 3.2.2 and section 3.3.2.3, respectively).
- Having input from patients in this phase may lead to meaningful interpretations. If involved, describe how patients' partners informed the interpretation of study results (see Section 2).

For **quantitative** studies, other considerations may include:

- Understanding whether participants responded in the anticipated way – this could include formal responses to comprehension questions or results from validity assessments (as considered in section 3.3.2.5); and whether this may impact or inform interpretation of results.
- Assessment and discussion of whether assumptions made, such as those related to heterogeneity (e.g. linearity, heterogeneity with respect to demographics, disease and/or psychological characteristics) or statistical assumptions (those required for modeling, if required for the method selected) hold true once data are collected (Bridges 2011; Mühlbacher and Johnson 2016).

### 3.3.3.5 Conduct: write-up

#### Considerations when reporting study results

Guidelines on reporting practices have been published relevant to qualitative (Hollin et al. 2020, Lacey and Luff 2007) and quantitative studies (Bridges et al 2011).

For both qualitative and quantitative studies:

- include consort-type diagrams describing the number of people invited to participate in the study, the number who agreed to participate and the number who completed the study.
- include an explanation of the patient preference study purpose that covers:
  - the decision(s) and decision-maker(s)
  - the type of preference-sensitive situation
  - whose preferences are of interest

- include a description of the study objectives
- include appropriate description and rationale for the method, instruments and design used
- the presentation of results should include a description of the sample population, a description of missing data and/or impact on interpretation of results, summaries of planned and any additional analyses as noted above
- the discussion of results should also include a transparent consideration of study limitations (including potential bias) and applications (including generalizability and/or representativeness, if applicable – see sample inclusion/exclusion criteria, [section 3.3.2.2](#)).
- include conclusions on what was learned from the study and future implications of the research.

### 3.3.3.6 Conduct: returning results to patient participants

The final step when conducting a patient preference study is the return of results to study participants, in plain language in accordance with plain language principles ([FDA, https://www.fda.gov/about-fda/plain-writing-its-law/plain-language-principles](https://www.fda.gov/about-fda/plain-writing-its-law/plain-language-principles)). Ideally, this step should be done in collaboration with patient research partners.

Since this is an operational aspect of a preference study, this activity is not described in detail, but a suggested structure for this communication is shown in [Table 3-15](#).

**Table 3-15 Suggested structure for communication of preference study results to study participants**

Suggested sections of the communication to participants of the patient preference study	Content
<ul style="list-style-type: none"> <li>• Why is this important for me to read? <b>(Instructions: In this section, frame the research issue and provide high level context and background for this study)</b></li> </ul>	<ul style="list-style-type: none"> <li>• Describe the treatment issue for this patient population</li> <li>• Why is this treatment issue of concern- describe the context</li> <li>• Why did we do this study?</li> </ul>
<ul style="list-style-type: none"> <li>• What did we find out? <b>(Instructions: In this section, summarize study results and provide key patient-relevant messages)</b></li> </ul>	
<ul style="list-style-type: none"> <li>• What can we learn from this? <b>(Instructions: In this section, report the study conclusions that are relevant for patients and reflect on them)</b></li> </ul>	<ul style="list-style-type: none"> <li>• What are the implications for this patient population?</li> <li>• Are there implications for other patient populations? If so, what are they?</li> <li>• For what purpose will these study results be used?</li> </ul>
<ul style="list-style-type: none"> <li>• How did we do this?</li> </ul>	<ul style="list-style-type: none"> <li>• Sample size</li> <li>• Patient inclusion/exclusion criteria</li> <li>• Methods</li> <li>• Countries recruited from; number of patients per country</li> </ul>
<ul style="list-style-type: none"> <li>• Where can I find out more information about this study? <b>(Instructions: Provide links to the</b></li> </ul>	

place where full study results are posted; provide citation to the published scientific results; etc.)

- Who were the study investigators?  
(Instructions: List the names, and affiliations of the study investigators here; Provide the name and email address of the key person to contact on the study team if a patient would like to reach out)

### 3.4 PREFER Framework component 3: applying preference data to inform medical product decision-making



A preference study is conducted to inform a decision, with the study purpose specifying the decision-maker and the decision being informed (b). Component 3 of the framework describes how preference study results can be applied to these decisions, covering a wide range of approaches for applying preference data and addressing some methodological specifics. The focus in this section is on preference studies supporting the type of preference-sensitive decisions described in [section 2.1](#). This section is not intended to be a complete overview of all possible approaches or methodological details.

The discussion in Component 3 includes three sub-sections: [section 3.4.1](#) describes common ways in which preference data can be applied to inform medical product decisions. [Section 3.4.2](#) describes the technical methods behind these applications. [Section 3.4.3](#) includes suggestions about the inclusion of preference data for informing regulatory, HTA or payer decisions in any corresponding public documents (e.g. European Public Assessment Report, product label) that describes the decision.

All these applications of preference data can be done on a complete sample level and on a subgroup level. In some cases, the applications may result in different decisions for different subgroups of patients.

#### 3.4.1 Applications of preference data to inform medical product decision making

There are many potential applications of preference data to inform medical product decision-making. This section discusses various potential applications consistent with the preference-sensitive decisions described in [section 2.1](#), with cross-references to the more technical aspects that are discussed in [section 3.4.2](#).

### 3.4.1.1 Industry and regulatory decisions about choice of key endpoints, HTA decisions about Relative Effectiveness Assessments

#### Type of preference study purpose discussed in this section

- **Industry decision** about the choice of endpoints for development studies. This might be prompted by:
  - a lack of established expectations about endpoints to study in this indication (e.g. drug development in an indication without approved treatments)
  - an interest in re-assessing established expectations about endpoints to study in this indication (e.g. a suspicion that the established endpoints fail to incorporate some aspects of the disease that are important to patients).
- **Regulatory decision** about the choice of key endpoints when making a benefit-risk assessment, based on either indirect or direct use of patient preference study results.
- **HTA/payer decision** about which endpoints to consider when making a relative effectiveness assessment (REA), based on either indirect or direct use of patient preference study results.

All these decisions can be informed by a single patient preference study, as shown in [Figure 2-3](#) in [section 2.1](#).

#### Technical methods applied to these decisions

An **industry decision about the choice of endpoints** for development studies can be made as follows.

- in a situation where there are not established endpoints for an indication: here, a typical approach would be to base the decision on the preference weights, i.e. preference data in isolation (see further details in [section 3.4.2.1](#)). This approach could also be used to understand the importance to patients of a non-health benefit such as convenience.
- in a situation where the preference study is used to re-assess established expectations about endpoints for an indication: one approach could be the use of choice share information (see further details in [section 3.4.2.3](#) and Table 3-16 for an example from a PREFER case study) derived from preference data in combination with hypothetical clinical data. This approach can help describe the relative importance of improvements in the established endpoints vs. improvements in the potential new endpoints.

Of note, there will frequently be discussion between industry and regulator and between industry and HTA body/payer about the choice of endpoints to include in a submission study.

Subsequent to the use of a preference study to inform the industry choice of endpoints, the regulatory decision about the choice of key endpoints would generally be driven by the endpoints used in the clinical study. As stated in the [EMA Day 80 assessment report template](#), the choice of key favourable effects “can often be achieved by including the primary efficacy endpoints and additionally those secondary endpoints that are considered to be of most clinical relevance (i.e., the key secondary endpoints)”. Furthermore, in the situation where the choice of clinical trial endpoints has been informed by preference study results, the regulator could also consider the preference study results as evidence to inform their decision about the choice of key endpoints (per the [ICH M4E\(R2\)](#) guideline, “Information about the patient perspective may be considered when describing the benefits”). This could include the patients’ views on non-health benefits such as convenience, which can be considered as key benefits ([ICH M4E\(R2\)](#)).

Similarly, where a preference study has been used to inform the industry choice of endpoints, the HTA/payer decision about which endpoints to incorporate into a Relative Effectiveness Assessment (REA) would generally be driven by the endpoints used in the clinical trial. Additionally, patient preference data can be used as a source of evidence when deciding which outcomes should be considered in the Relative Effectiveness Assessment (REA). The use of patient preference data directly in health technology assessments can for instance occur to (Bouvy et al, 2020):

- define/confirm the relevant clinical outcomes to be included in the REA
- quantify patients’ preferences towards non-clinical outcomes, such as improved convenience of a new treatment (e.g. mode of administration), impact on activities of daily living, etc..

Preferences for disease-specific health outcomes add to the information from preferences for generic outcomes in a REA (e.g. SF-36), especially if the disease-specific outcomes are defined based on what patients would most like to see improved. Understanding patients’ views on the relative importance of issues relevant to their disease or treatment allows decision-makers to assess to what extent the treatment under consideration meets the patients’ needs. This leads to a more accurate and patient-centred REA. It is a normative choice the HTA agency needs to make about whether disease-specific information is considered relevant or not for the decision-making.

**Table 3-16 Example of how a PREFER case study assessed patients’ views on choice of endpoints via choice share**

<b>Study:</b>	Additional industry study: <b>Novartis COPD study</b>
<b>How the study assessed patients’ views on the choice of endpoints:</b>	<b>Primary objective:</b> The level of importance that COPD patients place on symptoms such as cough and mucus secretion and consequences thereof, vs. more traditional endpoints such as breathlessness and exacerbations. <b>Technical method:</b> choice share methodology, using preference data combined with (hypothetical) clinical data.
<b>Conclusion (taken from the study report)</b>	The study showed patient preference studies are a useful way to assess which aspects of a disease matter most to patients. Analysis of changes in preference weights (utilities) for equivalent improvements in cough and mucus combined, are higher (more valued) than those for shortness of breath alone.
Details (taken from the study report):	Preference share for different health states  Preference share simulation was used to examine the effect of changes in health status on predicted patient preferences. In order to better understand the contribution of multiple attributes to predicted patient preferences, we set up a range of scenarios to examine two health status profiles. Starting point is the average health state profile of the patient population included in the patient preference study. The two health status profiles were incrementally changed from equivalence (= average patient health status) towards their defined ‘improved’ status, and the impact on predicted patient preference was observed.  <ul style="list-style-type: none"> <li>• “A” – the profile in which a single attribute (e.g. shortness of breath) is improved.</li> <li>• “B” – the profile in which a group of attributes (e.g. cough and mucus) are improved.</li> </ul> <p>The data were analysed in a way to come close to a real-world scenario where patients are impacted daily by excess mucus production, (chronic) cough (which usually occur together) and shortness of breath in addition to downstream impacts</p>



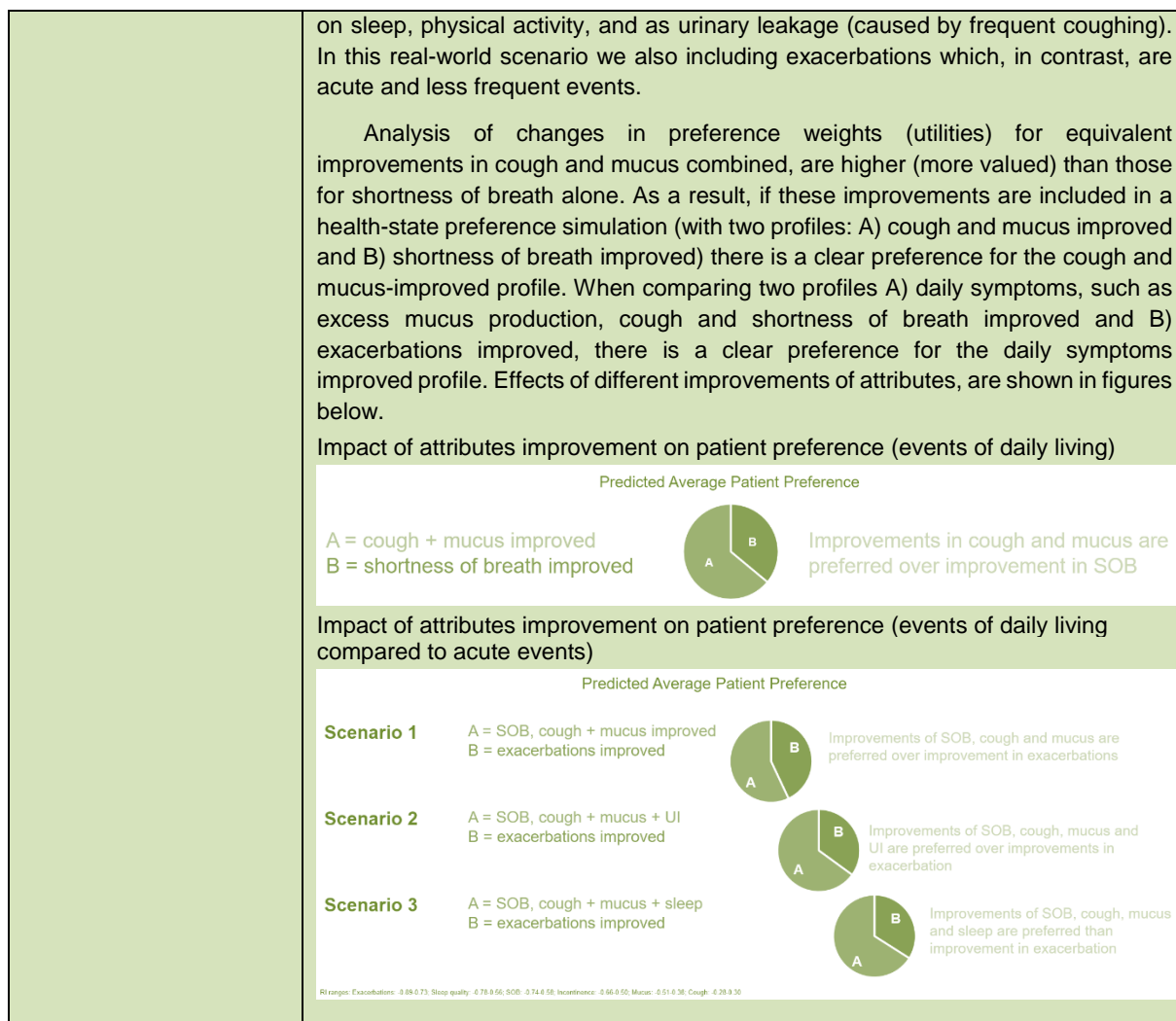
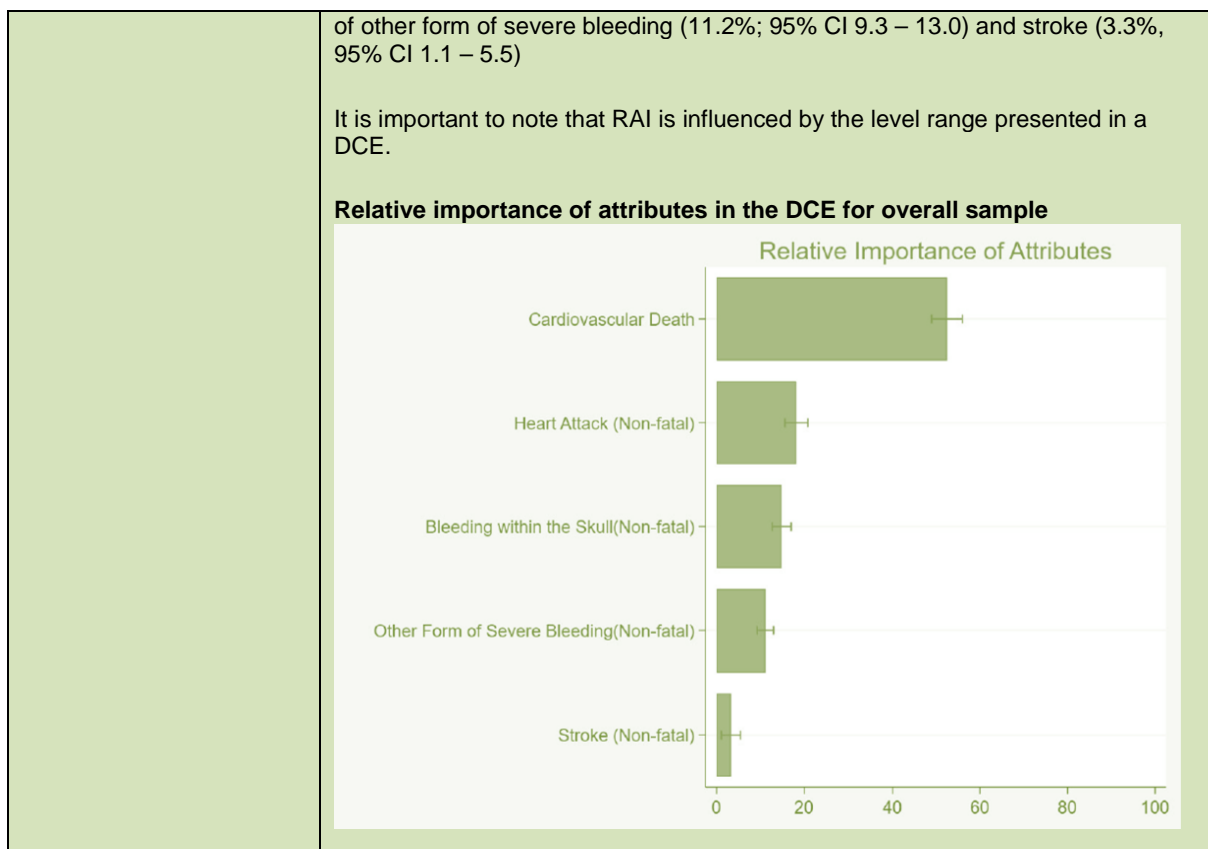


Table 3-17 shows an example of a study assessing preference weights, even though the main objective of this study was not about the choice of endpoints; rather, it was comparing preferences between two sub-groups of patients.

**Table 3-17 Example of a case study assessing patients’ views on relative importance of attributes**

<b>Study:</b>	Additional industry study: <b>MI study</b>
<b>How the study assessed patients’ views on the choice of endpoints:</b>	<b>Primary objective:</b> To compare preferences for antithrombotic treatment attributes of acute and chronic MI patients, with acute MI patients defined as those with their most recent MI within the year ( $\leq 365$ days) prior to study enrolment and chronic MI patients defined as those enrolled more than one year ( $>365$ days) after experiencing their most recent MI.  <b>Technical method:</b> preference weights.
<b>Details (taken from the study report):</b>	The most important attribute for patients was reducing the risk of cardiovascular death (52.4%; 95% CI 48.9 - 55.9), followed by risk of heart attack (18.2%; 95% CI 15.6 – 20.9], risk of bleeding within the skull (14.9%; 95% CI 12.7 – 17.0), risk



### 3.4.1.2 Industry decisions about meaningful effect size

#### Type of preference study purpose discussed in this section

- Industry decision about effect size.** For statistical hypothesis tests, an effect size is a quantitative measure of a clinically meaningful difference between treatments that forms the basis of the test. There are well-established effect sizes for most endpoints used in clinical trial hypothesis tests. However, for novel diseases, or for novel endpoints in well-studied diseases, there may not be established expectations about relevant effect size. Preference studies provide a novel means to define an effect size based on the patient perspective. This approach can be helpful when establishing a meaningful effect size for a new PRO instrument.

#### Technical method applied to these decisions

The approach uses a preference study and a well-established, clinically meaningful effect size for a different endpoint (see example discussed in [section 3.4.2.1](#)). The idea is to ‘borrow’ information on the well-established effect size from a well-understood endpoint and use it to establish an effect size of equal clinical importance for the novel endpoint. For example, if the endpoint #probability of disabling stroke# has a well-established, clinically meaningful effect size or a meaningful change threshold, that effect size or change threshold can be used to inform the effect size in the novel endpoint. The preference study must include both the novel endpoint and the endpoint with the well-established effect size. It is also critical that the attributes in the preference study can be mapped to

both trial endpoints; that is, there must be a defensible means to show that a given level or change in level in attributes corresponds to a given level or change in level in these endpoints.

### 3.4.1.3 PRO instrument scoring

#### Type of preference study purpose discussed in this section

- **Industry decision-making** when setting up a PRO instrument. Patient preference information can help interpret and weigh the changes in endpoints assessed with a PRO instrument by:
  - supporting the development of a scoring algorithm for a PRO instrument – see below
  - helping to define a meaningful effect size for a PRO endpoint – see [section 3.4.1.2](#).

Patient preference information can be associated with PROs in other ways. For example, a PRO could be chosen as an endpoint in a clinical trial based on preference study results showing that the topic captured via the PRO is patient-relevant. Using preference data to inform decisions about the acceptability of benefit-risk trade offs might include trade-offs relating to PROs. Preference studies can provide information about a meaningful effect size in a PRO endpoint. See the further discussion of these points within [section 3.4.1.1](#), [3.4.1.4](#) and [3.4.1.2](#), respectively.)

#### Technical method applies to this decision

- **How will the preference data inform this decision:** Typically, this activity will be based on the use of preference data in isolation (see further details in [section 3.4.2.1](#))

#### Preference studies to support the development of a scoring algorithm for a PRO instrument

Preference studies can gather information about the scoring algorithm for a PRO by describing the relative importance of different domains (e.g. the different constructs) and symptom levels (e.g. the relative importance of moving between mild, moderate and severe symptom levels within a construct). See [Table 3-18](#) for an example. Understanding patients' views on the importance of different domains can help in defining weights to be used when combining the responses to the domains into a single metric. Understanding patients' views on the importance of changes between symptom levels can help in defining a scoring algorithm. E.g. moving from a severe to moderate category could be much more important to patients than moving from a moderate to mild category; moving from a severe to moderate category in domain A could be much more important to patients than moving from a severe to moderate category in domain B ([Johnson et al, 2006](#); [Mohamed A, 2010](#)).

**Table 3-18 Example of how a preference study could inform a decision about PRO instrument scoring**

<b>Study:</b>	As described by Johnson FR, 2016, the study aimed to compare, in patients undergoing chemotherapy, a linear scoring rule to subjective importance for different domain and symptom levels of the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire- C30 (EORTC QLQ-C30).
<b>How the study used preferences to inform a decision about PRO instrument scoring</b>	The study found non-linearity in the importance placed by patients on changes within symptom levels across domains. For example, improvements from severe pain to mild pain, severe fatigue to no fatigue, and severe social limitations to moderate social limitations were all about twice as important as no work to limited work in the Role domain.

### 3.4.1.4 How a preference study about the acceptability of benefit-risk trade-offs can inform industry, regulatory and HTA/payer decisions

#### Type of preference study purpose discussed in this section

- **Industry decisions** about the acceptability of benefit-risk trade-offs to inform submission decisions
- **Regulatory decisions** about the acceptability of benefit-risk trade-offs at the time of market authorisation
- **HTA/payer decision** about the hypothetical uptake of a new treatment (where a decision about new treatment is within the context of the acceptability of a benefit-risk trade-off situation)

As shown in [Figure 2-1](#) within section 2, all these decisions can be informed by a single patient preference study.

An HTA body or payer decision in the context of acceptability of benefit-risk trade-offs will depend on regulatory approval. However, a preference study can nonetheless be useful to HTA/payer decisions in this context by providing information about hypothetical uptake of a new treatment, e.g. to inform budget considerations. Of note, the context for a regulatory decision could be different to the context for an HTA/payer decision, e.g. the regulatory decision might rely on a different assessment of the treatment landscape to the HTA/payer(s).

The approaches described in this section can also be applied to:

- **Industry decisions** about the acceptability of benefit-risk trade-offs to inform **development decisions**
- **Regulatory decisions** about the acceptability of benefit-risk trade-offs post-approval, in the event of a post-approval safety signal that prompts a re-think of a product's benefit-risk profile.

Industry decisions about acceptability of benefit-risk trade-offs to inform development decisions, and regulatory decisions about benefit-risk trade-offs post-approval, would typically be supported by a specific preference study. The timing of the specific preference study would be organised such that the study results are available to inform the relevant decision (see further discussion of preference study timing in [section 3.3.1.2](#)).

#### Technical methods applied to these decisions

An industry decision about submission and a regulatory decision about the acceptability of benefit-risk trade-offs at time of submission could be informed by:

- data displays combining both preference and clinical data (see further details in [section 3.4.2.2](#), and/or
- a side-by-side approach to preference and clinical data to support discussions on maximum acceptable risk for a specific level of benefit / minimum required benefit for a specific level of risk (see further details in [section 3.4.2.2](#)) (this approach is applicable only to simpler benefit-risk assessments with a smaller number of benefits and risks; a benefit-risk assessment with a larger number might be better suited to the mathematical combination of preference and clinical data), and/or

- information on choice share, stochastic multi-criteria acceptability analyses (SMAA) or multi-criteria decision analysis (MCDA) (see further details in [section 3.4.2.3](#)).

An HTA/payer decision about hypothetical uptake of a new treatment could be informed by information about choice share (see further details in [section 3.4.2.3](#)). Of note, the decision context relevant to the HTA decision could be different to the decision context relevant to the regulatory decision (see further discussion of the decision context in [section 3.2.1](#)). Differing decision contexts could mean that the clinical data appropriate to the choice share estimate for an HTA/payer decision could differ that for a regulator. For example, the choice share estimate for an HTA body could be based on clinical data associated with the new medicinal product vs. standard of care A, whereas the choice share estimate for a regulator could be based on clinical data associated with the new medicinal product vs. standard of care B. However, both choice share estimates can be based on the same preference study.

An **industry decision** about the acceptability of benefit-risk trade-offs to inform development decisions could be informed by:

- side-by-side approaches to preference and (hypothetical) clinical data to support discussions on MAR for a specific (hypothetical) level of benefit / MRB for a specific (hypothetical) level of risk (see further details in [section 3.4.2.2](#)).

A **regulatory decision** about the acceptability of benefit-risk trade-offs post-approval, in the event of a post-approval safety signal, could be informed by the same technical approaches as described for a regulatory decision about market authorisation. See [Table 3-19](#) for examples.

**Table 3-19 Examples of how preference studies informed decisions about the acceptability of benefit-risk trade-offs**

Study:		Case study: Chronic Pain (OA, CLBP)
Framework component 1: study purpose and objectives	Study purpose	<ul style="list-style-type: none"> <li>• <b>Decision and decision-maker:</b> regulatory at marketing authorization</li> <li>• <b>How the decision is preference-sensitive:</b> acceptability of benefit-risk trade-off situation</li> <li>• <b>Whose preferences are of interest:</b> patients with moderate to severe knee or hip osteoarthritis or chronic lower back pain, from the US and UK.</li> </ul>
	Study objectives	<ul style="list-style-type: none"> <li>• To quantify patients' preferences for attributes associated with pharmaceutical treatments that are both relevant to patients and that differentiate tanezumab from alternative classes of analgesics (including NSAIDs and opioids), and other NGF-inhibitor products, and</li> <li>• to quantify both the relative importance of each of these treatment attributes to patients and the trade-offs patients are willing to make among these attributes</li> </ul>
Framework component 3: applying preference data to inform decision-making	Application of preference study to inform decision-making	Conditional relative importance was assessed on the following attributes using DCE: 4% risk of severe joint damage, 25% risk of physical dependence, and 0.5% risk of heart attack. For the UK population, 25% risk of physical dependence was the most important of these three risks. 0.5% risk of heart attack was the second most important of these risks.

		The preference results were submitted 'in isolation' along with the clinical trial data via the Clinical Overview and as a separate study report.
	<b>Technical method used for application of preference data</b>	Relative importance
<b>Study:</b>	Preference study as described in <a href="#">Janssen Research &amp; Development Ad Com Briefing Document, 2019</a>	
<b>Framework component 1: study purpose and objectives</b>	<b>Study purpose</b>	<ul style="list-style-type: none"> <li>• <b>Decision and decision-maker:</b> FDA decision on marketing authorisation</li> <li>• <b>How the decision is preference-sensitive:</b> acceptability of benefit-risk trade-off situation</li> <li>• <b>Whose preferences are of interest:</b> patients with treatment-resistant depression</li> </ul>
	<b>Study objectives</b>	To provide information on how patients with TRD would regard the tradeoff between potential benefits of esketamine (improved mood, how quickly the medication works) versus short-term issues associated with dosing (dissociation, dizziness, supervision by a healthcare professional, wait time of 2 hours after dosing, and restrictions on driving) and potential long-term safety issues observed with ketamine abuse (cystitis and memory/cognitive difficulties)
<b>Framework component 3: applying preference data to inform decision-making</b>	<b>Application of preference study to inform decision-making</b>	<p>Preference weights were used to estimate the maximum acceptable risk of potential long-term risks associated with ketamine abuse that respondents would be willing to accept. In exchange for an improvement in depression symptoms from a MADRS total score of 40 to 20 (similar to the mean MADRS change observed within the esketamine clinical trials), patients with TRD were willing to accept a risk of:</p> <ul style="list-style-type: none"> <li>• permanent and severe bladder/cystitis problems &gt;5% (95% CI: &gt;5%→5%) (clinical trial sample) and 4.7% (95% CI: 3.4%→5.0%) (panel sample) (App 16 - Figure 1) (Note, the maximum acceptable risk that can be assessed with confidence is 5% as the maximum chance of severe bladder/cystitis problems shown in the survey was 5%.)</li> <li>• permanent cognitive impairment of 4.7% (95% CI: 3.5%→5.0%) (clinical trial sample) and 3.2% (95% CI: 2.4%→4.1%) (panel sample)"</li> </ul>
	<b>Technical method used for application of preference data</b>	Maximum acceptable risk
<b>Study:</b>	Preference study about obesity devices <a href="#">Maestro Rechargeable System</a> (see FDA press release <a href="#">here</a> ). <a href="#">Ho et al., 2015</a> , Incorporating patient-preference evidence into regulatory decision making, Surgical Endoscopy.	
<b>Framework component 1: study purpose and objectives</b>	<b>Study purpose</b>	<ul style="list-style-type: none"> <li>• <b>Decision and decision-maker:</b> FDA decision on marketing authorisation</li> <li>• <b>How the decision is preference-sensitive:</b> acceptability of benefit-risk trade-off situation</li> <li>• <b>Whose preferences are of interest:</b> adults patients with obesity.</li> </ul>
	<b>Study objectives</b>	Understand patients' views about the acceptability of risks associated with an implantable device for obesity.

<b>Framework component 3: applying preference data to inform decision-making</b>	<b>Application of preference study to inform decision-making</b>	“Additionally, the Agency looked at an FDA-sponsored survey relating to patient preferences of obesity devices that showed a group of patients would accept risks associated with this surgically implanted device for the amounts of weight loss expected to be provided by the device.”
	<b>Technical method used for application of preference data</b>	Maximum acceptable risk

### 3.4.1.5 How a preference study about the acceptability of uncertainty can inform industry, regulatory and HTA/payer decisions

#### Type of preference study purpose discussed in this section

- **Industry decisions** about the acceptability of uncertainty, to inform submission decisions
- **Regulatory approval decisions** relating to the acceptability of uncertainty
- **HTA/payer decision** about the hypothetical uptake of a new treatment (where a decision about a new medical product is within the context of an acceptability of uncertainty situation)

All these decisions can be informed by a single patient preference study (in the same way that a single patient preference study can support multiple benefit-risk assessments, as described in [Figure 2-3](#) within section 2). An example from a PREFER case study is shown in [Table 3-20](#).

A HTA body or payer decision relating to acceptability-of-uncertainty is only applicable once a product has gained marketing authorisation i.e. once the regulator has decided that the uncertainty is acceptable. However, a preference study about the acceptability of uncertainty can be useful to HTA/payer decisions that require information about the hypothetical uptake of a new treatment e.g. to inform budget considerations. Of note, the context for a regulatory decision could be different to the context for an HTA/payer decision, e.g. the regulatory decision might rely on a different assessment of the treatment landscape to the HTA/payer.

Two potential sources of uncertainty are:

- Statistical uncertainty i.e. wide confidence intervals. This might apply to estimates of efficacy and safety for a drug developed for an orphan indication, where relatively few patients were included in the clinical trial(s) for the drug.
- Lack of knowledge e.g. limited/no post-approval use of a drug. Regulatory/HTA/payer decisions will generally be made on products without post-approval use. However, lack of knowledge could be relevant to a trade-off decision e.g. when choosing between drug A with moderate efficacy, tolerable safety and five years of post-approval use vs. drug B with good efficacy, tolerable safety and no post-approval use.

Patients’ views on either/both type of uncertainty can be assessed via a preference study including attributes that address the uncertainty.

- Patients’ views on statistical uncertainty can be assessed with a preference study that includes an attribute describing the certainty in the estimates. For example, [Harrison et al \(2020\)](#) describe a DCE study where – in addition to attributes describing aspects of efficacy

and safety - one attribute was “certainty in estimates”, with levels of very little; limited and moderate.

- Patients’ views on statistical uncertainty can also be assessed with a preference study that includes efficacy and safety attributes where the description of the attribute levels includes a description of the associated uncertainty. For example, [Bansback et al \(2016\)](#) describe a DCE study where description of an attribute level was structured as “Between X and Y people, most probably X people out of 100, will [description of efficacy/safety outcome]”.
- Patients’ views on lack of knowledge can be assessed with a preference study that includes an attribute describing the level of knowledge. For example, the PREFER PAVING study included an attribute “uncertainty of long-term risks”; [Mohamed A \(2012\)](#) included an attribute of “how long the medication has been studied”, with levels of 1, 3 and 6 years; Hauber et al (2009) included an attribute of “what happens if you have bone damage or kidney damage” with one level describing uncertainty, namely “you don’t know if the problem can be treated successfully”, (as well as 2 further levels of “the problem can be treatment successfully” and “the problem cannot be treated successfully”).

**Technical methods applied to these decisions**

An industry decision about the acceptability of uncertainty, to inform submission decisions; and a regulatory approval decision relating to the acceptability of uncertainty could be informed by:

- information on choice share, net clinical benefit or MCDA/SMAA based on the mathematical combination of preference and clinical data ([section 3.4.2.3](#)).

A HTA/payer decision about hypothetical uptake of new treatment (where the decision about a new treatment is within the context of an acceptability of uncertainty situation) could be informed by:

- information on choice share based on the mathematical combination of preference and clinical data ([section 3.4.2.3](#)).

**Table 3-20 Example of how a preference study informed a decision about acceptability of uncertainty**

Study:	<b>PAVING study</b>
How the study used preferences to inform a decision with an uncertainty component	Primary objective (from the qualitative section of the study): To identify the outcomes and other features (attributes) of gene therapy and standard of care that are important to patients, and that will influence their choice
Outcome	Based on the results of the qualitative study, an attribute relating to uncertainty was taken forward into the quantitative study. (The attribute relating to uncertainty was kept constant throughout the quantitative section of the study, hence no further results about patients’ views on varying levels of uncertainty are available.)
Details (taken from the study report)	The ranking exercise with top-down and bottom-up identified attributes revealed that the five attributes most important to haemophilia patients are: annual bleeding rate (ABR), factor level, uncertainty of long-term risks, impact on daily life, and probability that prophylaxis can be stopped.



### 3.4.1.6 Direct elicitation of patients' preferences to support cost-effectiveness analyses

#### Type of preference study purpose discussed in this section

- **HTA/payer decision** about the value for money of a new treatment compared to current standard of care.

A patient preference study measuring preferences for health outcomes included in a disease-specific outcome measure (e.g. a disease-specific PRO) could be used for this purpose. This applies in particular to healthcare systems where reimbursement decisions are based on intra-indication comparisons of treatments (e.g. based on the efficiency frontier), such as in Germany. Comparisons between QALYs gained across indications are obviously not possible in this case.

To keep the possibility of comparing QALYs (or costs-per-QALY gained) across indications, a generic HrQoL instrument could be used in a patient preference study to allow for the calculation of QALYs with patient utility weights. The main argument for using patient preferences rather than public preferences is that the general public does not necessarily know what it means to be in a particular health state. Because it concerns generic health states, and a HTA is mainly interested in the impact of a deficiency on one of more generic health state dimensions, irrespective of the underlying condition of the patients, some agencies (e.g. TLV in Sweden) state that patients valuing a specific generic health state should not necessarily suffer from the condition targeted by the medical product. The only condition is that they experience the health state being valued. (reference: <https://www.tlv.se/download/18.2e53241415e842ce95514e9/1510316396792/Guidelines-for-economic-evaluations-LFNAR-2003-2.pdf>). Others, like the Portuguese agency, recommend that the health states are valued by people that are familiar with the evolution of the disease.

Usually, direct preference elicitation techniques like the time trade-off technique (TTO) or standard gamble (SG) are used for this purpose, because it is not feasible to derive full utility sets for generic health states from patients using indirect preference elicitation techniques like DCE. TTO and SG result directly in patient utility scores that can be used for weighing life years gained for the calculation of QALYs. This approach is applied in Sweden and Denmark (references: [https://www.eunetha.eu/wp-content/uploads/2018/01/Methods-for-health-economic-evaluations-A-guideline-based-on-current-practices-in-Europe\\_Guideline\\_Final-May-2015.pdf](https://www.eunetha.eu/wp-content/uploads/2018/01/Methods-for-health-economic-evaluations-A-guideline-based-on-current-practices-in-Europe_Guideline_Final-May-2015.pdf))

Some scholars argue that both public and patient preferences might be relevant for QALY calculations, and hence results with both sources of preference weights for life years gained should be applied and presented (Versteegh & Brouwer, 2016).

The discussion of the appropriateness of one approach or another to cost-effectiveness analyses is beyond the scope of PREFER.

#### Technical method applies to this decision

For the generation of a single metric from a disease-specific outcome measure, preference data need to be combined mathematically with the endpoints included in the disease-specific outcome measure (see section 3.4.2.3). An example is shown in Table 3-21.

For the calculation of (patient-utility weighted) QALYs, data on life years gained are mathematically combined with a utility value for the health state in these life years gained.

**Table 3-21 Example of how a preference study informed cost-effectiveness analysis**

<b>Study:</b>	<a href="#">Mühlbacher &amp; Sadler 2017</a> and <a href="#">Mühlbacher et al. 2017</a>
<b>How the study used preferences to inform cost-effectiveness analysis</b>	The DCE on patients' preferences for antiviral therapy of chronic hepatitis C allowed the derivation (by weighting of multiple outcomes of the antiviral therapy of cHCV) of an indication-specific and evidence-based aggregated measure to be used in health economic evaluations. In a cost-effectiveness analysis based on the Efficiency Frontier approach, patient preferences were combined with clinical data of interferon-free treatments (aggregation of overall benefit was ascertained using preferences and clinical data) and a net monetary Bbenefit was derived.
<b>Outcome</b>	The latest generations of interferon-free treatments are shown to yield a positive net monetary benefit and be efficient at current prices when taking into account patients' preferences and available clinical data.

### 3.4.2 Technical methods for the application of preference data

The previous [section 3.4.1](#) described common ways in which preference data can be applied to inform medical product decisions. This section describes the technical methods behind these applications.

There are numerous technical methods for applying preference data to inform decision making, which can be broken down into three broad classes:

- Preference data applied in isolation ([section 3.4.2.1](#))
- Preference data applied in parallel with clinical or other data ([section 3.4.2.2](#))
- Preference data mathematically combined with clinical or other data ([section 3.4.2.3](#))

For some decisions, such as choosing primary and secondary efficacy endpoints for an indication without established expectations about endpoints, the decision could be primarily based on preference data in isolation. For other preference-sensitive decisions, such as e.g. benefit-risk assessments during marketing authorisation or HTA and reimbursement decisions, preference, clinical data and potentially non-clinical data (e.g. convenience) are needed. In some cases, viewing clinical and preference data in parallel is sufficient to render a decision. In more complex cases, mathematical models that combine clinical and preference data into probabilistic summary metrics are helpful to gain an integrated assessment that informs a decision.

#### 3.4.2.1 Preference data applied in isolation

There is a variety of ways in which preference data could be used, without the need for clinical data, to inform medical product decision-making. This section discusses the use of preference weights, the use of preference information to derive a clinically meaningful effect size and the use of preference information to understand patients' views on Maximum Acceptable Risk / Minimum Required Benefit.

##### Preference weights:

These are defined as quantitative measures of the relative importance of the attributes included in the study ([MDIC 2015](#)). Qualitative assessments of relative importance from preference exploration methods may also serve in this capacity.

The interpretation of a weight is not always straightforward. The units and meaning of a weight depend on the preference elicitation method used, the definition of its attributes, and the levels used (if relevant to the method). For example, the relative importance in many methods is based on the full range of change in levels used for an attribute. In the table below ([Ho, 2015](#)), the 4.3 relative

importance (preference score) for the average amount of weight loss reflects a change from 0% to 30% weight loss, while the -3.2 relative importance reflects a change from 0 months to 60 months side-effect duration. The difference in sign indicates that increasing weight loss is regarded as a benefit, while increasing side effect duration is regarded as a harm. The relative magnitude of the weights indicates that the 0–30% weight loss increase is more important than the 0–60 months side-effect duration increase. The units of these two relative importances are 'per percent weight loss and per month side effect duration'.

However, in TTO and SG, the weights are in 'normalized units of utility per unit time', and are generally dependent on the time period used in the utility assessment.. In swing weighting, the weights are normalized to the full range of levels used in the method and requires careful interpretation. If all attributes are dichotomous, then the ranges are 'no' and 'yes', and the interpretation of relative importance is less challenging. A relative importance of 0.75 between attributes A and B means that a single instance of the event in attribute B is 0.75 times as important to patients as a single instance of attribute A, everything else being equal. However, if the attributes are continuous or categorical, interpreting relative importance depends on the ranges. For example, if attribute A's range is from 20% to 50% chance of that event, the relative importance of attribute A is over this 30% range only. Interpreting the weight as if it were measured for a dichotomous endpoint (reflecting a change from 0% to 100%) could underestimate the role of that attribute in a decision.

Finally, weights may not be constant over the full range of an attribute. For example, intuitively, the importance of a change from 0% risk of stroke to 5% risk is far greater than the importance of a change from 50% risk to 55% risk – once the chance of a severe adverse event becomes high enough, there is little relevance to the decision-maker of increases in that change – the treatment option is simply too bad. As noted in section XX, value functions are assessed in MCDA to explicitly characterize this non-linearity. In methods like DCE and probabilistic threshold techniques, the part worth implicitly combine both the weight and value functions. In the example by Ho (Ho 2015), the relative importance of a change from 0% to 10% average weight loss is 0.6, while for a change from 10% to 20% it is 1.4 (2.0–0.6) (Table 3-22). Similar non-linearities can be seen in other attributes in the same survey. When applying preference weights to clinical data, it is important that the ranges of the attributes studied in the preference survey are relevant to the ranges of the corresponding endpoints in the clinical data, and that the assessed weights are not applied outside the ranges for which they were assessed.

**Table 3-22 Estimate of preference scores from Ho**

**Table 1** Estimates of preference scores by attributes and levels

Attribute	Level	Preference score (SE)
Average amount of weight loss (TBWL)	30 %	+4.3 (0.52)
	20 %	+2.0 (0.11)
	10 %	+0.6 (0.15)
	5 %	+0.2 (0.23)
	0 %	Reference level
Weight loss duration	60 months	+4.3 (0.47)
	12 months	+2.0 (0.01)
	6 months	+1.4 (0.1)
Side effect duration	0 months	Reference level
	0 months	Reference level
	1 months	-1.0 (0.11)
	12 months	-2.0 (0.09)
Chance of side effects requiring hospitalization	60 months	-3.2 (0.31)
	None	Reference level
	5 % chance of hospitalization, no surgery	-0.2 (0.39)
	20 % chance of hospitalization, no surgery	-0.5 (0.35)
Dietary restrictions	5 % chance of hospitalization, with surgery	-0.6 (0.36)
	Eat 1/4 cup of food at a time	Reference level
	Wait 4 h between eating	-0.1 (0.29)
	Can't eat sweets or foods that are hard to digest	-2.2 (0.33)
Average reduction in dose of prescription drugs for comorbidity <sup>a</sup>	Eliminate need/risk	+3.2 (0.37)
	50 % dose/risk	+2.2 (0.29)
	No change	Reference level
Type of operation	Laparoscopic surgery	Reference level
	Endoscopic surgery	-0.5 (0.3)
	Open surgery	-2.5 (0.31)
Chance of dying from getting the device	0 %	Reference level
	1 %	-3.5 (0.13)
	3 %	-7.1 (0.15)
	5 %	-10 (0.37)

**Clinically meaningful effect size:**

For novel endpoints, there may be little experience from which to base a minimum difference in that endpoint which is clinically meaningful. Preference studies may allow “borrowing” information on effect sizes from well-characterized endpoints to develop a clinically meaningful effect size for a novel endpoint. This approach can be especially helpful when developing a clinically meaningful effect size for a new PRO instrument.

The general approach is similar to that used to compute a maximum acceptable risk or minimum required benefit (Figure 3-3). For a well-accepted effect size, calculate the equivalent change in preference. For the novel endpoint, calculate the change in the endpoint that corresponds to that same change in preference. The clinical impact of this change in the novel endpoints will be the same as that for the well-accepted effect size, and this change in the novel endpoint can be used as its effect size. In this example, the bottom-right arrow shows the well-accepted effect size for heart attack, starting from a baseline of zero chance (0% to 0.5% in this example). The top-right arrow is the change in preference associated with that effect size in heart attack. The top-left arrow left is the same change in preference, shown as an offset from the 0 years in the novel “time-to” endpoint.

The change in the novel endpoint corresponding to this change in preference is shown by the yellow arrow on the bottom-left, giving a two years effect size for the novel endpoint that has the same clinical importance as the 0.5% effect size for heart attack. In this figure, the effect sizes are from a baseline of zero. With a different baseline, results may differ, since as noted above, the change in preference for a given change in attribute may not be constant (i.e., non-linearly value functions) (section 3.4.2.3).

**Figure 3-3 Mock example of using a preference study to develop an effect size for a novel endpoint:**

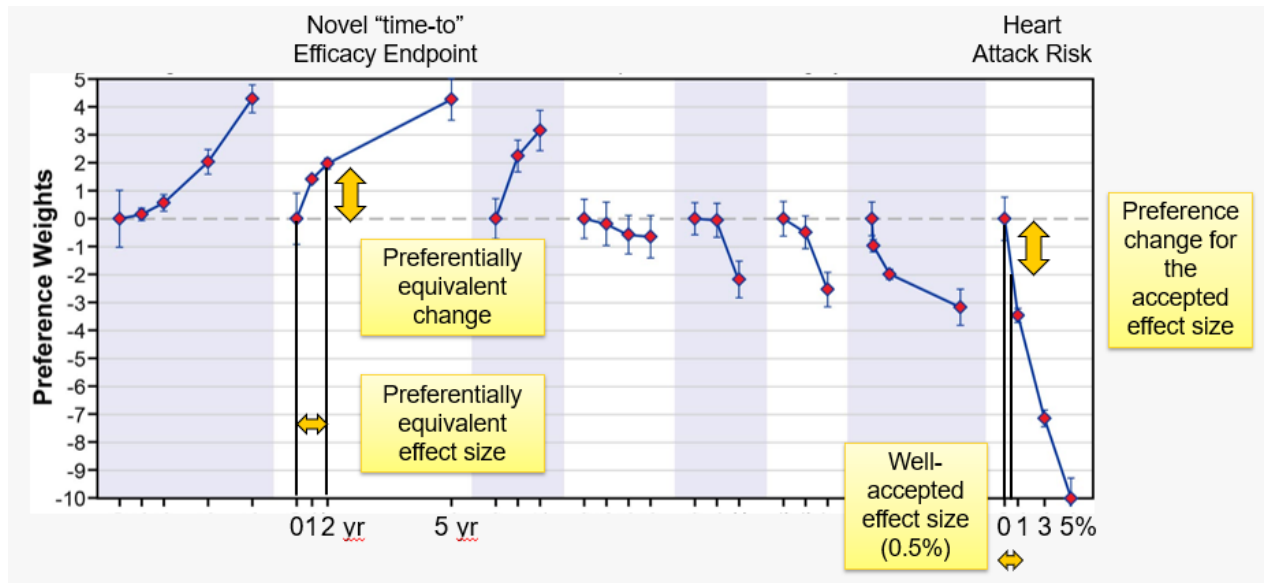


Figure based on [Ho, 2015](#).

**Maximum Acceptable Risk (MAR) / Minimum Required Benefit (MRB)**

Maximum acceptable risk (MAR) is defined as “the greatest increase in probability or magnitude of a harm a patient would accept to achieve or realize a given benefit” (adopted from the definition in [MDIC 2015](#)). This type of information can help inform decisions about the level of risk that would be tolerated – for a specific level of benefit – by patients within a particular disease area. There are numerous examples of this MARs in the preference literature (e.g. [Ho et al, 2015](#); [Johnson et al, 2019](#)).

The general approach to calculating a MAR is similar to the approach described in [Figure 3-4](#) for calculating a clinically meaningful effect size. For a given benefit, assess the change in preference associated with that benefit, then calculate the change in a risk endpoint associated with the same change in preference. That change in the risk endpoint is the maximum acceptable risk for the benefit.

While MAR is straightforward conceptually, it is rare to have only one adverse event of importance in a benefit-risk assessment of medical treatments. When there are several adverse events, a MAR could be computed for each, but it is misleading to consider whether a treatment meets each adverse event’s MAR separately for the purposes of benefit-risk overall. For example, if there are two adverse events and the first has a non-zero incidence, the MAR for the second adverse event will be an

overestimate of how much additional risk of the second adverse event is acceptable. Instead, the concept of MAR can be extended to address which combinations of multiple risks are acceptable for a given benefit or combination of benefits (Figure 3-4) (Angelyn et al, 2020).

**Figure 3-4 Example of maximum acceptable combination of risks for a given level of benefit.**

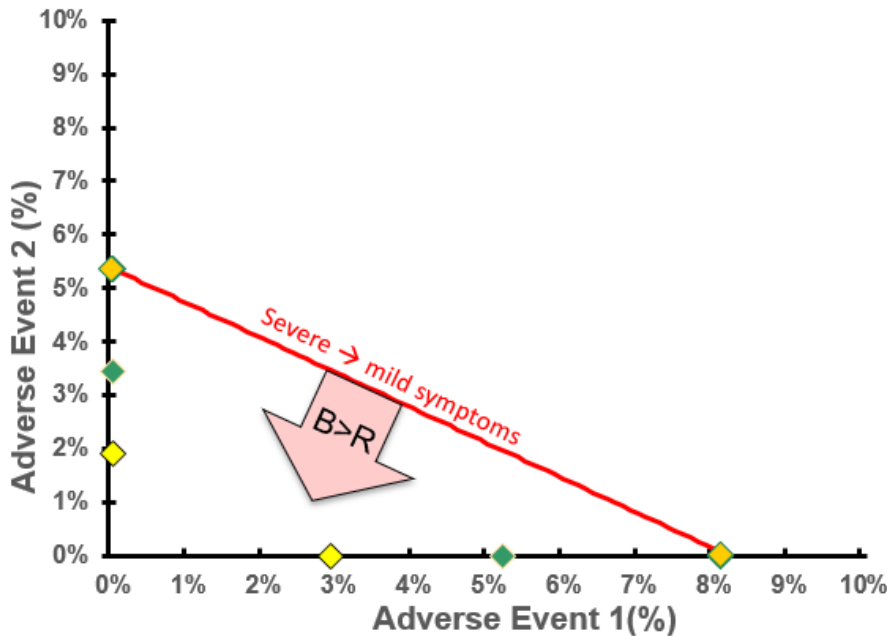


Figure 3-4 shows that a treatment that reduces severe symptoms to mild symptoms, the maximum acceptable risk for adverse event 1 is 8.2%. The maximum acceptable risk for adverse event 2 is 5.4%. If adverse event 2 did not occur, then the benefits outweigh risks on average as long as the incidence of adverse event 1 is below 8.2%. However, if the incidence of adverse event 2 is greater than zero, the MAR of 8.2% is no longer appropriate to use for adverse event 1. Instead, a joint MAR curve can be calculated. In Figure 3-7, the red line shows the joint MAR curve for adverse events 1 and 2, benefits exceeds risks for any combination of adverse events 1 and 2 that are below the red line. The joint MAR curve is linear in this example, but in can take on more complex forms in general. Additionally, the curves have an associated measure of uncertainty, which can be depicted graphically with a 95% confidence interval, allowing visual inspection of the degree to which benefit exceeds the joint adverse event incidences.

Minimum required benefit (MRB) is defined as “the smallest increase in probability or magnitude of a benefit a patient would require to accept a given risk” (adopted from the definition in MDIC 2015). This type of information can help to inform decisions about the amount of benefit that would be expected – for a specific level of risk. As with MAR, this concept can be extended to which combinations of multiple benefits would be acceptable for a given risk or combination of risks.

**Example of the approach of comparing clinical data to Maximum Acceptable Risk**

Results of a preference study can be expressed as a maximum acceptable risk for a specific treatment benefit. For example, in a study of preferences of multiple sclerosis patients (Johnson 2009), the

maximum acceptable annual risk for progressive multifocal leukoencephalopathy was 0.31% for a benefit of slow progression (where slow progression was defined as reducing the number of relapses in the next 5 years from 4 to 1 and increasing the time to next disability progression from 5 to 8 years). This maximum acceptable risk information can be readily compared to actual clinical data for a new multiple sclerosis drug.

Similarly, results of a preference study can be expressed as a minimum required benefit, and readily compared to actual data from a clinical trial.

### Practical use of MARs and MRBs

In clinical trials, the efficacy response is always heterogeneous. Not every patient benefits, and not all those who benefit achieve the same degree of benefit. Applying a single MAR to this population can be misleading, since the benefit used to define the MAR will not align with the diverse degree of benefit observed. One practical means of applying MAR to accommodate heterogeneity is to partition the population into groups with different degrees of benefit and use a different MAR for each. Benefit-risk can then be assessed in each group separately.

### 3.4.2.2 Preference data applied in parallel with clinical or other data

Viewing preference data alongside clinical and/or other data, or basing views of clinical data on preference data, can be particularly valuable for decision-making at the stage of marketing authorisation, an HTA/reimbursement decision or addressing a post-marketing safety issue. In many cases, these visualizations, followed by an integrated interpretation of the data, make a decision clear-cut. They also provide a means to defend a decision in a setting with multiple stakeholders, not all of whom may be intimately familiar with the preference data.

Approaches for using preference data side-by-side with clinical data include:

- Data display: There are numerous means by which tables or figures can depict such data together, such as:
  - effects tables or other tables in which clinical endpoints and preference data are shown in separate columns (see [Figure 3-5](#))
  - forest plots, in which the endpoints are shown in ranked order of decreasing or increasing preference weight (see in [Figure 3-6](#))
  - forest plots, in which the location of the bars depicting the endpoints are shown at locations proportional to their preference weight (see [Figure 3-7](#))
  - parallel bar plots, in which one plot depicts some measure of between treatment differences and an adjacent plot depicts preference weights (see [Figure 3-8](#))

This type of approach is sometimes referred to as deliberative or qualitative Multi-criteria Decision Analysis (MCDA), in which the decision-maker mentally integrates the clinical data and preference data, rather than a quantitative integration with MCDA (see [section 3.4.2.3](#) for a description of quantitative MCDA)

- Comparison of a new medical product's effect on risks and benefits to the Maximum Acceptable Risk (MAR) / Minimum Required Benefit (MRB) obtained from an appropriate

patient preference study. In practice, MAR and MRB are assessed with a measure of uncertainty (e.g. 95% confidence interval), so the comparison between MAR (MAB) and clinical data for a risk (benefit) may need to be done statistically.

**Examples of the data display approach**

There are various ways of showing clinical and preference data side-by-side. Some examples are shown below, though many alternatives are possible.

One approach is augmenting an effects table with preference data. Figure 3-5 shows an effects-table like format of mock data showing endpoint risk differences between treatments and patient preference importance weights. This could be extended with additional columns showing results for each study arm, textual information on the uncertainty, limitations of the data, and references to the sources for the data, providing the information shown in typical EMA effects tables (EMA Day 80 Assessment report template).

**Figure 3-5 Example table showing clinical data side-by-side with preference weights (mock data and weights)**

Endpoint		Rate Difference (95% CI) / 10,000 pat-years		Preference Weight
Benefits	All-cause mortality	-34.2	(-71.62, 3.29)	1.000
	Disabling stroke	-11.8	(-29.31, 5.81)	0.811
	Non-CNS systemic embolism	-14.9	(-24.29, -5.53)	0.420
	Non-disabling stroke	2.0	(-21.16, 25.19)	0.359
	Myocardial infarction	-20.9	(-47.33, 5.59)	0.275
Risks	Major bleeding	14.2	(-35.30, 63.73)	0.267
	Non-major clin relevant bleeding	43.8	(-50.23, 137.77)	0.159

Favors study drug

Favors comparator

A different approach is a figure showing clinical data ordered by patients' importance weights. Figure 3-5 shows data for the rate differences of benefits and harms of study drug vs. comparator with the endpoints ordered according to rank of health utility score from patients – essentially a hierarchy based on the weights. While this approach does not show the actual preference weights, it is often sufficient to support a decision. For example, this approach was used at an FDA Advisory Committee meeting (FDA 2011), where a figure showed that the risk differences favoured the study drug for the endpoints considered most severe by patients, while those that favoured the comparator were considered least severe by patients.



**Figure 3-6 Example of a forest plot with endpoints in order of decreasing health state utility**

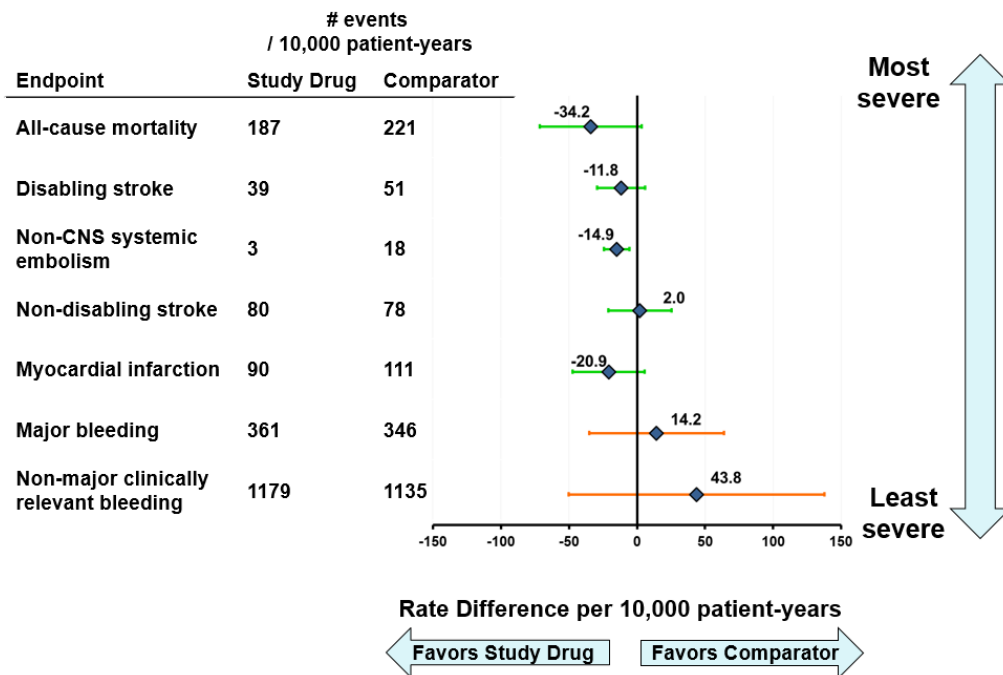
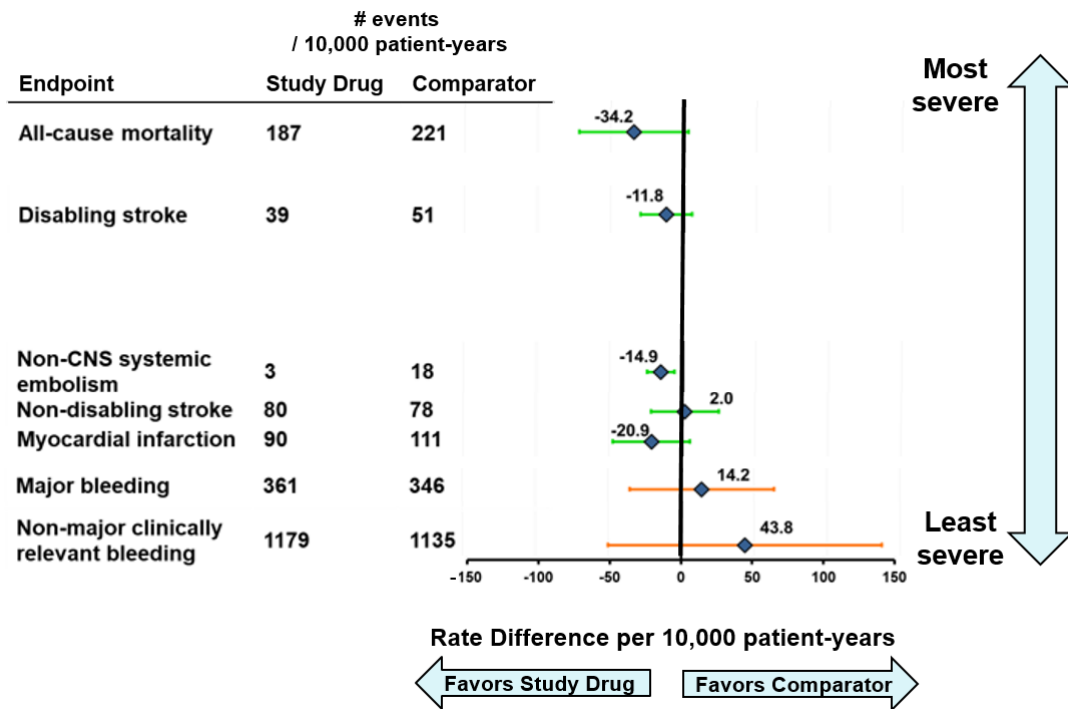


Figure 3-7 shows an example in which the location of the bars depicting the endpoints are shown at locations proportional to their mean preference weight. The advantage of this approach is that it avoids an issue associated with ranking, namely that ranking can hide large differences in preference, such as that between all-cause mortality and disabling stroke vs. the other endpoints in Figure 3-6. When several endpoints have similar weights, it is not always possible to position endpoints exactly where the weights are located without making the figure cluttered. However, the general position rather than the exact location based on weight is generally sufficient to support the decision. Additionally, the figure can be augmented with a table of the actual weights and their measures of uncertainty.

While the rank approach in Figure 3-5) hides the sizes of the difference in preferences between endpoints, it makes for a much simpler figure than proportional positioning approach in Figure 3-6. Additionally, if there is debate about the values of the weights, a rank-based approach may lessen the degree of disagreement between different stakeholders and simplify decision-making. For example, in Figure 3-8, a benefit-risk decision would likely be the same for most people regardless of the order of the first five endpoints. The clinical and preference data determine which visualization approach best serves the needs of the decision-makers.

**Figure 3-7 Example figure showing location of bars at locations proportional to their preference weights**



It is also possible to combine the figure and effects table approaches. Figure 3-8 shows an example effects table augmented with two bar plots, in which one plot shows the rate difference between study drug and comparator, and the adjacent plot shows the preference weights (point estimates).

**Figure 3-8 Example parallel bar plot showing endpoints in order of decreasing weight**

The red/blue bar plot depicts rate difference, with blue bars favouring the study medicinal product and red bars favouring comparator. The green bars reflect preference weight magnitude, with longer bars showing those endpoints are of greater importance to patients.

Endpoint	# events / 10,000 patient-years		Rate Difference (95% CI) / 10,000 pat-years	Rate Difference	Preference Weight
	Study Drug	Comparator			
All-cause mortality	187.8	222.0	-34.2 (-71.62, 3.29)	[Blue bar]	1.000
Disabling stroke	38.5	50.2	-11.8 (-29.31, 5.81)	[Blue bar]	0.811
Non-CNS systemic embolism	4.5	19.4	-14.9 (-24.29, -5.53)	[Blue bar]	0.420
Non-disabling stroke	78.8	76.7	2.0 (-21.16, 25.19)	[Red bar]	0.359
Myocardial infarction	90.8	111.7	-20.9 (-47.33, 5.59)	[Blue bar]	0.275
Major bleeding	359.6	345.4	14.2 (-35.30, 63.73)	[Red bar]	0.267
Non-major clin relevant bleeding	1180.4	1136.6	43.8 (-50.23, 137.77)	[Red bar]	0.159

### 3.4.2.3 Preference data mathematically combined with clinical or other data

Approaches that mathematically combine preference with clinical or other data could be used for medical product decision-making at the stage of marketing authorisation, HTA/reimbursement decision or post-marketing safety issue. These approaches include:

- Choice share
- Net clinical benefit
- Multi-criteria Decision Analysis (MCDA) / Stochastic Multi-Criteria Acceptability Analysis (SMAA) models
- Patient-preference based QALYs

These approaches are outlined below. There are numerous references with details on methodology and best practice.

- **Choice share:** Choice share is the probability that a treatment with a given profile will be chosen from among a set of alternatives, each with a different profile. In population terms, choice share estimates the proportion of the population for a treatment that, when provided equal access to several alternative treatments, would choose that treatment. The concept is similar to market share, but they differ in the types of information used in the estimate and the intent of the estimate. Choice share is generally based only on clinical endpoints and potentially tolerability measures (e.g. formulation, dosing frequency). Market share also may account for access, time on market, cost, insurance, and other issues. Choice share reflects benefit-risk amongst treatments and can be used to assess whether a substantial portion of the population for a treatment would choose to use that treatment, while market share is generally used for commercial decisions. If the treatments are study drug and no drug, choice share reflects whether patients would choose to use the study drug at all, and it has been used in this manner in at least one regulatory benefit-risk decision to date ([Ho, 2015](#)).

Choice share is calculated by mathematically combining preference and clinical data for a selected set of treatments. Using DCE as an example, choice shares can be estimated by simulating the utility of a set of alternatives using the standard deviations in a random parameter logit model and then calculating the proportion of times that a given alternative has the highest utility from among the set. A key consideration for choice shares is that they require the assumption of the form of the utility function that gives the probability distribution for the utility of an alternative. Within the five methods addressed below, DCE, BWS case 2 and swing weighting can be used to estimate utility functions. The threshold technique usually starts with two meaningful alternatives, so the choice shares are the proportions choosing each of those alternatives; however, a utility function like that for DCE can be used for the threshold technique. The degree to which choice share would change with different functional form for utility functions is an open research questions.

- Choice share can be assessed repeatedly, as new treatments become available or existing treatments are discovered to have higher or lower rates of known adverse events. This may be particularly useful in HTA and reimbursement decisions. However, if a new adverse event for the new or comparator treatments is discovered to be relevant and the existing preference study did not include that adverse event, the existing study cannot be used to re-assess choice share.

- **Net clinical benefit:** There are a variety of metrics that could be classified as a type of 'net clinical benefit' (NCB). Here, NCB is defined as a weighted sum of between-treatment differences in two or more dichotomous endpoints ([Singer 2009](#); [Hsu 2015](#); [Barnett 2018](#)). More formally, this NCB is the sum over all endpoints of the between-treatment differences in rate (e.g. incidence proportions, exposure time rate) for an endpoint multiplied by the corresponding weight (preference weight) for that endpoint. Related measures such as the win ratio and desirability of outcome ranking that use the ranking of endpoints or preference weights can also be considered NCB endpoints, though these are not covered here ([Evans 2015](#), [Pocock 2012](#)). Cardiovascular trials often use NCB endpoints, given the very low incidence of key events that are assessed in such trials (e.g., myocardial infarction, stroke, cardiovascular death).
- **Weighted sum NCB** is a simplified version of MCDA or SMAA methods in which all endpoints are dichotomous and their value function are assumed to be linear; i.e., a given change in rate is associated with the same change in preference, regardless of the baseline rate. Conceptually, this simplification is valid provided any non-linearity of each value function is small over the range of rates for the corresponding endpoint in both treatments. A more serious limitation of weighted sum NCB is that all endpoints must be dichotomous. In contrast, MCDA and SMAA can use any form of measurement for endpoints (e.g. dichotomous, categorical, continuous, probabilistic).

The interpretation of NCB in weighted sums depends on how the weights are scaled. For example, if a weight of 1.0 corresponds to death, then the weighted sum NCB is the preferential equivalent difference in risk or rate of death between the treatments. As with clinical trial endpoints, both types of NCB can be calculated with a 95% confidence interval and be used for formal statistical hypothesis tests.

The results of an NCB analysis include:

- Point estimates and 95% CI of the NCB measure between two treatments
- Decomposition of the components of the NCB endpoint, showing the contribution of each endpoint to the total (accounting for both weights and clinical data)

There are numerous sensitivity analyses that can be performed with NCB:

- Single-way weight sensitivity analysis: each weight is individually varied over a range (e.g. 25% to 75% percentile), and the change in NCB is assessed. Any changes that switch which treatment has positive benefit-risk suggests that the result is sensitive to the uncertainty in that weight. Visualization (e.g. tornado plot) are typically used. This is similar to the weight sensitivity analysis conducted in MCDA (described below).
- Monte Carlo simulation sensitivity analysis in which all weights are simultaneously varied over their ranges and the distribution of NCB is assessed. The probability mass below (or above) zero represents the probability that benefits exceed risks, taking into account all uncertainty in preference weights.
- Monte Carlo simulation sensitivity analysis in which all weights are simultaneously varied over their ranges and all clinical data are varied over their range. The distribution of NCB is assessed. The probability mass below (or above) zero represents the probability that benefits

exceed risks, taking into account all uncertainty in preference weights and clinical data. This is identical to the sensitivity analysis conducted in SMAA (described below).

- **MCDA / SMAA:** Multi-criteria decision analysis (MCDA) is a complete process for making a decision that spans framing the decision problem through to communicating the decision. The preliminary steps of MCDA and SMAA are similar to qualitative preference studies, i.e. identifying the relevant criteria for the decision to be made. In HTA and reimbursement decision making, the approach solves the problem that QALYs and cost-utility analysis do not capture all factors that are important to patients, providers or policy-makers, such as non-health outcomes or process characteristics. Such attributes may include attributes like waiting times and modes of access to healthcare, mode and frequency of treatment administration, and impacts on family members. Patient preference studies can provide insights into the relative importance and potential trade-offs between important patient-relevant outcomes, including health outcomes, non-health outcomes and process attributes. These can be mathematically combined and provide a patient-oriented MCDA within a larger policy-oriented MCDA ([Broekhuizen 2017, 2018](#)). The focus here is on the patient-oriented application of MCDA. Stochastic Multi-Criteria Acceptability Analysis (SMAA) is a specific type of quantitative MCDA, which allows for uncertainty in all inputs and provides probabilistic results. Within both quantitative MCDA and SMAA there are steps to generate a value function, which translates the values of attributes into a normalized (e.g. 0 to 1) scale, and weights that are applied to these normalized scales for each attribute.
- **MCDA** - Any MCDA, be it quantitative or qualitative, starts with the same three steps: defining the decision problem, selecting the decision criteria and constructing the performance matrix ([Thokala 2016](#)). The performance matrix presents an assessment of an intervention against each of these criteria, such as health outcomes (e.g., life years gained), non-health outcomes (e.g. ability to work) and process attributes (e.g., inconvenience of treatment, frequency of hospital visits).

Quantitative MCDA uses a value measurement model to interpret the performance matrix. This approach continues with five further steps after the initial three steps. Firstly, preferences are elicited to specify a value function for each criterion, which translates a technology's performance on that criterion into a score, e.g., between 0 and 100 or between 0 and 1. Secondly, preferences regarding the relative importance of criteria are measured and used as criterion weights. Thirdly, the performance scores for each criterion are multiplied by the relative weight of that criterion, and the weighted scores are summed to obtain an overall value for the intervention ([Baltussen 2019](#)). Interventions are then ranked on the basis of these overall values. Fourthly, uncertainty analysis is performed to understand the level of robustness of the results ([Broekhuizen 2015](#)).

- **SMAA** - Constructing an SMAA model is nearly identical to constructing an MCDA. The main differences are that uncertainty in all the attributes or criteria and in all the weights are incorporated. SMAA also allows for a covariance structure between weights, so that they need not be preferentially independent as is required in MCDA. SMAA can generate the same results as MCDA, but in a probabilistic manner. For example, rather than single scores for each alternative treatment, there are probability density functions over the scores. Instead of alternative A being better than alternative B by some fixed amount on the MCDA output's

scale, an SMAA can generate the probability that treatment A's benefits outweigh its risks when compared to treatment B.

Because there is always considerable uncertainty in the clinical endpoints used in MCDA/SMAA models, SMAA is generally better aligned with the needs of medical treatment decisions.

- **QALYs:** In HTA "Quality-adjusted life years" (QALYs) are frequently used as an effectiveness measure in cost-utility analyses. QALYs are calculated by weighting life years gained with a value reflecting the health-related quality of life (HrQoL) in these years. The HrQoL value is expressed on a 0 to 1 scale, where 0 reflects the value of the state 'dead' and 1 the value of the state 'perfect health'. The values are obtained from preference studies, either from patients or from the general public. By weighing the life years gained from a treatment with the HrQoL value in these years, the number of QALYs gained from treatment is obtained. A year of life lived in perfect health is worth 1 QALY.

QALYs are a composite endpoint of HrQoL and life years gained from a treatment compared to its best alternative.

### 3.4.3 Incorporating preference information into industry, regulatory and HTA/payer documents

PREFER suggest that if preference study results are intended to inform a regulatory or HTA body/payer decision, then the corresponding data should be included in the industry submission package, the regulatory or HTA body assessment documentation and potentially the drug label. This approach would provide more transparency about how patients' perspectives have informed decision-making, and - for preference study information included in labels - would provide information to patients and prescribers that could assist their decision-making.

Most of the proposed approaches align with current processes documented either in [ICH M4\(R2\)](#), [EMA Day 80 Assessment Report](#), or [FDA reviewer guidance documents \(ICH M4\(R2\); EMA; FDA 2019\)](#). Suggestions related to placement within labeling reflect the approach taken within the [FDA Patient Preference Information guidance \(2016\)](#), which states that preference data supporting approval should be described in the device label, and advises that the inclusion of preference data in decision summaries can be helpful to both healthcare professionals and patients who need to make tricky benefit-risk trade-off decisions.

#### 3.4.3.1 Inclusion of preference study results in an industry submission package

Preference study results could be incorporated into a sponsor's submission package in two places:

- **High-level preference study results: within the Clinical Overview (eCTD section 2.5)**

Preference study results intended to support a marketing authorization application could be included in a Clinical Overview. For example, the results of a preference study assessing patients' views on patient-relevant outcomes could be included in [section 2.5.1](#) 'Product Development Rationale' and/or "therapeutic context" to characterise the unmet need. Furthermore, as described earlier in this document ([section 3.4.1.1](#)), a preference study assessing patients' views on patient-relevant outcomes could inform the choice of endpoints included in a submission study, and these submission

study endpoints could have an influence on the selection of key outcomes described in [section 2](#). The results of a preference study assessing patients' views on an acceptability of benefit-risk trade-off issue, or an acceptability of uncertainty issue, could have an influence on the choice of benefits in CTD section 2.5.6.2 'Benefits' and the choice of risks in CTD section 2.5.6.3 "Risks", as well as being included in CTD section 2.5.6.4 "Benefit-risk assessment". Preference study report: within CTD section 5.3.5.4 "Other study reports" of the eCTD.

### **3.4.3.2 A preference study report could be included within eCTD section 5.3.5.4. Inclusion of preference study results in the European Public Assessment Report (EPAR)**

For a regulatory decision where preference data played a key role in the approval of the product and/or may make a change to the use of the product by an individual prescriber, PREFER suggest that preference study results could be included in the appropriate section(s) of the European Public Assessment Report.

For example, for a preference study providing patients' views about the most important attributes of a specific disease / medical product, the preference study results could be mentioned to support the description of unmet need the section describing "Available therapies and unmet medical need". Furthermore, as described earlier ([section 3.4.1.1](#)), results from this type of preference study could inform the choice of endpoints included in a submission study, and these submission study endpoints could have an influence on the choice of favourable effects included in an Effects Table. For a preference study providing patients' views about the acceptability of a benefit-risk trade-off or the acceptability of uncertainty, the preference study results could be included in the section describing "Balance of benefits and risks" and/or in the section describing "Additional considerations on the benefit-risk balance". One further option would be the inclusion of preference weights into the effects table in the section "Effects table".

Prior to inclusion in the European Public Assessment Report, preference results could also be included in a Day 80 Assessment Report. This would be consistent with the advice in the [EMA Day 80 Assessment report template](#) about consideration of patient input.

### **3.4.3.3 Inclusion in the labelling documents (e.g. SmPC)**

For a regulatory decision where preference data played a key role in the approval of the product and/or may be of relevance for the prescriber and the patient when deciding on the prescription, PREFER proposes that preference data could be included in the 'Clinical efficacy and safety' sub-section of SmPC section 5.1 'Pharmacodynamic properties'. (Alternatively, preference data could be included in a new 'Patient Experience'" sub-section of the SmPC section 5.1.)

The SmpC could include details of the preference data informing the regulatory decision, structured as follows:

- **Summary of the situation prompting the preference study**
- For an 'acceptability of trade-off' or 'acceptability of uncertainty' scenario, a summary of the situation could be provided by describing the study purpose or the primary research question.
- **Description of the preference study design and population**

This would be aligned with the approach typically taken in the description of clinical data, in which the SmPC template ([European Commission Volume 2C Notice to Applicants - SmPC, 2009](#)) expects information on “the main characteristics of the patient population”.

- **Summary of the preference study results**

This would also be aligned with the approach typically taken in the description of the clinical data, where the SmPC template expects this section to provide ‘evidence from relevant studies.

An example of how preference information has been described in a U.S. label is shown in [Table 3-23](#).

**Table 3-23 Example of preference data included in labelling information**

<p><b>Study used as a basis for preference information in the label</b></p>	<p>As described by Rummel M , 2015, patients with diffuse large B-cell lymphoma were asked for their preference for different formulations of rituximab (subcutaneous or intravenous) at the end of a cross-over study in which patients had direct experience of both modes of administration.</p> <p>(This is a preference study based on direct patient experience, rather than patients' stated preferences.)</p>
<p><b>Description of patient preferences within the U.S. label</b></p>	<p>Previously untreated adult patients outside of the United States with CD20+ diffuse large B-cell lymphoma (DLBCL) or CD20+ follicular non-Hodgkin’s lymphoma (FL) Grades 1, 2, or 3a, were randomized to receive a standard chemotherapy regimen (CHOP, CVP, or bendamustine) and either RITUXAN HYCELA 1,400mg/23,400 Units at Cycles 2–4 (after the first cycle with intravenous rituximab) or a rituximab product by intravenous infusion at Cycles 1–4. After the fourth cycle, patients were crossed over to the alternative route of administration for the remaining 4 cycles. After Cycle 8, 477 of 620 patients (77%) reported preferring subcutaneous administration of RITUXAN HYCELA over intravenous rituximab and the most common reason was that administration required less time in the clinic. After Cycle 8, 66 of 620 patients (11%) preferred rituximab intravenous administration and the most common reason was that it felt more comfortable during administration. Forty eight of 620 patients (7.7%) had no preference for the route of administration. Twenty nine subjects of 620 (4.7%) received Cycle 8 but did not complete the preference questionnaire.</p>

**3.4.3.4 Inclusion of preference study results in published information about HTA body / payer decisions**

For an HTA/payer decision where preference data played a key role, and where the HTA body/payer publish documentation that provides a background to their decision-making process, PREFER would also suggest that the preference study results could be included in such published documentation.

In particular for HTA, an additional assessment element on patient preferences could be added to the HTA Core Model (<https://eunetha.eu/hta-core-model/>) in the clinical effectiveness domain. This domain currently includes assessment elements such as impact of the intervention on health-related quality of life, morbidity and patient satisfaction, which could be valuably complemented by information from patient preference studies. In the HTA Core Model, patient satisfaction currently



refers to “patients’ overall perception of the value of the intervention and their satisfaction with the treatment”. It is used to assess acceptability of the intervention and prediction of overall uptake (referred to as ‘choice share’ in the PREFER Framework), and assessed by means of surveys, qualitative research, observational studies and trials. Patient preference studies could be added as a type of study that can inform overall uptake estimates. In several other domains of the HTA Core Model reference is made to patient preferences (i.e. in the ‘economic’, ‘ethical’, ‘organisational’ and ‘patients & social aspects’ domains), but never to patient preference studies.

PREFER recommends to include a specific assessment element in the clinical effectiveness domain, to allow assessors to describe how patient preference studies informed relative effectiveness assessments in a more specific way, in addition to the relevance of patient preferences for the other HTA domains.

## 4 Methods

### 4.1 Background on the method selection process

This section presents five methods (DCE, best-worst scaling [BWS] case 1, BWS case 2, threshold technique [TT], and swing weighting) that can be applied in conjunction with the framework to assess patients’ perspectives. The resulting data could then be incorporated into sponsor, regulatory, HTA, and/or payer decision-making.

#### 4.1.1 Identifying preference methods

Accounting for the particular research questions planned for the PREFER case studies (Table 4-1) and their popularity for benefit-risk assessment, five methods (DCE, Best-Worst Scaling Case 1 and 2, Threshold Technique, and Swing Weighting) were selected as most likely to meet most decision-makers’ needs during all stages of the MPLC. In the following sections, each method is described in terms of:

- Background and overview to the method
- Assessment of the suitability of the method for answering different research questions
- Assessment of the method from the perspective of different stakeholders
- Description of practical considerations of conducting a study with the method
- Method-specific principles of preference elicitation, instrument design, and data analysis tailored from the general framework
- Tests for and evidence of validity specific to the method.

#### 4.1.2 PREFER case studies and methodological research questions

Within PREFER, numerous research questions relating to preference studies were identified based on input from stakeholders, and then distilled into those ranked as highly important by experts (task 2.8). This distillation together with identification of methods described led to the final focus of the core case studies with respect to methodological research questions and methods included (see Table 4-1).

**Table 4-1 Summary of methodological research questions and methods in PREFER case studies**

Methodological research question	#	Number of times the method is used across PREFER case studies	#
Comparing methods	8	Probabilistic Threshold Technique	2
Impact of changes in numbers, type, and definitions of attributes	1	Best-Worst Scaling case 1	2
Generalizability of preferences across populations	7	Best-Worst Scaling case 2	1
Impact of educational material on understanding of tasks and preferences	5	Swing Weighting	3
Can psychosocial constructs provide insight in individual preferences	10	Q-methods	1
What tests assess if patients can perform a cognitive task	9	Discrete Choice Experiment	10

## 4.2 Discrete Choice Experiment (DCE)

### 4.2.1 Introduction

DCE has been increasingly used to quantify patients' preferences for health outcomes, health services, and medical treatments (Soekhai 2019a; Clark 2014; De Bekker-Grob 2012). Hauber (2013) described a conceptual framework for applying these methods to benefit-risk decisions. The Center for Devices and Radiological Health at the FDA has demonstrated how data from a DCE can quantify patient preferences in a manner that can support regulatory approval decisions (Ho 2015), and these data were used to support the regulatory approval of a device to treat obesity (Al-Faruque, 2015). In addition, DCE results can be used to support HTA and payer decisions (Mühlbacher, 2015; Mühlbacher and Johnson, 2017).

DCE is a utility-theoretic method that can be used for eliciting preferences for medical interventions (Hauber 2013). The DCE is based on the hedonic principle that products or services are evaluated based on their attributes and how the products perform on each attribute (attribute level), and that an individual's choice of a product or service is a function of the levels of the attributes that define it.

Attributes and attribute levels are chosen to represent the health outcomes and features of medications, devices, and health care services that are relevant to a treatment decision. Before a patient can make a choice among alternative treatments, in which the attributes and levels are varied, it is important that the attributes and levels are, a priori, expressed in terms of a relevant context (i.e., a vignette) that is understandable by the patient. Once the attribute and attribute levels have been determined, the attribute levels are used to create sets of hypothetical scenarios or treatment profiles. Each respondent is presented with a series of choices among sets of hypothetical treatment or product profiles and the pattern of choices made over the series of choice questions can be used to statistically infer the rates at which respondents are willing to trade off each attribute for the others. The hypothetical profiles and sets of hypothetical profiles are determined by an experimental design with known statistical properties (Johnson 2013) that allows the estimation of a unique preference parameter for each attribute level and potentially interactions among attributes.

It is noted that the DCE and BWS multi-profile case (also called "case 3") are closely related (Flynn et al., 2007). While DCEs usually require respondents to select their preferred (best) profile, BWS case 3 requires them to also select their least preferred (worst) profile. Because DCE and BWS case 3 share many parallels in terms of the research question, theoretical underpinnings, choice of attributes and levels, experimental design, and analyses of best data, they will be treated as equivalent and only BWS case 1 and case 2 are explored in this document.

Best-worst scaling case 3 is another type of preference-elicitation technique that is similar to the DCE. In a DCE, the respondent is asked to choose the preferred alternative from a set of two or more options as described above. In best-worst scaling case 3, the respondent is asked to choose the most and least preferred alternatives from a set of three or more options. The DCE and best-worst scaling case 3 are constructed and analysed similarly and yield similar types of results. For the purpose of this document, we treat the DCE and best-worst scaling case 3 to be equivalent.

Each preference parameter indicates the relative contribution of each attribute level to the probability of choosing an alternative with that attribute level from all possible combinations of attribute levels.

McFadden (1978) has shown that, appropriately modelled, the DCE is consistent with random utility theory and that the resulting parameter estimates are measures of the marginal utilities of the set of attribute levels. These marginal utility estimates can be used to estimate the marginal rates of substitution (rates of trade-off) among attributes, the importance of one attribute relative to all other attributes included in the DCE conditional on the ranges of the levels of the attributes, and to estimate choice share (e.g. see Ho 2015).

These measures can thus be used to estimate:

- the relative importance of treatment attributes,
- the maximum level of treatment-related risk that patients would be willing to accept to achieve a given level of treatment benefit or an improvement across a group of benefit attributes (Hauber 2015; van Houtven 2011),
- the minimum level of treatment benefit patients would require to accept a given set of treatment-related risks (Hauber 2013) and
- choice shares - the probability that the combinations of attribute levels defining a given treatment are preferred to the attribute levels defining a different treatment or standard of care (which can be interpreted as the probability that the benefits of that treatment exceed the risks relative to an alternative treatment or standard of care).

When out-of-pocket cost or some other payment vehicle is included as an attribute in the DCE, the results can be used to calculate the marginal economic (i.e. monetary) value to patients of changes in attribute levels or the economic value of a treatment relative to an alternative treatment or standard of care (Ryan & Wordsworth, 2003).

In some preference scenarios, it may be relevant to offer the respondent an “opt-out” option, when neither of the options offered is attractive/acceptable (Determann, 2019). This option may, for example, be “neither”, or “standard of care”, or “current treatment”. Including an opt-out alternative may make the context of the choice more realistic if no treatment or no change in current treatment is a realistic option for a patient. In addition, an opt-out alternative may be necessary if the objective of the study is to predict uptake of the medical product.

DCEs yield cross-sectional choice data for which there are multiple responses for each respondent. The theoretically correct method for analysing such data is a limited-dependent variable model in which each choice is regressed on the characteristics of the alternatives in the choice set (i.e., the levels of the attributes for each profile in the choice set). The basic utility-theoretic model for analysing these data is the conditional multinomial logit model (McFadden, 1978), although this assumes that all respondents have the same preferences and that each choice from a single respondent is independent of all other choices from that same respondent. To account for heterogeneity of preferences across the sample and the panel nature of the data, alternatives to the conditional multinomial logit models are often used. Two commonly used alternatives to the basic model are the random-parameters logit model and latent class finite mixture models (Hauber 2016; De Bekker-Grob 2012, Soekhai 2019a). The random parameters logit model assumes a continuous distribution of preference heterogeneity across the sample by estimating both mean marginal preference parameters and a parameter for the standard deviation of each preference weight across the sample. The latent class finite mixture model assumes a discrete distribution of preferences

across the sample by identifying segments within the sample with similar patterns of choice and thus similar preferences. Separate mean coefficients are estimated for each attribute level of each segment and the probability that each respondent in the sample is characterized by the pattern of choices in each segment is calculated.

Estimating preference heterogeneity can be used to simply control for unobserved preference heterogeneity within the sample (heterogeneity due to unobserved covariates) or it can be used to determine whether specific characteristics of individuals in the sample (e.g., demographic characteristics, disease state and treatment experience) explain systematic differences in patients' preferences. These results can then be used to determine the extent to which the preference measures generated by the discrete-choice experiment differ for different subgroups within the sample.

## **4.2.2 DCE-specific aspects of framework component 2 “design: method selection and analysis planning”**

### **4.2.2.1 What type of research question is appropriate for a DCE?**

A DCE is typically used in decisions where the alternative treatment options are characterized by multiple attributes, where some attributes favour one treatment and other attributes favour other treatments, and clinical judgment is insufficient to defensibly assess the benefit-risk trade-offs amongst them. The output from a DCE can quantify the relative importance of these attributes and which trade-offs are acceptable. DCEs are especially helpful when these trade-offs are complex, such as when the relative importance depends on the baselines for the attributes (technically, non-linear value functions) or when there are dependencies between the preferences of attributes (technically, preferential dependence).

One good example of this is given in the paper by [Bridges \(2012\)](#). A DCE was used to evaluate patients' preferences for treatment outcomes in advanced non-small cell lung cancer (NSCLC). The research question was clearly posed as “What are the benefits patients judge sufficient to compensate for different levels of the risks associated with therapy for NSCLC?” The DCE had multiple attributes, each with multiple levels and investigated the benefit-risk trade-offs .

One advantage of including multiple attributes in a patient-preference study that vary simultaneously is that interactions among the attributes can be investigated. In the example above, the two-way interaction between disease symptoms and progression-free survival was significant, which implied that if the disease was severe enough, patients would choose a shorter lifespan.

Another good example is given in the paper by [Janssen \(2018\)](#) in which the research question considered whether benefits and harms of diabetes medication played a role in patients' treatment decisions. One particular feature of this example is that it was intended to illustrate the process of explicitly following the steps given in the ISPOR Good Practice checklist described by [Bridges JF et al. \(2011\)](#).

### **4.2.2.2 When is a DCE an appropriate method?**

Situations that suggest a DCE would be appropriate include:

- When there may be interactions between the preference of attributes (preferential dependence). Assessing interactions requires that the choice tasks depict multiple attributes simultaneously. Most methods (e.g., swing weighting, threshold technique, standard gamble, time trade-off) can only compare one attribute directly against another or against a single health state, so any dependencies (interactions) among the attributes are ignored. A DCE can consider all attributes simultaneously.
  - When the assessment needs to be made in a context similar to that when a patient is asked by a physician to choose between several treatment options. The physician first sets up the context in the form of a vignette (the patient's medical history and current condition) and then describes the potential benefits and risks of the alternative options in terms of their attributes. While the choice tasks in a DCE involve hypothetical scenarios, these scenarios can be designed to be very similar to those in patient-physician shared decision-making.
1. When it is of interest to obtain patients' preferences for treatments or services that are possible but not yet available. Because of the hypothetical nature of the choices, DCEs naturally allow for the use of benefits and risks, and their combinations, that are not relevant to available treatments but are critical for those under development or being considered. DCEs also allow for the levels of these benefits and risks to include those expected for future treatments. For the same reason, DCEs allow computing benefit-risk trade-offs between current and future potential treatments.
    - When needing to deal with complex patterns of uncertainty (e.g., by using mixed models, Bayesian methods, etc., which are now well-developed for the analysis of DCE data)
  2. When patients vary in a population in terms of their preferences (i.e., there is preference heterogeneity) and it is important to identify both decision-relevant sub-groups within which there is relative homogeneity of preferences as well as the heterogeneity within these subgroups. Preference sub-groups can be particularly important to identify sub-groups that are willing to accept more risk than others or require a greater benefit to accept a given risk than others.

When the assessment needs to be made in a context similar to that of a patient being asked by a physician to choose between several treatment options. Considering these capabilities, a DCE can be considered as one of the most complete methods currently available to assess preferences ([Soekhai 2019a](#)) in the sense that it can incorporate nearly all of the features that other quantitative methods have, but all within the same approach. However, conducting a DCE may be resource intensive, put a high cognitive burden on patients, and require complex analysis methods that may be difficult for end-users to understand. These points and others are listed in the following sub-section.

#### **4.2.2.3 When is a DCE not an appropriate method?**

There are situations where a DCE would be inappropriate or excessive:

- the research question lends itself to a simpler method (e.g., if there is just one benefit and one harm, in which case the probabilistic threshold technique or other methods are simpler, and/or generally quicker to complete)
- too little is known about the specific benefits or risks in the treatment alternatives (in which case a better option may be a less rigorous but faster and less resource intensive method,

reserving the DCE for later when more detailed information about the attributes and their ranges are known)

- the cognitive burden on patients would be excessive
- there is insufficient time or budget to conduct a DCE
- the research question, even if complex, can be defensibly addressed using clinical judgment (true for any preference assessment method)
- there is a lack of available expertise to plan, design, run and analyse a DCE (statistician, survey methodologist, psychologist, etc.)
- the trade-offs a specific individual is willing to make are required.

#### **4.2.3 Points to consider in choice of preference elicitation method: i.e. why DCE and not something else?**

DCE is often the preferred method when the objective of the study is to estimate benefit-risk trade-offs among multiple benefits and risks simultaneously. When the research question only requires a ranking of attributes, estimates of the relative importance of attributes, or the trade-off between two attributes, then simpler or more direct methods may be enough.

##### **4.2.3.1 Points to consider regarding the analysis planning for DCE studies**

- The ISPOR Task Force report ([Hauber 2016](#); [Lancsar 2017](#)), give descriptions of the various models that can be fitted and what software can be used to fit these models and describe the advantages and limitations of various analysis approaches.
- The choice of modelling approach depends on the research questions, the study design and any constraints in terms of quality and quantity of data.

[Hauber \(2016\)](#) provide a checklist of questions to consider when justifying the choice of analysis method, describing the analysis and interpreting the results.

##### **4.2.3.2 Expected timeline for conducting a DCE**

Expected timeline for conducting a DCE strongly depends on contracting between the relevant parties, how long it takes to develop the DCE, including determining the attributes and their levels, which often involves a pre-period of qualitative research (e.g., online and structured interviews), and how long it will take to recruit subjects and collect the DCE survey data. The total length of this period could vary, from two to three months for simple studies to up to more than 12 months, for joint qualitative/quantitative studies. Therefore, expected overall timelines can be highly variable for each individual study, which should be taken into account when planning to incorporate a DCE into the overall clinical program.

As a rough estimate, most sponsor-conducted qualitative/quantitative preference studies done in collaboration with an academic group or a consulting company take one to two years.

### 4.2.3.3 Points to consider for DCE internal validity testing

There are numerous methods that can be considered informative on internal validity in preference studies. Some tests for validity and reliability primarily apply to DCEs and similar survey-based methods ([Johnson 2019b](#)), including:

- **Stability:** With a repeated choice set, testing if the same objects are consistently chosen as best and/or worst
- **Within-set dominated pairs:** Including a choice task where one alternative is unambiguously better for all attributes and assessing whether the respondent chooses the better alternatives
- **Across-set dominated pairs:** A generalization of the within-set dominated-pairs test that is based on two choice tasks (see [Johnson 2019b](#))
- **Transitivity:** Testing whether, if object X is worse than object Y, and object Y is worse than object Z, then object X is worse than object Z in a subsequent choice set (holding all other attribute levels constant)
- **Attribute dominance (non-compensatory preferences):** Respondents should be willing to accept a reduction in one desirable attribute in return for a sufficiently large compensating increase in another desirable attribute. Attribute dominance is an observed non-compensatory pattern in which respondents choose the alternative with the better level of one attribute in all or nearly all choice questions
- **Straight-lining or flat-lining:** Testing if respondents make a selection based on object location rather than information

As noted in the Framework, “failing” these tests is not always a definitive indication that survey results are not valid. For example, the alternatives in a repeated choice task may have very similar utilities, causing respondents to be uncertain which is better and to answer the two tasks differently. Similarly, the two alternatives in a within-set dominated pair may have very similar utilities, causing respondents to be uncertain which is better. Attributes may appear to be dominant if the ranges of levels for other attributes are insufficient to cause trade-off behavior. At times, these issues can be identified in advance by good pretesting or a pilot survey, though not in all cases.

## 4.2.4 DCE-specific aspects of framework component 2 “design: sample definitions – justifying the sample size”

### 4.2.4.1 Points to consider regarding the sample size of a DCE

The sample size depends on the complexity of the DCE design, the magnitude of the between-patient variability and the desired precision of the estimated effects. The complexity of the design increases as the number of attributes and the number of levels of each attribute increases (in which case it may be necessary to reduce the number of choice tasks given to each subject to reduce the cognitive burden). A desire to estimate interactions between attributes will also increase the complexity of the DCE.

Methods for constructing a DCE design are given in [Johnson \(2013\)](#).



- If a formal sample size calculation is needed to test a particular hypothesis, the methodology given by [De Bekker-Grob \(2015\)](#) may be used; however, the information required to apply this methodology may not be available a priori, especially in cases where there is no information on which to predict utility differences (i.e., effect sizes.).
- In situations where formal hypothesis testing is not of direct interest, which is often the case (or the information needed for the method mentioned above is not available), the sample size is calculated based on previous experience and established rules of thumb (see, [Marshall 2010](#) and [Bridges 2012](#), for example). If a formal power and sample size calculation is needed, it is often done as a supplementary analysis of a completed DCE to gain useful information for the planning of future studies.
- [Yang, Johnson \(2015\)](#) give a meta-analytic review of sample sizes used in 32 DCE patient preference studies. They consider more than just hypothesis testing but, in particular, consider the empirical joint effects of sample size and study design characteristics on utility difference precision. They give an empirical formula for sample size: a sample size of 250 subjects is not atypical.

#### **4.2.5 DCE-specific aspects of framework component 2 “design: preference question design – consideration of the appropriate number of attributes and attribute levels; patient burden issues”**

##### **4.2.5.1 Points to consider regarding the choice of attributes**

The choice of attributes should be determined by the research question. The number of attributes should be limited to those that are required to answer the research question and the number that respondents can actually consider simultaneously.

A balance needs to be struck between what is important to the respondent and what is relevant to the policy - or decision-making environment and guided by the research question.

When choosing the attributes, at least two perspectives need to be considered:

- that of benefit-risk science - what are the key benefits and risks of the product/treatment and
- what is important to patients.

If the DCE is to be conducted for a benefit-risk evaluation of a specific product, the choice of attributes must recognize that the wishes of patients need to be consistent with what the product can offer. This will require consultation with clinical experts, qualitative researchers and usually, the consideration of the results of preliminary studies.

##### **4.2.5.2 Points to consider regarding the choice of attribute levels**

The number of levels should be sufficient to capture the trade-offs of interest (the ranges of levels should be large enough to induce trade-offs (e.g., risk probabilities cannot be so low that respondents ignore them when making their choices) and include the levels that have been seen or would be expected in the real world. The appropriate number of levels also depends on the nature of the model

to be fitted. Two levels allow a linear trend in the levels to be detected, three levels allow a quadratic trend to be detected, and so on.

Choosing the number of levels involves multiple considerations, including the number of levels for the other attributes (there is evidence that having a larger number of levels for one attribute than for others may draw greater attention to that attributes) and types of level. There are three types of attribute level:

- numeric (i.e., time, probability, ) that are continuous and naturally ordered;
- categorical and naturally ordered (e.g. mild, moderate, severe); and
- categorical and not naturally ordered (e.g. red, blue, and green).

The considerations involved in selecting levels will depend in part on which of these types, or mixture of types, applies to the DCE under consideration. There are often good reasons to include unrealistic levels in a DCE; however, dramatically unrealistic levels might lead to over - or under-estimation of actual preferences. The distances between the levels should allow the recovery of any level of interest between them (assuming that the levels are numeric).

The number of levels will have a significant impact on the complexity of the experimental design and ultimately the sample size required and/or the number of questions each respondent will need to answer.

#### **4.2.5.3 Points to consider in the set-up of choice profiles (experimental design)**

The experimental design refers to the specific combinations of benefits and risks used in the choice tasks that responders complete. Since most benefit-risk problems entail far more combinations than can possibly be asked in a preference survey, a carefully-designed experimental design ensures that the set of combinations for benefits and risks is covered completely, uniformly and in sufficient density to measure preferences while not requiring an excessive number of tasks for each responder. There are many resources on this topic, including a best practice guidance (e.g. [Johnson 2013](#)) and several software packages to generate good experimental designs (e.g. SAS).

#### **4.2.6 Summary**

From the most promising elicitation methods, DCE was selected for qualification given the desirability of using a method with a strong theoretical background, and one appropriate for eliciting trade-offs in a multi-attribute preference context. Further support for it is its increasing use in quantifying preferences in health research, within regulatory benefit-risk related decisions, and the applicability of results to HTA and payer decision-making ([section 3.1](#)).

### **4.3 Swing Weighting**

#### **4.3.1 Introduction**

SW is a preference elicitation method that obtains respondents' trade-offs for changes between attributes. The trade-offs are elicited directly from individuals in a manner which enables for the analysis of individual-level preferences. This is in contrast to some other preference elicitation

methods such as discrete choice experiments (DCEs), which elicit preference statements that are used as inputs to a preference model, which then provides the trade-offs as outputs (i.e., trade-offs are elicited indirectly), and may not allow for as precise individual-level analyses (Tervonen et al., 2017). While individual-level predictions (from model estimates) can be obtained from DCEs using choice models that account for heterogeneity (de Bekker-Grob et al., 2020), it does not enable individual-level trade-off data for each individual in the sample as can be obtained from SW.

The typical SW procedure consists of two stages. In the first stage, respondents are asked to rank importance of changes in attributes (i.e., “swings”) from the highest to the lowest. This is followed by a second stage wherein the respondents are asked to judge the relative value of the attribute swings. The most common method is by assigning a value of 100 to the highest ranked attribute, and then asking respondents to express the value of the second highest ranked attribute swing as compared to the highest ranked swing, with a value between 0 and 100. The process is then repeated for all attributes and the resulting weights normalised to sum to a constant, typically 1 or 100, to obtain trade-off weights that express the relative importance of attribute scale swings (Angelis and Kanavos, 2017, Department for Communities and Local Government: London, 2009, Medical Device Innovation Consortium, 2015, Mussen et al., 2007, Olson, 1996). Other approaches to obtaining weights include taking a total of 100 points then dividing it as a total to divide among all swings.

The basic SW procedure, which is typically conducted in an interviewer-facilitated in-person setting, only captures the trade-offs respondents make between attributes. SW is often paired with a scoring procedure to capture preferences for changes within attributes; that is, potential non-linearities of the partial value functions (Marsh et al., 2016). In scoring, respondents make value judgements on the incremental changes in each attribute to determine the partial value function (Department for Communities and Local Government: London, 2009). This is in contrast to methods such as DCE, BWS case 2 and the threshold technique, in which the preferences reflect a combination of the importance weight (corresponding to the full swing) and the value function.

The terms SW and multiple-criteria decision analysis (MCDA) are often used interchangeably, although there is a distinct difference between the two approaches. While SW refers to the elicitation of trade-offs, MCDA is a decision-making framework that enables individuals and groups to come to a consensus by assessing multiple benefits and risks by combining judgements and data. SW, or other forms of assessing weights, is a step within MCDA. Additionally, MCDA applies decision theory to decisions with multiple objectives, enabling for the appraisal of treatments with multiple attributes and combining them into a single overall appraisal (Keeney and Raiffa, 1993). SW and data from other preference elicitation methods capturing trade-offs, such as DCE, can be used to inform MCDA assessments (Marsh et al., 2014).

MCDA methods including SW have been widely used in the public and private sector, such as for policy decision-making in areas such as transport, education, environment (Gregory et al., 2012). The original swing weighting technique, SMART (Simple Multi-Attribute Rating Technique), was proposed as a method for eliciting multi-attribute utility in an individual or a group within a public policy context. It was purported that multi-attribute utility measurement enables decision-making or regulatory agencies to shift their focus from the actions being regulated to the values these actions served (Edwards, 1977, Edwards and Barron, 1994, Olson, 1996). The healthcare sector has been slower to adopt the MCDA approach (Diaby et al., 2013, Marsh et al., 2014). A 2014 literature review

of MCDA within the healthcare industry showed that only 7.3% of MCDAs conducted used SW to elicit weights (Marsh et al., 2014)

#### 4.3.1.1 Use of Swing Weighting within healthcare sector

Within healthcare, SW has been used for a range of purposes that can have an impact from a medicines regulatory and access perspective (Marsh et al., 2014, European Medicines Agency (EMA), 2008, European Medicines Agency (EMA), 2009, European Medicines Agency (EMA), 2012), as well as from societal and health policy perspectives (Nutt et al., 2010, Nutt et al., 2014) and for eliciting patient preferences (Marsh et al., 2017b, Vermersch et al., 2019).

In the regulatory context, swing weighting has been used to inform quantitative benefit-risk assessments (BRA) (Tervonen et al., 2019a) with stakeholders such as regulators, experts and clinicians. The European Medicines Agency (EMA) have conducted a benefit-risk project and concluded that MCDA is a suitable framework for quantitative benefit-risk assessment uses (European Medicines Agency (EMA), 2008, European Medicines Agency (EMA), 2009, European Medicines Agency (EMA), 2012).

In the health technology assessment (HTA) context, SW has been used to assess the weight and value placed on the burden of disease, therapeutic impact, safety profile, innovation level and socio-economic impact which has then been used to create generic value models that can be adapted and applied across different decision-making contexts (Angelis and Kanavos, 2017, Felli et al., 2009, van Valkenhoef et al., 2012). For example, MCDA was used to evaluate an integrated care programme for people with multi-morbidities in eight European countries (Rutten-van Molken et al., 2018). Ordinal SW was used to elicit weights for the different assessment criteria from 5 different stakeholder groups: patients, partners and other informal caregivers, professionals, payers and policy makers. An MCDA framework has also been developed from the Sustainable Integrated Care Models for Multi-Morbidity: Delivery, Financing and Performance (SELFIE) project, aimed at improving person-centred care for people with multi-morbidities and which uses both SW and DCEs to elicit weights (Leijten et al., 2017)

SW can also be used to inform public health strategy by eliciting values from expert groups (Nutt et al., 2010, Airoidi et al., 2011). In the UK, members of the Independent Scientific Committee on Drugs participated in an assessment of the harms caused by misuse of alcohol and illicit drugs, with the aim of informing UK healthcare and public policy (Nutt et al., 2010). This assessment has been replicated in the EU with similar findings (van Amsterdam et al., 2015). Similarly, an international expert panel used SW to assess the different types of harm from nicotine-containing products (Nutt et al., 2014). SW has also been used to elicit benefit-risk preferences of treatments for physicians and other experts (Marsh et al., 2017b, Vermersch et al., 2019), which may have practical implications such as to help prescribers make objective decisions about appropriate treatments to recommend to individual patients (Vermersch et al., 2019).

There has been a paucity of published studies using SW to elicit patient preferences. This may reflect the origins of the MCDA methods, which were traditionally used to produce a consensus in decision-making, rather than for analysing population-level preferences (Tervonen et al., 2017). Some studies have applied an online, swing weighting-inspired procedures (Marsh et al., 2019, Postmus et al., 2016, Postmus et al., 2018, SriBhashyam et al., 2019).

MCDA has also been used in clinical practice as personal decision support tools to facilitate shared-decision making (Kaltoft et al., 2018, Wagner et al., 2018). Other than the standard approach described above, there are also some variations of the SW method, including imprecise SW and choice-based matching. also known as adaptive swing weighting. Imprecise SW enables decision-makers to provide imprecise perceived value estimates using a range estimate rather than a single point estimate (Tervonen et al., 2015). This approach aims to account for the behavioural biases from elicitation techniques that result in a single exact weight (Weber and Borcherding, 1993). Choice-based matching (Postmus et al., 2016, Postmus et al., 2018) also known as adaptive SW (Marsh et al., 2019, SriBhashyam et al., 2019), involves a ranking exercise followed by a thresholding approach to preference elicitation.

#### 4.3.1.2 Guidance on Swing Weighting

A range of guidance documents are available on using SW, which are predominantly focused on the use of SW within an MCDA in regulatory, HTA and expert decision-maker settings. The ISPOR MCDA Task Force guidance has a two-part guidance, with part 1 providing an overview and definitions of the steps involved in decision-maker MCDA processes and part 2 providing good practice guidelines (Marsh et al., 2016, Thokala et al., 2016). In this guidance, the weight and scoring of different decision criteria are discussed as one of the MCDA steps, and the SW technique is mentioned as one approach to eliciting weights from stakeholders.

Tervonen and colleagues (Tervonen et al., 2015) have also provided guidance on how MCDA can be incorporated into a BRA; in addition to the methods and weight elicitation using SW process, there is also guidance on the inclusion of imprecise/incomplete data into a BRA. Phillips (Phillips, 2017) has also outlined an eight-step framework for constructing an MCDA model with best practice principles for use of MCDA in healthcare decisions, including the weighting of criteria such as using a SW technique.

Beyond the healthcare domain, the UK Department for Communities and Local Government has also released an MCDA manual for the appraisal of policy options that explains the SW process (Department for Communities and Local Government: London, 2009).

#### 4.3.2 SW-specific aspects of framework component 2 “design: method selection and analysis planning”

SW may be more appropriate for studies where a small sample size and trade-off data is desired. It is thought to provide the same precision of population preference estimates as a DCE with smaller sample sizes (Tervonen et al. 2019b). However, SW has higher responder burden (Tervonen et al. 2017) and risks eliciting lower quality choice data in studies where interviewer-led confirmatory checks are not carefully implemented (Keeney 2002). The inability to have interviewer-led checks is particularly important when large samples are needed, such as to obtain a representative sample of a large population or when a measure of preference heterogeneity is needed. Past studies have indicated preferences elicited with SW to be stable even when the elicitation process was replicated with new respondents in a different country (Nutt et al. 2010; van Amsterdam et al. 2015).

### 4.3.2.1 What type of research question is appropriate for SW?

Swing weighting elicits trade-offs among attributes in exact format. Therefore, it is appropriate for supporting benefit-risk evaluations that have been established as preference-sensitive. SW can be used to elicit preferences to use in a quantitative benefit-risk model.

### 4.3.2.2 When is SW an appropriate method?

Situations that suggest SW would be appropriate include:

- When it is useful for understanding individual patients' preferences. SW provides individual-level preference data that does not require any modelling, unlike DCE/BWS (Tervonen et al., 2017). For example, Postmus and colleagues used SW has been used to explore the distribution of individual preferences for multiple myeloma treatments (Postmus et al., 2018), to conduct a patient-centered BRA of aortic stenosis treatments (Marsh et al., 2019), and to establish the minimum acceptable benefit of treatment in exchange for the treatment risks (SriBhashyam et al., 2019).
- When it is only feasible to obtain small to medium sample sizes (5 to 50 respondents), such as in studies of rare diseases (SriBhashyam et al., 2019, Tervonen et al., 2017). SW may also be suitable for studies with complex attributes that would benefit from the presence an interviewer, such as attributes that are difficult to understand, that respondents are unfamiliar with, or that have implications that may be difficult to understand (Tervonen et al., 2017).
- For small or pilot studies with short timelines that do not allow for a separate instrument pre-testing phase. When applied in a workshop setting with experienced facilitators, SW enables the construction of attributes and elicitation tasks, although it is best practice, however, is to pilot the elicitation task before fielding it with the target respondents. Where there is an incomplete or long list of attributes; workshop elicitation enables the development of the final attribute list and preference elicitation within a single workshop session.

In contrast to some of the other preference elicitation methods, limited econometric expertise is not an issue as the experimental design and analysis are less complex than with methods requiring preference modelling.

### 4.3.2.3 When is SW not an appropriate method?

There are situations where SW would not be appropriate, such as studies where trade-offs (i.e. preference data) are not required, or for studies with large sample sizes or large populations; a standard SW study is resource intensive, as it requires interviewer-led elicitation (Tervonen et al. 2017). SW in small samples also may not be representative of patient population preferences in conditions with a large, heterogenous patient populations.

Further, SW may not be appropriate where there is a lack of experienced interviewers or where it is not feasible to train interviewers for the elicitation. SW requires considerable expertise as the interviewers need to ask confirmatory questions to validate respondents' preference statements, understand the method and moderate the workshop (Marsh et al. 2017a).

SW appears to be more cognitively burdensome than DCE for respondents (PREFER 2020a). In some situations, it may not be possible for respondents to either participate in long SW workshops (e.g. patients who may be critically ill) or have difficulty processing the SW task or providing the relevant numerical responses (e.g. patients with cognitive difficulties) (Medical Device Innovation Consortium, 2015).

Standard SW typically assumes an additive value model, which is based on the assumption that the different attributes are preferentially independent, i.e. that improvements in one attribute does not affect preferences for change in other attributes ((European Medicines Agency (EMA), 2008, Keeney, 1994, Keeney and Raiffa, 1993). DCE, in contrast, naturally allows for preferential dependence through the use of interaction terms in the regression. Although SW can be applied with non-additive (multilinear and multiplicative) preference models (Keeney and Raiffa 1993), they are difficult to apply in practice.

#### 4.3.2.4 Points to consider regarding the analysis planning for SW studies

'Analysis of SW data is relatively simple and can be done with basic descriptive statistics (Medical Device Innovation Consortium, 2015). It does not require complex modelling, as preference parameters are directly elicited from respondents in the required format (Tervonen et al., 2017). Population preferences can be estimated from SW data using Dirichlet regression (Tervonen et al., 2019b) that also enables the evaluation of the impact of respondents' characteristics on their preferences.

The robustness of SW can be assessed using various techniques such as the stochastic multicriteria acceptability analysis (SMAA) which enables the quantification of uncertainty of a benefit-risk decision due to imprecise swing weight estimates (Lahdelma and Salminen, 2001, Tervonen and Figueira, 2008, Tervonen et al., 2019b, Tervonen et al., 2011). SMAA can also account for uncertainty in the clinical data. Software is available that specifically supports the elicitation and analysis of SW data including ADDIS ([www.drugis.org](http://www.drugis.org)) (van Valkenhoef et al., 2013) and HiView (<http://www.catalyzeconsulting.com/software/hiview3/>), although simple Excel sheets are typically sufficient for SW elicitation workshops.

#### 4.3.2.5 Expected timeline for conducting a SW study

The typical timeline for an SW study is highly dependent on the length of time for various factors, including:

- contracting between the relevant parties
- time for determining the attributes and their levels (e.g. selected by the study team and/or stakeholders in a workshop, informed by literature reviews and qualitative research, or simply presenting respondents with a longer list of attributes)
- piloting
- recruiting respondents, the target sample size
- implementing the SW workshops and the number of workshops (e.g. one-off or multiple with each respondent)

The total duration could vary from a few months for simple SW studies with no qualitative research or piloting, to more than 12 months for more extensive studies. Analysis time may also vary depending on whether analyses are conducted on a population or individual level. The variability of timelines from study to study should be considered when planning to incorporate SW into the overall clinical program.

#### **4.3.2.6 Points to consider for SW internal validity testing**

SW elicitation tasks are cognitively demanding and therefore SW is typically conducted with individual in a workshop or focus-group setting and facilitated by an interviewer. Some studies have also implemented swing weight online via a survey ([Rutten-van Molken et al. 2018](#), [PREFER 2020a](#), [PREFER 2020b](#)). It is important to explain to respondents how to respond to the SW tasks prior to the main preference elicitation tasks. Respondents should be given an opportunity to deliberate and change their responses if needed.

The consistency of the weights and scores elicited should be tested throughout the elicitation exercise by eliciting qualitative reasons for respondents' choices or preferences; this enables the interviewer to gauge whether the respondents' understanding of elicitation tasks is consistent with how their responses will be used ([Marsh et al. 2016](#)). Performing consistency checks should also be conducted, whereby the interviewer reports back their interpretation of the respondents' responses in a different format for confirmation ([Department for Communities and Local Government: London 2009](#), [Goetghebeur et al. 2012](#), [Phillips 2017](#)).

#### **4.3.2.7 Points to consider for sample size**

Sample size is not typically a concern in SW given the small sample needed. A larger sample size is required when establishing population preferences—although the sample size requirements in these cases are likely to be significantly lower than for similarly complex DCEs ([Tervonen et al. 2019b](#)). This may pose an implementation challenge given the labour-intensive nature of the preference elicitation ([Tervonen et al. 2017](#)). Online survey-based SW loses the interaction between facilitator and subjects and reduces the opportunity for respondents to address questions about the attributes and elicitation tasks.

#### **4.3.3 Points to consider when setting up the preference question design**

There are several points to consider when setting up the preference question design of a SW study such as the number of attributes, mode of elicitation, implementation of SW tasks and minimising bias. While SW can account for a greater number of attributes than some other preference elicitation methods, too many attributes may also be cognitively burdensome for respondents. The preference question design should also consider whether the study objective is to obtain individual-level preference data or population-level consensus data.

SW is typically conducted as interviewer-facilitated focus groups or workshops but they may also be conducted as individual interviews if a group-setting is not feasible. A workshop or focus-group setting facilitates knowledge sharing, allowing respondents to clarify the tasks being posed and also facilitates discussion between respondents ([Marsh et al. 2017a](#)). Some SW variants have used an online survey format, without interviewer facilitation. Results from the PREFER case study on glucose monitoring devices for diabetes suggest that online SW using the standard procedure without the



best practices of interviewer facilitation and confirmatory questions is likely to lead to low data quality. However, the potential for bias from the facilitator should also be considered, and this can be minimised by ensuring that the facilitator understands the objectives of the exercise and is thoroughly trained on the facilitation approach (Phillips, 2017).

SW tasks can be implemented in various ways, for example, the second phase of SW elicitation can be implemented as comparisons in rank order (first ranked compared to second ranked, second ranked to third, etc.), or by comparing all attributes to the most important one. Attributes should also be designed and selected so they are preferentially independent, i.e. the preference for one attribute should not depend on the preference for another attribute.

Because the attribute weights obtained from a SW exercise represent the importance of the attribute swing, the range of the swing also needs to be considered. Scale ranges that are too large or too small may not elicit meaningful preferences (Phillips, 2017). For example, when considering the cost of different treatments, if the difference between the most and least costly treatments is small, then cost may not be considered as important, whereas a larger cost difference may be considered more important; though too large a range may cause respondents to underweight the importance of a unit change in the attribute. Accounting for the range of the swing is critically important when comparing SW weights with those obtained from other methods.

If SW is combined with other techniques such as scoring to account for non-linearity in the value function, then the appropriate number of levels per continuous attribute should be considered. Past studies have shown piecewise three-piece linear functions (i.e. four levels) to be sufficient for many downstream comparative analyses (Durbach and Stewart, 2009, Stewart, 1995) (Phillips, Nutt et al., 2010, van Amsterdam et al., 2015).

## 4.4 Best-Worst Scaling: Case 1

### 4.4.1 Introduction

The origination of BWS is generally attributed to market researchers, namely Jordan Louviere, who originally referred to the method as 'maximum difference scaling' or 'maxdiff', and this nomenclature is still popular in some disciplines and preference software packages such as Sawtooth (Finn and Louviere 1992; Louviere, Flynn and Marley 2015). The method is supported by the psychological premise that individuals can identify extremes—the best and the worst—when presented with a series of options. Therefore, BWS may be more difficult for respondents than a traditional DCE where respondents are only required to select one option (Whitty and Oliveira Gonçalves 2018). However, BWS is arguably easier for respondents than conducting the rating or ranking of options required by conjoint analysis methods where the responder must interpret the levels for several attributes of two treatments simultaneously. Best-worst scaling case 1 is often seen as an alternative to rating scales with anchored end points such as Likert or related questions and eliminates scale-interpretation issues (i.e., is one person's 6/10 equivalent to another person's 7/10?). Although it can take longer for the researcher to develop BWS tasks and longer for respondents to complete a BWS question, BWS requires respondents to answer fewer questions and likely produces, in general, more robust and reliable data than simple ordinal response categories or scale responses (Burton et al., 2019).

In BWS case 1, a list of objects (e.g. attributes of a medical product) is created and each respondent is presented with a series of choice sets containing a subset of these objects. These

a list of objects (e.g. attributes of a medical product) is created and each respondent is presented with a series of choice sets containing a subset of these objects. These objects are also sometimes called items, criteria or attributes. Because there are no levels, BWS case 1 can typically include a larger number of attributes than a traditional DCE. As with DCEs and other preference methods, it is commonly assumed that respondents would pick the object that provides the most utility (the most preferred or best, under utility maximization) and the least utility (the least preferred or worst). Respondents are usually asked to select their most and least preferred from a subset, although variations may include best and worst, most and least important, or most agree with and least agree with. The framing of the choice question is determined by the research question, and different frames and phrasing will reveal different information about respondents' preferences. From the responses made over a series of questions, it is possible to determine the ranking of the objects on an underlying latent scale.

([Cheung et al., 2016](#)) Best-worst scaling rating scores can be used to understand the following:

- The most and least relevant treatment outcomes to patients ([Yu et al., 2015](#))
- The relative importance of benefits and risks of treatment ([Yuan et al., 2014](#))
- Priorities for research into new health technologies ([Gallego et al., 2012](#))
- Concerns for health and non-health consequences of risky activities ([Marti, 2012](#))
- The importance of different adverse events on physicians' treatment decisions ([Cozmuta et al., 2014](#))

Simple count analysis enables BWS to estimate individual utility estimates, which can then be aggregated to reveal average preferences from a select sample. Individual utility estimates cannot be estimated with discrete-choice models.

Best-worst scaling case 1, like BWS case 2, is a ranking method, and the approaches are related but distinctly different and can be used to answer different research questions (see Section 2).

#### 4.4.2 Why Best-Worst Scaling Case 1?

The complexity of a BWS case 1 is a continuum, meaning a study can be relatively simple and quick and easy to implement. Simple BWS case 1 studies can be analyzed using basic count models to give "real time" individualized results, which may be useful for integrating into shared decision-making tools ([Flynn et al. 2015](#)). Reviews have shown BWS case 1 has risen in popularity during recent years ([Cheung et al. 2016](#); [Mühlbacher et al. 2016](#)). These studies are also very transparent and it is easy for stakeholders from all backgrounds to follow the study steps from aims and methods to results and conclusions. The simplicity comes at the expense of being able to answer research questions on acceptable trade-offs. As such, BWS case 1 is frequently used as a complementary method. For example, [Mansfield et al. \(2019\)](#) used BWS case 1 alongside a DCE to understand the preferences of metastatic melanoma patients' for efficacy (progression-free survival), risk of side effects, and mode and frequency of treatment administration. The separate BWS experiment allowed more in-depth investigation into the preferences for the treatment administration attribute by investigating the ranking of nine levels, which is more than could reasonably be incorporated in the DCE.

Most applications tend to involve studies where the research question requests a ranking or relative importance of many objects or attributes (e.g. [Silverman et al. \(2013\)](#) conducted a BWS case 1 study

to understand the ranking of 39 treatment features related to osteoporosis medicines.) In these cases, BWS case 1 may be sufficient.

#### **4.4.3 Best-Worst Scaling Case 1-Specific aspects of framework component 2 “Design: method selection and analysis planning”**

##### **4.4.3.1 What type of research question is appropriate for a BWS Case 1?**

Best-worst scaling case 1 is typically used when a ranking is required to understand preferences for objects that could include attributes of a product, trial end points, benefits, and/or risks. Analysis of a BWS case 1 study can reveal the order of these objects that can be particularly useful when there are many objects to appraise.

An example of BWS case 1 being used to understand the ranking of many objects is provided by [Husni et al. \(2018\)](#). A BWS case 1 study was conducted to understand patients’ and physicians’ preferences for the many symptoms of psoriatic diseases. The research question was to understand the perceived bother of psoriatic disease manifestations and to compare the perceptions of patients and physicians. The study included 20 objects and BWS case 1 was used to rank these in order of importance (of relative bother). For patients, the most bothersome items were painful, inflamed or broken skin, very closely followed by joint pain, soreness or tenderness. The least bothersome item was difficulty choosing clothing. The physicians assessed joint pain, soreness, or tenderness as most bothersome, followed by discomfort while doing everyday tasks.

An advantage of using BWS case 1 as opposed to Likert questions or a visual analogue scale is that there is no need for ‘calibration’, and it encourages discrimination among important objects. In the example cited above, all outcomes may, understandably, have been rated by patients as ‘very bothersome’ or ‘5 out of 5’, revealing little about their relative importance or ordering to the analyst. Because BWS forces a complete ranking of all objects, the method reveals more information about the ordering of respondents’ preferences.

Another notable example of BWS case 1 is provided in an article by [Hauber et al. \(2017\)](#). The research question in this study considered the relative importance of various treatment risks. The BWS exercise was conducted alongside a DCE (presented in one survey) to understand the relative importance patients with anemia placed on avoiding seven potential problems of a blood transfusion. Respondents received seven choice sets that asked them to select the most and least bothersome from a subset of three transfusion attributes. Analysis of the data revealed that patients were most bothered by having lung damage and getting a serious infection because of a transfusion, and they were relatively least bothered by needing to arrange transport to a hospital or center to receive a transfusion.

##### **4.4.3.2 When is BWS Case 1 an Appropriate Method?**

Situations that suggest BWS case 1 would be appropriate include:

- When there is a limited sample size
- When individual utility estimates are important (e.g. tailoring a patient-physician discussion or shared decision-making tool in a clinical setting)

- Where there is a need to reduce information (e.g. eliminating less important benefits or risks from a long list of trial end points or reducing attributes in another preference study)
- For collecting data supplementary to a preference study (e.g., to understand the ordering of levels or attributes that could not easily be incorporated into a DCE; or to quantify attitudes, perspectives, and perceptions that may be used to explain preference heterogeneity)
- When simple analytical models are required (e.g. when the research team has limited experience, or the decision maker requires simplistic analysis for transparency)
- When there is a need to understand the relative importance of objects without associated levels

#### 4.4.3.3 When is BWS Case 1 not an appropriate method?

“Best” is not a synonym for “acceptable,” and “worst” does not mean “unacceptable.” Therefore, BWS case 1 (and case 2, to a lesser extent) is limited in its ability to look at thresholds such as MAR, willingness to pay, or demand (preference shares), which require discrete choices.

There are other situations where BWS case 1 would be inappropriate or insufficient.

- when the research question is multifaceted and thus requires a more complex method (e.g. there is an interest in understanding relative importance in addition to trade-offs, thresholds and/or demand)
- when there are multiple product features of interest and it is important to understand preferences for the attribute levels
- when seeking to understand the trade-offs individuals are willing to make between objects
- when the ranking of many of the objects is self-evident to most responders, in which case BWS case 1 will give ranking of the objects but its utility estimates will be misleading

#### 4.4.3.4 Points to consider regarding the analysis planning for BWS case 1 studies

The subset of objects presented in the choice tasks are created using an experimental design. As with DCEs, an experimental design allows the researcher to reduce the criteria to a reasonable amount for a respondent to consider when making a choice. Most BWS case 1 studies use balanced incomplete block designs (BIBD) ([Cheung et al. 2016](#)) where the ‘blocks’ are the subsets of objects presented in the choice set. Note that in BWS designs, a BIBD block refers to the set of objects in each choice question and not to the subset of questions from a split design (as is common in the DCE literature). When the same number of objects appear in each choice task, and each object appears the same number of times and an equal number of times with every other object, the design is ‘balanced’.

Approaches for analyzing BWS data vary in complexity, and the simplest analytical approach can include direct counts of the number of times an item was selected best or worst. As with responses to DCEs, respondents to BWS studies provide answers over multiple choice sets, which creates cross-sectional data. Analyses may therefore closely match the analysis of DCE data where the worst data are appended and provided a ‘-1’, and are then analyzed ‘sequentially’ (respondents chose the best

from a set then the worst from the remaining objects) using discrete-choice models. Alternatively, maximum-difference models assume an individual makes a choice by simultaneously selecting the option with the biggest difference in utility (i.e. identifying the best and worst objects simultaneously); these are rarely implemented in practice and will not be described in detail in this section.

There is limited guidance for BWS case 1 best practices from task forces, policy makers or other established bodies. Furthermore, guidance for conducting DCEs or other preference-based methods may not be applicable. However, some useful resources include the following:

- The textbook *Best-Worst Scaling: Theory, Methods and Applications* ([Louviere, Flynn and Marley, 2015](#))
- For experimental design, popular software programs can convert BIBD into BWS case 1 questions (e.g. see *BWS R package* [[Aizaki, 2015](#)])
- The choice of modeling approach depends not only on the research question and data collected but also on underpinning psychological theory about how individuals are believed to have made choices (i.e. sequentially or simultaneously). For sequential choice making, [Cheung et al. \(2019\)](#) provides a useful illustration of five analytical methods including count analysis, multinomial logit, random parameter logit and latent class models ([McFadden and Train, 2000](#)), as well as hierarchical Bayes estimation.

#### 4.4.3.5 Expected timeline for conducting a BWS case 1 study

As with other stated preference methods, the time to conduct the study strongly depends on the recruitment of respondents to the survey, the extent of piloting or prior qualitative research, and the arrangements between interested parties. There is a body of work suggesting BWS case 1 could be less time-consuming than a DCE study because of opportunities to simplify procedures for both the experimental design and analysis, but the study's steps are somewhat more intensive than simple Likert-style questions collecting the strength of preferences. As a rough estimate, most sponsor-conducted BWS case 1 studies done in collaboration with an academic group or a consulting company take approximately one year.

#### 4.4.3.6 Points to consider for BWS case 1 internal validity testing

A BWS case 1 study can use internal tests for validity as used in other quantitative preference methods, such as DCEs, and outlined in the framework. For example, a BWS case 1 study can incorporate tests for the following:

- **Stability:** With a repeated choice set, testing if the same objects are consistently chosen as best and/or worst.
- **Transitivity:** Testing whether, if object X is worse than object Y, and object Y is worse than object Z, then object X is worse than object Z in a subsequent choice set (holding all other attribute levels constant).
- **Straight-lining or flatlining:** Testing if respondents make a selection based on object location rather than information.

As highlighted in the framework and DCE sections, failing these tests is not always a definitive indication that a survey respondent was irrational or inconsistent, nor does it indicate the results of a study are invalid. For example, if two objects had similar utilities (i.e. the respondent was indifferent between two objects being best or two objects being worse), then the respondent's choice may effectively be random, which could result in different answers to the same question and "failure" of a stability test.

In a BWS case 1 study, it may also be useful to investigate the face validity of the survey using qualitative research methods (e.g. interviews) to ascertain if respondents are answering in line with theory or a priori expectations. Face validity may also be explored quantitatively in post hoc analysis. For example, [Yuan et al \(2014\)](#) tested whether respondents had a good understanding of the clinical outcomes (items), paid close attention to the survey, and took the exercise seriously by identifying those who chose an outcome other than disabling stroke or moderately disabling stroke as worse than death in any single question.

#### **4.4.4 Best-Worst Scaling Case 1- specific aspects of framework component 2 'design: sample definitions—justifying the sample size'**

##### **4.4.4.1 Points to consider regarding the sample size of a BWS case 1 study**

Reviews of BWS studies suggest, on average, case 1 studies have approximately 260 respondents, with some studies completed with samples of fewer than 100 (see [Imaeda, Bender and Fraenkel, 2010](#); [Torbica et al., 2014](#); [Meyfroidt et al., 2015](#); [Ross et al., 2015](#)). In BWS there are more observations as the choice task is expanded, thus the method can often be conducted with smaller sample sizes than a DCE study. As with any regression model, as the number of parameters increases, so does the required number of observations. Therefore, studies with many objects require a larger sample size, not only because of the number of parameters, but also because these studies may have included designs to reduce the complexity of the task for respondents (i.e. fewer objects in a choice set, fewer choice sets or using subsets of the full design by blocking). Similarly, more complex models typically require more observations. For example, a random-parameters logit model with all attributes/levels included as random parameter estimates both a mean and a standard deviation, doubling the number of parameters compared with a simple multinomial logit model.

If a formal sample size calculation is required, rules of thumb and calculations from the DCE literature can be used, see [Marshall et al., \(2010\)](#), [de Bekker-Grob et al., \(2015\)](#), and [Yang et al., \(2015\)](#), where each item is considered an attribute, and the number of levels of each attribute is set equal to two - present or absent. Alternatively, general rules of thumb for simple regression analyses suggest 20 to 25 observations per parameter (i.e. per attribute), which may be sufficient ([Adelman et al., 2020](#)).

#### **4.4.5 Best-Worst Scaling Case 1-Specific Aspects of Framework Component 2 “Design: Preference Question Design—Consideration of the Appropriate Number of Attributes and Attribute Levels; Patient Burden Issues”**

##### **4.4.5.1 Points to consider regarding the choice of objects**

As with other preference methods, the choice of attributes should primarily be determined by the research question. In BWS case 1, it is important to have a comprehensive list of objects that reflect all features likely to be important to the individual's decision. Firstly, if the most important object is absent, the researcher may erroneously conclude the remaining objects are important when they are all relatively trivial. Secondly, it is impossible to infer the ranking of a new object post-data collection. For a DCE, although missing attributes cannot be added post hoc, the value of a missing numerical level may be inferred from existing levels (making assumptions on the functional form of utility). Likewise, for Likert or simple scale responses, an additional independent survey question could be developed without the need to run the whole survey again.

However, researchers selecting objects should be cognizant of the number of objects given their experimental design. Balanced incomplete block designs do not exist for all objects. It could be advantageous to remove or combine items when possible ([Louviere, Flynn and Marley, 2015](#)).

##### **4.4.5.2 Points to consider in the choice context**

A key consideration in BWS case 1 is the choice context presented to individuals. Despite the method's name, the labels 'best' and 'worst' are not mandatory descriptors, rather descriptors should fall on opposite sides of a scale. For some questions, it may be more reasonable to ask about the 'most preferred' and 'least preferred' or the 'most important' and 'least important' objects. However, there is a balance between framing the choice context to best answer the research question, and respondents' interpretation and understanding. For some situations (e.g. risky endpoints in a trial), respondents in a BWS case 1 may feel all objects are the least preferred or no single object is least important. For these situations, careful explanation of the choice context may assist respondents in making a choice. However, in some instances a forced choice may be unrealistic, and BWS case 1 is not an appropriate method - researchers should choose an alternative method to allow for indifference (e.g. TT) or the opportunity to opt-out (e.g. DCE).

##### **4.4.5.3 Points to consider in the set-up of choice sets (experimental design)**

The experimental design refers to the specific combinations of objects presented in the choice tasks that responders to a BWS case 1 study are asked to complete. Most BWS case 1 studies use BIBD, where 'blocks' are the subsets of objects presented in the choice set ([Cheung et al., 2016](#)). For some numbers of objects, there is no BIBD. Like BIBD, Youden designs ensure every object occurs in every block (set) an equal number of times. They also ensure each object occurs in each position (row) an equal number of times ([Hess and Daly, 2014](#)). For certain numbers of objects, a BIBD may exist. Alternatively, orthogonal main effects plans as used in DCEs, can be used to create blocks of objects to present in the choice tasks ([Flynn et al., 2015](#))([Cheung et al., 2016](#); [Mühlbacher et al., 2016](#)).

## 4.5 Best-Worst Scaling: Case 2

### 4.5.1 Introduction

BWS case 2 has similarities to both BWS case 1 and DCE/BWS case 3. Like a DCE, alternatives are described in terms of both attributes and levels, and like BWS case 1, respondents are asked to select some variant of the best or worst attribute level (most or least important, etc.). Unlike a DCE, individuals are presented with a single profile from which they must state which attribute level is best or worst (Louviere, Flynn and Marley, 2015). The profiles are determined by an experimental design. As with BWS case 1, BIBD designs are possible but most studies typically use designs similar to a DCE (e.g. fractional factorial designs) (Cheung et al., 2016). A review of BWS studies found two-thirds of case 2 examples used orthogonal main effects plans. This is because introducing attribute levels means there are many more possible combinations than in BWS case 1; therefore, BIBD can quickly become too complex.

The analysis of BWS case 2 also has parallels to both BWS case 1 and DCE. A simple count analysis can be performed, where the number of times an attribute level is selected as worst is subtracted from the number of times an attribute level was selected as best, with an adjustment for the number of times the attribute level appeared. This simple count analysis provides individual level utility estimates and can be aggregated across respondents to capture the preferences of the sample. This approach is popular in BWS case 2, and almost a quarter of studies presented result in this way (Cheung et al., 2016). As with BWS case 1, data can also be analyzed using maxdiff (simultaneous) or discrete-choice models (sequential), depending on the analyst's underlying psychological assumptions regarding respondents' choice formulation. In BWS case 2, weighted least squares is also used frequently (reported in 15% of studies). The analytical approaches and related considerations are described in more detail in section 4.5.3.

One of the first applications of BWS case 2 in health was a study eliciting preferences for health states (Szeinbach et al., 1999), although it was described as a 'maximum difference conjoint analysis'. Patients were presented with six attributes (e.g. domains of health-related quality of life) described by one of three levels (e.g. no problems, some problems, moderate problems) and were asked to indicate which attribute level would be the hardest ('worst') and easiest ('best') to live with. Since this application, there have been many more examples in a range of areas, but the method remains particularly popular for valuing health states and outcomes (see the ICECAP and the Child Health Utility [CHU-9D, EQ-5D Youth version (EQ-5D-Y)] literatures [Coast et al., 2008; Ratcliffe et al., 2012; Dalziel et al., 2020]).

The evidence regarding whether case 2 BWS and DCE yield comparable results is mixed. A recent review found that there is agreement between the results of BWS case 2 and DCE. Specifically, among nine empirical studies comparing DCE with BWS case 2, the results of the two methods were concordant (Whitty and Oliveira Gonçalves, 2018). Some studies suggest that BWS case 2 and DCE (Broekhuizen 2015a and 2015b) are not significantly different (Potoglou et al., 2011), whereas others found they yielded different preference estimates with poor performance of mid-ranked attributes/levels (Severin et al., 2013).



## 4.5.2 Why BWS case 2?

The decision to choose BWS case 2 is likely to be driven primarily by the research question and a consideration of the sample (i.e. the cognitive burden to respondents). However, reviews of the literature suggest BWS case 2 is the most popular method, with slightly more applications than case 1 and many more applications than case 3 in health ([Cheung et al. 2016](#); [Mühlbacher et al. 2016](#)).

Best-worst scaling case 2 offers more insights than the ranking of objects in BWS case 1 but can still be analyzed in a way that is simpler than a DCE (e.g. count analysis). For this reason, BWS case 2 can be particularly useful when the researcher (or the audience) has mixed experience with preference research but seeks to understand preferences for multiple attributes and levels. A well-designed BWS case 2 study with certain properties (e.g. level balance) can be analyzed with models varying in complexity. This means BWS case 2 may be useful where the research question is fluid because the final models may be either simple (and transparent) or more complex.

## 4.5.3 Best-Worst Scaling Case 2- specific aspects of framework component 2 'design: method selection and analysis planning'

### 4.5.3.1 What type of research question is appropriate for a BWS case 2 study?

As noted in [Section 4.5.2](#), there have been many applications of BWS case 2 for health states or health outcomes valuations, and BWS case 2 has been used to understand and compare preferences for benefits and risks too. For example, [Knox et al. \(2002\)](#) elicited preferences for the attributes of contraceptive products in a study with women and general practitioners. The authors found heterogeneity, with women rating heavy periods with increased pain and cost (\$A60) as the 'worst', while the worst attribute level for general practitioners was the level of contraceptive effectiveness (10/100 annual pregnancy rate).

### 4.5.3.2 When is BWS case 2 an appropriate method?

Situations that Factors that suggest BWS case 2 would be appropriate include:

- When information is needed about the relative importance of levels, as well as attributes, thus offering more information than BWS case 1.
- When a reduced cognitive burden is important. Because respondents see a single profile for some samples (rather than the multiple profiles they would see in a DCE), this may be less cognitively burdensome, particularly if there are many attributes of interest. The reduced cognitive burden is cited as a reason for the method's popularity in health-state valuation studies with older people and children ([Coast et al., 2008](#)).
- When an alternative utility frame is proposed, [Coast et al. \(2008\)](#) uses BWS case 2 to understand values of an instrument to measure capabilities. Because BWS case 2 does not require trade-offs, the authors suggest it aligns better with Sen's Capabilities Approach, which focuses on capability rather than functioning and does not require trade-offs between attributes or levels.
- When real-time individualized preference data is sought (e.g. for shared decision-making), count analysis can provide an immediate ranking of attribute levels.

### 4.5.3.3 When is BWS Case 2 not an appropriate method?

When the following exists or occurs, BWS case 2 would be inappropriate or insufficient.

- The research question is multifaceted and lends itself to a more complex method. For example, when the researcher is interested in the relative importance of attributes and attribute levels, in addition to demand, MAR, or willingness to pay. Designing a BWS case 2 study to create data for these end points may be challenging and require other survey questions in addition to the choice sets.
- There is an interest in estimating utility in terms of MAR, minimum acceptable benefit (MAB), willingness to pay, or some other common value.
- There is a need to assess a profile's performance relative to another (e.g. old treatment or no treatment). To estimate unconditional demand, the study would need a follow-up question (i.e. would you choose in real life? Yes or no.). In these instances, including a follow-up question raises two issues: (1) how to model these responses with the best-worst data and (2) why not use a method more adept to modeling unconditional demand.
- When the ranking of many of the objects is self-evident to most responders, in which case BWS case 2 will give ranking of the objects but its utility estimates will be misleading

### 4.5.3.4 Points to consider regarding the analysis planning for BWS case 2 studies

- If the design was not balanced, BWS case 2 count analysis must take account of the number of times the attribute level occurred in the profile sets (i.e. the number of opportunities there were to select it as best/worst).
- In count analysis, it may also be useful to look at the sum of the squared difference between the best and worst scores across the attribute levels for an individual because this indicates the variance in responses (i.e. the consistency of the respondent's preferences) (Louviere, Flynn and Marley, 2015).
- As with BWS case 1, the choice of modeling approach for case 2 also depends on the underpinning psychological theory about how individuals are believed to have made choices, either sequentially (marginal model) or simultaneously (so called maxdiff).
- As described in the BWS case 1 section, models of sequential choice are typically easier to estimate, are more common, and are available in common statistical software packages. The sequential analysis of BWS case 2 data is similar to BWS case 1 in that a "1" indicates the best and a "-1" indicates the worst attribute level. Therefore, much of the guidance from the International Society for Pharmacoeconomics and Outcomes Research Task Force on statistical analysis of DCEs is applicable to BWS case 2 when a sequential choice is assumed.
- Weighted least squares is a less popular estimation approach, but a review of BWS case 2 studies in 2016 found it was reported in ~15% of articles (Cheung et al., 2016). Weighted least squares regression is an alternative to estimating a conditional logit model with maximum likelihood estimation (Louviere, Flynn and Marley, 2015). In a BWS case 2 study estimating preferences for dermatology consultations, data were analyzed using conditional logit models with maximum likelihood estimation and weighted least squares estimation with

the results, suggesting a high level of agreement between the approaches (Flynn et al., 2008). It has been suggested that weighted least squares may be a useful first step for identifying outliers (i.e. attribute levels at the extremes) (Louviere, Flynn and Marley, 2015).

#### 4.5.3.5 Expected timeline for conducting a BWS case 2 study

The timeline for a BWS case 2 study is likely to be comparable to a DCE, as many of the study steps are similar. For example, BWS case 2 also requires the identification of attributes, attribute levels, and creation of an experimental design. Similarly, the analytical models may be as complex as a DCE, and investigations into preference heterogeneity could be as extensive. As with all preference methods, the study schedule will depend on how long it will take to recruit participants, how refined the research question is (i.e. the need for qualitative investigations), and contracting between the relevant parties.

#### 4.5.3.6 Points to consider for BWS case 2 internal validity testing

As noted in the framework, there are various methods that can be used to understand the internal validity of a stated preference studies. For BWS case 2, some methods are not appropriate (e.g., there is no opportunity to look at dominated pairs specifically). However, techniques that may be incorporated into a study design include the following:

- **Stability:** With a repeated choice set, testing if the same objects are consistently chosen as best and/or worst.
- **Transitivity:** Testing whether, if object X is worse than object Y, and object Y is worse than object Z, then object X is worse than object Z in a subsequent choice set (holding all other attribute levels constant)
- **Trading behavior:** Testing for dominant preferences by measuring the times an attribute is selected as best or worst. For example Ryan, Krucien and Hermens (2018) conducted a BWS case 2 study for health state valuation and created a score to identify lexicographic behavior, which was measured from 0 (never selects an attribute as best or worst) to 100 (always selects an attribute as best or worst), where dominance was defined as a score of 50 or higher.
- **Face validity:** exploring respondents' reactions and understanding of the experiment by using qualitative interviews or selective free-text comments within the survey. Whitty and Oliveira Gonçalves (2018) also found that a number of BWS case 2 studies used self-reported measures of difficulty to assess understanding.

### 4.5.4 BWS case 2-specific aspects of framework component 2 'design: sample definitions— justifying the sample size'

#### 4.5.4.1 Points to consider regarding the sample size of a BWS case 2 study

Best-Worst Scaling case 2 studies have been conducted with relatively small samples (fewer than 50 respondents [Jones, Hawkins and Brown, 2015]) and with much larger samples (more than 1,000 respondents in some health-state valuation studies [Ratcliffe et al., 2016]). Because no sample size calculations exist specifically for case 2, researchers can use the rules of thumb from the DCE

literature ([de Bekker-Grob et al., 2015](#)). Researchers should be aware that increasing the number of attributes and levels will typically increase the number of respondents needed because the models estimated will require more observations to achieve a certain level of significance for each parameter. Usually though, for estimation purposes, for each respondent, the total number of attribute levels plus one answered question is necessary.

#### **4.5.5 BWS case 2-specific aspects of framework component 2 'design: instrument design—consideration of the appropriate number of attributes and attribute levels; patient burden issues"**

##### **4.5.5.1 Points to consider regarding the choice of attributes**

As with other stated preference methods, the attributes describing the profiles are determined by the research question. Similar to DCEs, researchers must balance the acquisition of more information with cognitive burden and carefully consider what is needed to be known and how much a respondent can reasonably consider at once. Because a respondent considers only one profile at a time, a BWS case 2 study can arguably accommodate more attributes than a traditional DCE where two or more alternatives are typically considered.

In answering the research question, the attributes of a BWS case 2 study will consider (1) the features of treatment that are important to the decision-maker (e.g. a regulator, a HTA body or a physician), and (2) features important to the respondent's choice. These two perspectives do not need to be contradictory, and including an attribute important to the decision-maker (e.g. a particular treatment feature) and finding it is of relatively low importance to patients may be a study finding in itself.

The identification of appropriate attributes has parallels to the DCE literature, and as such researchers may seek to consult with experts or conduct qualitative research to identify or reduce attributes ([Hollin et al., 2019](#); [Shields et al., 2020](#)).

##### **4.5.5.2 Points to consider regarding the choice of levels**

The levels in a BWS case 2 study should reflect the possibilities for the attributes of the alternative. Choosing unrealistic levels may encourage abnormal trading behavior (i.e. always picking an attribute as best/worst). As with DCEs, levels can be numeric, ordered-categorical, or categorical (and not naturally ordered). When including non-numerical attribute levels, there is no opportunity to investigate functional form or include these as continuous parameters in the analysis of preference. Therefore, the choice and quantity of levels in a BWS case 2 study can potentially have a greater impact on the complexity of the experimental design and the required sample size.

##### **4.5.5.3 Points to consider in the set-up of choice sets (experimental design)**

Similar to BWS case 1, a key step in setting up the choice sets for a BWS case 2 study is determining the question frame. 'Best' and 'worst' can be framed as 'most bothersome' and 'least bothersome' or the 'most important' or 'least important' levels, but these then result in different interpretations. Another variation is 'best' and 'second best' but this frame (best-best scaling) requires a different analysis (i.e. rank-ordered models) in addition to a different interpretation. Because BWS case 2 is limited in its ability to estimate demand, the choice sets may also include a follow-up question (e.g. Would you choose this profile? Yes/no).

In terms of the experimental design, BWS case 2 design can follow the BIBD designs in BWS case 1. BIBD designs may be difficult as the number of attributes and or levels increases. In these instances, fractional factorial designs, such as those used in DCEs, may be used. For a description of experimental designs, see [section 4.5.3](#). Indeed, most BWS case 2 studies used orthogonal main effects plans (OMEPS) ([Cheung et al., 2016](#); [Mühlbacher et al., 2016](#)).

## 4.6 Threshold Technique

### 4.6.1 Introduction

Threshold-based approaches to preference elicitation are a type of indifference method that aims to find different combinations of attributes (e.g. benefits and risks) offering the same level of utility ([Soekhai et al. 2019](#)). Typically, these methods vary the value of one attribute in an option until the participant is 'indifferent' to the alternatives. In addition to TT, threshold approaches include the standard gamble, time tradeoff, and contingent valuation ([Soekhai et al. 2019](#)). Arguably, threshold-based approaches have been the predominant preference elicitation method for most HTA decisions where weights for health states have conventionally been derived from time trade-off valuation studies.

In a TT study, a respondent – typically a patient or physician – is presented with a choice between two healthcare options. One of the options is the 'reference option', which is the baseline against which an alternative is compared with. This reference option is usually the current standard of care (e.g. current treatment or no treatment). The second is the 'target option' and usually presents both an increase in the benefit and an increase in burden relative to the reference option, therefore, requiring respondents to make a tradeoff when choosing their preferred alternative.

The alternative in a TT, as in other preference methods, is defined by its attributes and the levels of these attributes. The attributes are typically the benefits and risks associated with a new treatment, although studies have also considered preferences for other attributes (e.g., waiting time, number of clinic visits) ([Hauber and Coulter, 2020](#)).

Studies considering thresholds for probabilities (i.e., the chance of a benefit or the risk of a harm) are referred to as 'probabilistic threshold technique'. When the key attribute of interest is a measure of burden (e.g. the risk of harm, time, or cost), the estimated threshold is the level of the additional burden that exactly offsets the incremental benefit provided in the target option. Conversely, if the key attribute is a benefit (e.g. chance of a benefit, improvement in quality of life, survival time), the estimated threshold reveals the minimum additional benefit that the target must provide to offset the incremental burden of that option.

In the initial question of a TT series, the key attribute is typically set to have the same level in both the reference and target options. If the reference option is chosen first, the key attribute of the target is made better and the question is repeated. If the target is chosen first, the key attribute of the target is made worse and the question is repeated. The procedure is repeated until the researcher can identify the threshold level at which the respondent is indifferent between the reference and target options. Because each TT series is focused on eliciting the threshold for a particular attribute, the TT exercise should be repeated to understand thresholds for other attributes or target treatments. Therefore, to some extent, TT can estimate preferences for multiple attributes.

The threshold can be a specific value or an interval (range) within which a respondent's threshold for a particular attribute lies. For responses where the range is known but the exact value is unknown, interval regression is used to model the categories. These models take account of the upper and lower limits of the range in which the threshold value is known to lie. Interval-regression models reveal the mean threshold value of the key attribute for a given level of another attribute holding all else constant.

These results of the analysis can thus inform the following:

- the maximum risk a patient would be willing to accept for a certain level of benefit
- minimum reduction in one risk that would make another risk worthwhile

Sample size permitting, explained preference heterogeneity can be investigated by estimating separate sets of thresholds for different subgroups of interest. Unlike fixed-effects, discrete-choice models, interval-regression models can easily include many covariates to explore whether and how respondent characteristics influenced the mean threshold for each attribute. Unexplained preference heterogeneity can be investigated by examining the distribution of threshold values, which in turn can be considered alongside individual characteristics in subgroups or covariate analyses.

#### **4.6.2 Why TT?**

The TT is becoming an increasingly popular approach for quantifying preferences, and a review of the method found 43 examples published between 1991 and 2016 ([Hauber and Coulter, 2020](#)). Probabilistic TT is particularly suited to estimating individuals' MAR and MAB., and the method is therefore promising for decisions regarding benefit-risk tradeoffs. The Center for Devices and Radiological Health at the FDA has used TT to quantify preferences for health technologies. For example, data from a TT study quantifying preferences for ear tube placement were used to support regulatory approval of the Tula ear system ([FDA, 2019](#)). The data demonstrated that an in-office procedure with a success rate of at least 68% (i.e. the 'threshold') was preferred to placement under general anesthesia with a success of 99%.

Most applications tend to involve estimating thresholds for benefit-risk tradeoffs, particularly to understand MAR and MAB ([Hauber and Coulter 2020](#)). Although studies have incorporated multiple thresholds ([Tomlinson et al. 2011](#)), a review showed that most have focused on one or two key attributes of interest ([Hauber and Coulter 2020](#)). The relative simplicity of a TT makes the design and analysis relatively accessible for researchers or decision makers from different backgrounds. Experimental design, regression models and results tend to be easier to interpret than those in more complex preference methods.

#### **4.6.3 TT-specific aspects of framework component 2 'design: method selection and analysis planning'**

##### **4.6.3.1 What type of research question is appropriate for a TT study?**

Typically, TT is used when there are a few attributes or tradeoffs of interest. In situations where there are many attributes, a DCE (if fewer than approximately 8 attributes) or other preference method (e.g. BWS case 1, if there are many) may be more suitable. As with other stated preference methods, these are particularly useful when treatments are new or when clinical judgement alone is

insufficient for decision makers. Threshold questions may also be useful when a large amount of heterogeneity is anticipated. For example, where the levels of a DCE study maybe too broad. Although TT cannot incorporate as many attributes as BWS case 1, responses indicate demand rather than the relative order of preferences alone.

A good example of a multiattribute TT is described in a paper by [Devereaux et al. \(2001\)](#). The authors conducted a probabilistic TT study with patients and physicians to understand the threshold for the minimum reduction in the risk of stroke and the maximum increase in risk of a bleed acceptable for people with atrial fibrillation considering treatment with antithrombotic drugs. The authors identified heterogeneity between patients and physicians, notably in the thresholds for risk of bleed, where patients were more tolerant.

Another good example is provided in the IMI PREFER case study by [van Overbeeke et al. \(2020\)](#). The study investigated the tradeoffs patients with hemophilia were willing to make when choosing between prophylactic factor replacement therapy (PFRT) and a new gene therapy. The survey design incorporated three series of TT questions to understand the MAB to switch to gene therapy in terms of annual bleeding rate, change to stop prophylaxis, and quality of life. The PFRT and gene therapy were also described by their evidence base (the study follow-up duration), as PFRT is an established treatment (30 years of side-effect monitoring), while gene therapy is relatively new (only 10 years). In addition to MAB thresholds, the study also estimated the proportion of patients who would accept gene therapy under certain scenarios. Further analyses revealed significant preference heterogeneity.

#### 4.6.3.2 When is TT an appropriate method?

The following factors may indicate that a TT would be appropriate:

- When pairwise tradeoffs are all that is needed.
- When the population of interest is anticipated to be small, for example, because the condition is rare or because of resource (budget and time) constraints. In these instances, estimating preferences for a key attribute in a TT study may be more appropriate than a more complex method with an insufficient number of observations to estimate models with any statistical confidence.
- When the sample's cognitive function is unknown, highly variable, or potentially limited (e.g. patients with brain disorders or very young children). There is an emerging evidence base suggesting TT may be more suitable in samples unable to make the complex tradeoffs required by DCEs or the ranking needed in a BWS study ([Peay et al., 2019](#); [Poulos et al., 2020](#)). Because the method can be conducted with smaller sample sizes, there are more opportunities for interviewer-assisted data collection where respondents can ask for clarification.
- When there is a need or desire to understand the preferences of an individual. Because TT is a direct elicitation method ([Hauber, Fairchild and Johnson, 2013](#)), each respondent reveals their threshold. In DCE studies, thresholds are derived from the choices made over multiple choice sets.
- When there is an interest in predicting patient choice and/or forecasting demand. When there is a new (target) treatment to compare with an existing (reference) case, a TT can directly

incorporate the scenarios within the experiment (e.g. the initial question could be based on true values). It is also possible to identify individual characteristics associated with picking one treatment over another. Unlike with indirect preference methods, there is no need for an additional “direct elicitation” question or post-analysis simulation.

#### **4.6.3.3 When is TT not an appropriate method?**

The following are examples of situations where TT would be inappropriate or insufficient:

- When the research question lends itself to a more complex method because there are many key attributes or tradeoffs of interest that would result in too many TT exercises for any individual to complete.
- When interactions between the attributes are important. A TT study cannot easily accommodate interactions – that is when the threshold of an attribute depends on the level of more than one other attribute. If an aggregated measure of net benefit is required, a method that can simultaneously estimate preferences for multiple attributes and interactions may be more appropriate. DCEs and related methods are most adept at incorporating designs to estimate between attribute interactions (Hauber and Coulter, 2020)(Tomlinson et al., 2011)(Hauber and Coulter, 2020).

#### **4.6.3.4 Points to consider regarding the analysis planning for TT studies**

As of 2020, there is no specific guidance for the analyses of TT data from reputable bodies such as the ISPOR Task Forces. As with many quantitative studies, the exact analytical approach will depend on the study research question, the study design and constraints from the sample size.

In the interval regression, there are two dependent variables that represent the lower and upper bounds of the interval. For respondents at the extreme lower end (e.g. not willing to accept any risk or willing to wait any time) both the upper and lower bounds will be 0. For respondents at the extreme upper end, the lower bound of the interval will be equal to the maximum level presented in the survey and the upper bound to the feasible maximums (e.g. 100% for risk). If respondents directly state their threshold, both the upper and lower bounds of the interval will be equal to their statement. For respondents in-between (e.g. those who accept increases in risk in some cases but not in others), the lower bound value is equal to the lowest minimum value the respondent accepted, while the upper bound value is equal to the lowest minimum value the respondent rejected.

Multi-collinearity is another point of consideration in the analysis when investigating preference heterogeneity by incorporating covariates into the regression model. Researchers should avoid including covariates that are collinear by assessing correlation between characteristics of interest. To do this, a correlation matrix could be produced with covariates selected for inclusion in the regression model only if the correlation is sufficiently low.

#### **4.6.3.5 Expected timeline for conducting a TT study**

As with all preference methods, the expected timeline for conducting a TT study depends on practical and logistical constraints, including contracting between parties, recruitment of the sample population, the need for ethical approval and the extent of engagement with decision-makers. Generally, the timeline of a TT study is shorter than a DCE study, primarily because there tends to



be a simplified approach to the selection of attributes and levels, there is no experimental design and simpler regression models are typically required in the analysis.

#### 4.6.3.6 Points to Consider for Threshold Technique Internal Validity Testing

The PREFER framework notes methods for understanding the validity of a preference study, and many of these could be informative in a TT study (including stability, straight-lining or flat-lining). For TT specifically, further tests could be used to explore the following:

- **Monotonicity:** A simple test for validity used to confirm that higher levels of benefits are preferred to lower levels of benefit, and that lower levels of burden are preferred to higher levels of burdensome attributes over different TT series.
- **Anchoring effects:** In a TT study, the starting levels of the key attribute are specified by the researcher. This means respondents have the potential to ‘anchor’ to this starting point when answering subsequent questions. It has been suggested that methods like TT may be more susceptible to starting point biases than DCEs, but the evidence is varied (Robinson, Spencer and Moffatt, 2015; Rodríguez-Míguez, Pinto-Prades and Mosquera-Nogueira, 2019). Tests for anchoring could be explored by varying the starting values. If the starting point is based on the true expected value, then the hypothetical scenario reflects the real-world decision context (Hauber and Coulter, 2020).
- **Shift-framing effects?** Like anchoring, shift-framing effects are present when the respondents’ choices are influenced by the difference in the levels between the target and reference options presented in the first question. Larger (smaller) initial differences between the starting values may result in larger threshold values for the key attribute and vice versa. Tests for anchoring could be explored by varying the difference between starting values in the initial target and reference options.

Some TT studies have also investigated the test-retest reliability of responses by repeating the TT exercises with the same sample a few weeks or months later (Percy and Llewellyn-Thomas, 1995; Brundage, Davidson and Mackillop, 1997; Kennedy et al., 2008). However, in the wider preference literature, it has been noted that there are several caveats associated with using test-retest reliability as an indicator of validity, notably that preferences in the interluding period could be affected by external shocks, experiences, or simply the act of thinking about the choices made in the original exercise (Mørkbak and Olsen, 2015).

It may also be useful to explore the face validity of responses, to ensure respondents’ understanding and to check for protest responses. A protest response may occur when a respondent is trying to influence the decision maker by failing to reveal their true threshold.

### 4.6.4 TT-specific aspects of framework component 2 ‘design: sample definitions—justifying the sample size’

#### 4.6.4.1 Points to consider regarding the sample size of a TT study

There is no specific power calculation to determine sample size in TT studies without knowing the expected threshold value a priori. Most TT studies are conducted with 100 or fewer respondents, and substantially smaller samples (between 20 and 42 respondents) have been used successfully in

previous studies ([Dales et al., 1999](#); [Sung et al., 2004](#); [Steures et al., 2005](#); [Gupta et al., 2016](#)). Although there is a lack of clear guidance on sample size estimation, it is generally assumed that a minimum of 50 responses per TT choice set is needed to estimate a threshold value in each threshold exercise.

#### **4.6.5 TT-specific aspects of framework component 2 ‘design: instrument design – consideration of the appropriate number of attributes and attribute levels; patient burden issues’**

##### **4.6.5.1 Points to consider regarding the choice of attributes**

Two key points of consideration in a TT are (1) the target and reference treatments and (2) the attribute(s) to be varied ([Kattan, 2009](#)). The target treatment is typically the new intervention, and the reference treatment is the current standard of care, which could also describe no treatment. The key attribute(s) of interest are those that differentiate the target treatment and provide information about the thresholds of interest, and this is typically the attribute to be varied.

Although a TT study can include multiple attributes, this results in more question series. Too many series may induce respondent fatigue, resulting in poorer quality preference data.

Points to Consider in the Range of Levels

The levels in a TT study reflect the range of theoretical minimums and maximums for the attribute and intervals in between. For some studies, respondents may have a preference above the maximum presented (e.g., a very high risk threshold). In these instances, respondents may be asked to state their threshold directly.

If the level of a fixed attribute is 100%, respondents in the TT study are essentially presented with a modified standard gamble choice. In this instance, the respondent is required to trade off certainty with chance, and if Kahneman and Tversky’s “Certainty Effect” holds, the target option may be under-weighted and MAR estimates downwardly-biased ([Kahneman and Tversky, 1979](#); [Rodríguez-Míguez, Pinto-Prades and Mosquera-Nogueira, 2019](#)).

##### **4.6.5.2 Points to consider in the range of levels**

The experimental design refers to the specific attributes and levels used in the choice tasks that responders complete. Because each TT series typically varies only one key attribute, the design is determined by the respondent’s first choice. The levels in the initial choice question are therefore important and can minimize potential biases. When using TT to estimate preferences for multiple attributes, researchers should randomize the TT series to avoid ordering effects.

These can be divided into three categories: methodological factors, participant factors, and feasibility factors and are summarized in [Table 4-2](#). See also Applications of preference data to inform medical product decision making.

**Table 4-2 Categories of points to consider when selecting a patient preference elicitation method**

Category	Description	Points to Consider
<b>Methodological</b>	Factors related to the specific requirements and capabilities of the method. Methodological factors primarily address the ability of the method to answer the research question and the extent to which the quality of the data can be assessed. Methodological factors can also influence the suitability of a method for different types of participants and the feasibility of implementing the method given resource constraints of the study sponsor.	<ol style="list-style-type: none"> <li>1. Does the method provide estimates at the level of the individual, the sample, or both?</li> <li>2. Can the method account for and/or test for sources of preference heterogeneity?</li> <li>3. What minimum sample size is required to implement the method?</li> <li>4. Can the method elicit preferences for individual treatment features or attributes? If so, how many features or attributes can be evaluated?</li> <li>5. Can the method elicit preferences over multiple levels of each feature or attribute?</li> <li>6. Does the method elicit relative importance, trade-offs, or both?</li> <li>7. If the method can be used to estimate trade-offs, can it estimate trade-offs among multiple attributes simultaneously (or are trade-offs estimated pairwise)?</li> <li>8. Can the research method accommodate interactions between treatment features or attributes?</li> </ol>
<b>Participant</b>	Factors related to the characteristics of the patient population, the ability of individual patients to participate in the preference elicitation exercise, and the ability of the available patient population to provide the type and amount of information required by the method.	<ol style="list-style-type: none"> <li>1. Do patients have the cognitive capacity to successfully complete the preference elicitation exercise?</li> <li>2. Is the sample large enough to provide enough information to implement the method?</li> </ol>
<b>Feasibility</b>	Factors related to the expertise required to implement the method, conduct the study, and analyze the data.	<p>Is specific methodological expertise to:</p> <ul style="list-style-type: none"> <li>• to design the data collection materials?</li> <li>• to implement the study and acquire the necessary data?</li> <li>• to analyze data and interpret the results of the analysis?</li> </ul>

## 5 Points to Consider for methods selection

The previous section described multiple methods for eliciting patient preferences and includes ‘points to consider’ when evaluating the suitability of each method. The purpose of this section is to consolidate and categorize the points to consider and to evaluate the five methods described in the previous section relative to these points to consider to the extent possible.

In 2015, the Medical Device Innovation Consortium (MDIC) noted that: “Designing and implementing a preference study is dependent on numerous considerations, including the level of existing knowledge about benefits and risks in a particular clinical situation, the ability of each method to provide the type of patient preference information needed for the particular benefit-risk assessment, and the resources and experience of the organization undertaking the study”. Designing and implementing a patient preference study does not follow a cookbook process but requires judgment on the part of the organization undertaking the study.” (MDIC Benefit-Risk Framework, p51). Although a great deal of work on methods and applications of patient preference methods has been conducted since the MDIC report was published, these statements remain true (Van Overbeeke E, 2019). Choosing the appropriate method for eliciting patient preferences in any situation, whether the information is intended to inform regulators, HTA bodies, or other decision makers, is context specific and depends on a number of factors (Whichello et al., 2020). Therefore, rather than providing a decision tree or other algorithm for determining which method is most appropriate in which situation, we have identified a set of points to consider when selecting a quantitative patient preference method.

These can be divided into three categories: methodological factors, participant factors, and feasibility factors and are summarized in Table 5-1. See also section 3.4.3.3: Applications of preference data to inform medical product decision making.

**Table 5-1 Categories of points to consider when selecting a patient preference elicitation method**

Category	Description	Points to Consider
<b>Methodological</b>	Factors related to the specific requirements and capabilities of the method. Methodological factors primarily address the ability of the method to answer the research question and the extent to which the quality of the data can be assessed. Methodological factors can also influence the suitability of a method for different types of participants and the feasibility of implementing the method given resource constraints of the study sponsor.	<ul style="list-style-type: none"> <li>• Does the method provide estimates at the level of the individual, the sample, or both?</li> <li>• Can the method account for and/or test for sources of preference heterogeneity?</li> <li>• What minimum sample size is required to implement the method?</li> <li>• Can the method elicit preferences for individual treatment features or attributes? If so, how many features or attributes can be evaluated?</li> <li>• Can the method elicit preferences over multiple levels of each feature or attribute?</li> <li>• Does the method elicit relative importance, trade-offs, or both?</li> <li>• If the method can be used to estimate trade-offs, can it estimate trade-offs among multiple attributes simultaneously (or are trade-offs estimated pairwise)?</li> <li>• Can the research method accommodate interactions between treatment features or attributes?</li> </ul>
<b>Participant</b>	Factors related to the characteristics of the patient population, the ability of individual patients to	<ul style="list-style-type: none"> <li>• Do patients have the cognitive capacity to successfully complete the preference elicitation exercise?</li> </ul>

	participate in the preference elicitation exercise, and the ability of the available patient population to provide the type and amount of information required by the method.	<ul style="list-style-type: none"> <li>• Is the sample large enough to provide enough information to implement the method?</li> </ul>
<b>Feasibility</b>	Factors related to the expertise required to implement the method, conduct the study, and analyze the data.	<p>Is specific methodological expertise to:</p> <ul style="list-style-type: none"> <li>• to design the data collection materials?</li> <li>• to implement the study and acquire the necessary data?</li> <li>• to analyze data and interpret the results of the analysis?</li> </ul>

## 5.1 Methodological factors

Methodological factors are related to the specific requirements and capabilities of the method, and primarily address the ability of the method to answer the research question and the extent to which the quality of the data can be assessed. Methodological factors can also influence the suitability of a method for different types of participants and the feasibility of implementing the method given resource constraints of the study sponsor. Each of these factors, along with the performance of each method described in Chapter 4 and the performance of each method on each methodological factor is summarized to the extent possible given available evidence in [Table 5-2](#).

**Table 5-2 Assessment of methods relative to methodological points to consider**

Points to Consider	DCE	BWS1	BWS2	Threshold	SW
Does the method provide estimates at the level of the individual, the sample, or both?	Sample	Both	Both	Both	Both
Can the method account for and/or test for sources of preference heterogeneity?	Yes	Yes	Yes	Yes	Yes
What minimum sample size is required to implement the method?	Tends to be larger - typically greater than 100	Depends on the method used to analyse the data -- typically at least 50, but could be smaller	Depends on the method used to analyse the data -- typically at least 50, but could be smaller	Tends to be somewhat smaller -- typically at least 50	Tends to be much smaller - can be 5-10 in facilitated groups, could be larger if survey approaches are used
Can the method elicit preferences for individual treatment features or attributes? If so, how many features or attributes can be evaluated? <sup>2</sup>	Yes, Tends to be more limited	Yes, Tends to be less limited	Yes, Tends to be more limited	Yes, Tends to be less limited	Yes, Tends to be less limited
Can the method elicit preferences over multiple levels of each feature or attribute?	Yes	No	Yes	No	No
Does the method elicit relative importance, tradeoffs, or both?	Estimates rates of tradeoff that can be used to calculate relative importance	Estimates relative importance which can be used to infer rates of tradeoff	Estimates relative importance which can be used to infer rates of tradeoff	Estimates rates of tradeoff that can be used to calculate relative importance	Estimates relative importance which can be used to infer rates of tradeoff
If the method can be used to estimate tradeoffs, can the method estimate tradeoffs among multiple attributes simultaneously (or are tradeoffs estimated pairwise)?	Yes	--	Yes	No	--
Can the research method accommodate interactions between treatment characteristics?	Yes	No	Yes	No	No

Note: DCE = discrete choice experiment; BWS = best-worst-scaling; SW = swing weighting

### **5.1.1 Does the method provide estimates at the level of the individual, the sample, or both?**

Some preference elicitation methods are designed to capture preferences of a sample of patients rather than inferring the preferences of the sample by aggregating information about the preferences of the individual patient. In most cases, preference-elicitation methods that require an experimental design (e.g., DCE, BWS) yield estimates of preferences at the sample or population level rather than at the individual level. While there are post-estimation methods for inferring individual preferences from sample or population-level estimates, the experimentally designed preference elicitation methods first require estimation of preferences at the sample level (Greene, Hensher, and Rose, 2005). However, there are simple-sum approaches that can be applied to BWS (especially BWS Case 1BWS1) to approximate preferences for each individual. In contrast, methods such as TT and SW yield preference information at the level of the individual patient and then aggregate these results across the sample to also provide sample-level preference estimates.

### **5.1.2 Can the method account for and/or test for sources of preference heterogeneity?**

All of the methods described in the previous chapter can account for preference heterogeneity. For those methods that provide estimates at the individual level, heterogeneity can be assessed directly by observing differences among individuals. For those methods that provide only sample-level estimates, heterogeneity must be modeled statistically (for example using random parameters logit, latent class analysis, or interactions between respondent characteristics and attribute levels in the utility function). Relating preference heterogeneity to characteristics of individual patients is more direct when the method provides preference estimates at the individual level. When heterogeneity is modeled statistically, the extent to which individual characteristics explain preference heterogeneity can be included in the model but is indirect. It is important to note that preference heterogeneity is related to representativeness but is not the same. Specifically, preference heterogeneity can often be evaluated even in small samples which would not be considered representative or even with larger samples that are not necessarily representative of the patient population.

### **5.1.3 What minimum sample size is required to implement the method?**

Sample size calculations are problematic in preference analysis because preference elicitation exercises are often used to estimate preferences for multiple outcomes simultaneously, resulting in multiple, often interrelated effect sizes to be estimated. In addition, sample size calculations require prior expectations on the magnitude of the effect size and the variability (heterogeneity) of the effect across the sample. Therefore, researchers must often rely on loose rules of thumb when determining sample size. Methods that require an experimental design (DCE, BWS) typically require larger minimum samples than methods that do not require an experimental design (threshold technique and swing weighting that can often be conducted with fewer respondents). In addition, for BWS Cases 1 and 2, the minimum required sample size may depend on the method used to analyse the data. Simple-sum approaches tend to require smaller sample sizes than parametric approaches to estimation. It is important to note that while some methods often require larger sample sizes than other methods, the specific minimum sample size required for any study is a function of other study

characteristics such as task difficulty and survey length and thus can vary among different studies using the same method.

#### **5.1.4 Can the method elicit preferences for individual treatment features or attributes? If so, how many features or attributes can be evaluated?**

All of the methods described in the previous chapter can be used to elicit preferences for individual treatment features or attributes. However, the number of features or attributes that can be evaluated differs across methods. DCEs and BWS Case 2 typically require respondents to consider the full set of attributes simultaneously. In addition, DCEs and BWS Case 2 tend to capture more information about each attribute because these methods include multiple levels of each attribute. Therefore, the number of attributes that can be included in a DCE or BWS Case 2 is typically more limited than can be included in a BWS Case 1, threshold technique, or swing-weighting exercise.

#### **5.1.5 Can the method elicit preferences over multiple levels of each feature or attribute?**

Typically, DCE and BWS case 2 include multiple levels for each attribute and therefore can capture the impact of changes in the levels of an attribute within a range. In contrast, BWS Case 1 effectively defines each attribute as a dichotomous variable – either present or absent – and therefore is limited to estimating the importance of an attribute rather than the importance of a change in the level of that attribute. The threshold technique is designed to identify the change in one attribute that exactly offsets a pre-specified change in the level of another attribute. Therefore, the threshold technique effectively sets a change in one attribute to be constant while allowing the levels of one attribute to vary. Swing weighting examines the relative importance of a swing between two levels (e.g. lowest to highest) and often, though not always, excludes intermediate levels.

#### **5.1.6 Does the method elicit relative importance, trade-offs, or both?**

Generally speaking, BWS case 1, BWS case 2, and Swing Weighting estimate relative importance directly. While trade-offs could potentially be inferred by the ratio of relative importance estimates, BWS case 1, BWS case 2 and Swing Weighting do not estimate trade-offs directly. DCE, BWS case 2, and Threshold Technique estimate trade-offs that can then be used to calculate relative importance.

#### **5.1.7 If the method can be used to estimate trade-offs, can the method estimate tradeoffs among multiple attributes simultaneously (or are tradeoffs estimated pairwise)?**

As noted above, DCE, BWS Case 2, and threshold technique can be used to estimate trade-offs. DCE and BWS Case 2 estimate trade-offs among multiple attributes simultaneously, so all attributes must be considered simultaneously in the preference-elicitation question. This limits the number of attributes that can be included in the preference-elicitation exercise. Threshold technique elicits pairwise trade-offs between two attributes, so it must be repeated to elicit each pairwise comparison. Because each pairwise comparison is estimated separately, the number of attributes that can be included in a threshold technique exercise is not limited by the fact that all attributes must be considered simultaneously; however, because the exercise must be repeated for each pairwise



comparison, there is a limit to the number of questions a single respondent can answer and thus a limit to the number of attributes that can be evaluated by any single respondent.

### **5.1.8 Can the research method accommodate interactions between treatment characteristics?**

To estimate interactions between attributes – the non-independent, cumulative effect of simultaneous changes in the levels of two attributes, the attributes must be considered simultaneously with each other and with other attributes in the study. Therefore, only those methods that require the simultaneous evaluation of multiple attributes – DCE and BWS Case 2 – enable attribute interactions. In addition, it should be noted that incorporating attribute interactions may impact participant factors because doing so may result in requiring a larger minimum sample size or requiring respondents to answer additional choice questions.

## **5.2 Participant factors**

Participant factors related to the characteristics of the patient population, the ability of individual patients to participate in the preference elicitation exercise, and the ability of the available patient population to provide the type and amount of information required by the method.

### **5.2.1 Do patients have the cognitive capacity to complete the preference elicitation exercise successfully?**

There are several reasons why patients may have difficulty completing a preference-elicitation exercise, which may be related to cognitive function, education, literacy and numeracy. Some methods may be easier for such patients to complete than others. In addition, the preference elicitation materials should be developed to be appropriate to the patients in the sample. In some cases (e.g. when the patient is cognitively impaired or a child), the patient may lack the ability to complete a patient preference exercise and preferences must be elicited instead from a surrogate such as a parent or care partner.

### **5.2.2 Is the sample large enough to provide enough information to implement the method?**

In some cases, the patient population is small or it is difficult to access patients who meet the target criteria. In such cases, it may not be feasible to use a method such as DCE that requires an experimental design and a relatively large sample size. Likewise, when larger samples of patients are available, methods that are geared toward smaller samples such as swing weighting may not be desirable because other methods are able to elicit preferences from a more representative sample more efficiently.

## **5.3 Feasibility Factors**

Feasibility factors related to the expertise required to implement the method, conduct the study, and analyze the data.

### **5.3.1 Is specific methodological expertise required to design the data collection materials?**

Most preference elicitation methods require experience and expertise but the type of expertise may differ among methods. For example, designing a DCE survey requires expertise in experimental design and survey methods. SW requires expertise in developing interview guides, or facilitator guides and meeting materials for decision conferencing. In all cases, an understanding of the analysis of the data that will be generated by the data collection material is critical.

### **5.3.2 Is specific methodological expertise required to implement the study and acquire the necessary data?**

Survey methods require experience in survey implementation and may require expertise in survey programming if the survey will be administered online. Swing weighting requires expertise in interviewing and decision conferencing.

### **5.3.3 Is specific methodological expertise required to analyze data and interpret the results of the analysis?**

Expertise in statistical analysis will be required for any study, but methods that require more complex statistical analysis, especially those based on an experimental design (e.g. DCE and BWS) may require specialized statistical expertise.

## **5.4 PREFER case studies**

Two completed case studies were used as examples to demonstrate retrospectively how the points to consider for methods selection can be used. The outcome of the learnings are summarized in [Table 5-3](#) and [Table 5-4](#).

**Table 5-3 Learnings from the Pain Case Study on Points to Consider Regarding Methods Selection**

Methods Compared: Discrete-Choice Experiment (DCE) and Best-Worst Scaling (BWS) Case 1

Point to Consider	DCE	BWS Case 1
<b>Does the method provide estimates at the level of the individual, the sample, or both?</b>	The objective of the study was to estimate preferences from a representative sample of patients. Either method would enable this.	
<b>Can the method account for and/or test for sources of preference heterogeneity?</b>	Both methods enabled preference heterogeneity to be accounted for and the source to be tested.	
<b>What minimum sample size is required to implement the method?</b>	The minimum sample size in each country was determined for the DCE using Yang et al 2015 and then applied to the BWS case 1. It is unclear whether a smaller sample size would have been required for the BWS case 1 alone.	
<b>Can the method elicit preferences for individual treatment features or attributes? If so, how many features or attributes can be evaluated?</b>	The DCE can elicit preferences for individual treatment features and was used to elicit preferences for efficacy, one key risk from each of three treatment classes, mode of administration, and out-of-pocket cost	The DCE can elicit preferences for individual treatment features and was used to elicit preferences for efficacy, one key risk from each of three treatment classes, mode of administration, and out-of-pocket cost
<b>Can the method elicit preferences over multiple levels of each feature or attribute?</b>	The DCE can elicit preferences over multiple levels of each feature, enabling the elicitation of preferences for changes in the levels of all attributes.	The DCE can elicit preferences over multiple levels of each feature, enabling the elicitation of preferences for changes in the levels of all attributes.
<b>Does the method elicit relative importance, tradeoffs, or both?</b>	Trade-offs between efficacy and risk and between all attributes and cost were estimated, as was the relative importance of attributes.	Trade-offs between efficacy and risk and between all attributes and cost were estimated, as was the relative importance of attributes.
<b>If the method can be used to estimate tradeoffs, can the method estimate tradeoffs among multiple attributes simultaneously (or are tradeoffs estimated pairwise)?</b>	Estimated trade-offs among all attributes simultaneously.	N/A
<b>Can the research method accommodate interactions between treatment characteristics?</b>	Yes, although no attribute interactions were specified or estimated.	No
<b>Concluding statement on choice of method:</b> The two methods served different purposes in this study and each was appropriate to its purpose. Specifically, the DCE provided information about trade-offs for multiple attributes simultaneous that enabled the estimation of both trade-off measures (e.g. maximum acceptable risk and willingness-to-pay), as well as the relative importance of each attribute. Because the DCE was limited in the number of attributes that could be included, it focused on the key attributes identified by the study team. Additional attributes were included using BWS case 2, which enabled the inclusion of many more attributes. For the attributes included in the BWS case 1, only relative importance could be estimated; however, this was considered sufficient by the study team because the attributes in the BWS case 2 were of secondary importance to the research question.		

**Table 5-4 Learnings from the Multiple Myeloma (MM) Case Study on Points to Consider Regarding Methods Selection**

Methods Compared: Discrete-Choice Experiment (DCE) and Swing Weighting (SW)

Point to Consider	DCE	SW
<b>Does the method provide estimates at the level of the individual, the sample, or both?</b>	<b>Both;</b> DCE provides both individual attribute weights as well as population attribute weights. For the latter, mixed logit (MXL) and latent class (LC) models will be used to estimate the distribution of the attribute weights in the population.	<b>Both;</b> SW can provide both individual and sample estimates. SW is a direct valuation technique, meaning that as opposed to DCE, the answers to the SW questions directly provide the individual relative weight for each participant and each attribute. SW does not require statistical modelling to calculate per-participant relative attribute weights. In addition to individual participant estimates, it will be sought to estimate the distribution of the attribute weights in the population. According to Tervonen et al <sup>1</sup> , the Dirichlet distribution is particularly interesting for this purpose given its support is the simplex (i.e., the full feasible space of attribute weights when they are normalized to sum to unity). Therefore, this study will explore the use the Dirichlet distribution <sup>2</sup> to estimate the relative importance of the attributes in the population.
<b>Can the method account for and/or test for sources of preference heterogeneity?</b>	<b>Yes;</b> the MXL and LC model allow analysing preference heterogeneity. In addition, cluster analysis will be used to cluster the individual attribute weights obtained from the MXL model to detect patient groups with homogeneous preferences.	<b>Yes;</b> heterogeneity can be analysed using standard statistical methods external to SW such as cluster analysis.
<b>What minimum sample size is required to implement the method?</b>	De Bekker-Grob et al. <sup>3</sup> describe a technique for calculating upfront the minimum sample size for parameter estimates in DCEs. However, to use this approach, the following needs to be known: i) significance level; ii) statistical power level; iii) statistical model used in the DCE analysis; iv) initial belief about the parameter values and v) the DCE design (defined by the number of choice sets, the number of alternatives per choice set, the number of attributes, and the combination of the attribute levels in each choice set). Since these elements were not known prior to launching the survey, it was not possible to estimate the sample size using the technique described by de Bekker-Grob et al. A practical solution described by de Bekker-Grob et al. that does not require any sample size calculations is to <b>maximize the sample size</b> in order to facilitate in-depth analysis. Therefore, this study seeks to recruit <b>as many participants as possible</b> .	<b>Not relevant for estimation of individual attribute weights through SW;</b> no minimum sample size needs to be established upfront since SW captures per-patient preferences in an exact manner and the required sample size does not depend on the magnitudes of utility <sup>2</sup> .
<b>Can the method elicit preferences for</b>	<b>Yes;</b> DCE is based upon the premise that choices among sets of alternative profiles are triggered by differences in the levels of the	<b>Yes;</b> SW directly captures individual patient preferences for attributes. Respondents of the SW experiment are asked for their

Point to Consider	DCE	SW
<b>individual treatment features or attributes? If so, how many features or attributes can be evaluated?</b>	attributes that define these profiles. A typical DCE question is “which of the (two) treatments do you prefer”. DCE allows researchers to quantify the impact of changes in attribute levels on choice by asking respondents to make choices among sets of profiles in a series of choice questions. The estimates of these impacts reflect the strength of preference for changes in attribute levels. The difficulty of completing DCE questions increases when the number of attributes increases. However, when there are too few attributes, there is the risk of excluding a potentially important attribute to patients’ decisions. In this study, eleven patient-relevant and plausible attributes were identified in the qualitative phase. De Bekker-Grob et al. <sup>3</sup> describe that DCE studies showing between four and six attributes were common in the published healthcare-related DCE studies in 2012. To accommodate for eleven attributes without burdening the patients, a partial-profile design was chosen that shows <b>4 of the 11</b> attributes in each DCE task.	preferences for moving from the worst to the best levels (swings) on each attribute. In order to obtain attribute values, two types of SW questions are asked. One of the SW question types asks patients to rank the changes from worst to best level for each of the attributes according to their relative importance. The other SW question type asks patients to assign a weight (“point allocation”) to each of these changes. The answers to these two questions directly provide the relative weight for each of the attributes (namely the importance of each change from worst to best level for each of the attributes) per participant. As this study aims to compare attribute weights identified via SW vs DCE, the two SW questions will be developed using the same attributes as used for the DCE questions. For the same reason, the levels for the SW questions will be the same levels used in the DCE questions.
<b>Can the method elicit preferences over multiple levels of each feature or attribute?</b>	<b>Yes.</b> The DCE questions ask respondents to choose between hypothetical MM treatment profiles and each hypothetical treatment profile is defined by a set of attributes. Each attribute has levels over which it can vary, and the hypothetical treatment profiles are defined by different combinations of attribute levels. The type and number of the levels selected in this study considered: i) that the levels needed to be clinically and patient-relevant, i.e. expected to be seen in clinical development and clinical practice, ii) the expected linearity between participants’ choice behavior and changes in attribute levels and, iii) the impact of increasing the number of levels on the required sample size <sup>3</sup> . Since this study aims to compare DCE with SW, and the SW questions only elicit preferences for changes between two attribute levels, two levels for each of the attributes in the DCE choice sets were included.	Yes; the SW questions can elicit preferences for an improvement (swing) between two attribute levels.
<b>Does the method elicit relative importance, tradeoffs, or both?</b>	<b>Both. Regarding relative importance,</b> Conditional Logit (CL), Latent Class (LCA), Hierarchical Bayes (HB), and Mixed Logit (MXL), can be fitted to the data to explore which one has the best fit. The 95% confidence interval and standard error for the mean relative attribute importance for each of the attributes (weights) of the best fitting model will be obtained using standard statistics such as standard t statistics. <b>Regarding trade-offs,</b> DCE allows determining the Maximum Acceptable Risk (MAR) and the Minimum	<b>Both. Regarding relative importance,</b> one of the SW question types asks patients to rank the changes from worst to best level for each of the attributes according to their relative importance. The other SW question type asks patients to assign a weight (“point allocation”) to each of these changes. The answers to these two questions directly provide the relative importance for each of the attributes (namely the importance of each change from worst to best level for each of the attributes) per participant. <b>Regarding trade-</b>

Point to Consider	DCE	SW
	Acceptable Benefit (MAB) that patients express for different treatment alternatives; MAR will indicate the maximum amount of risk (in this study; specific MM symptoms and MM treatment side-effects) they may be willing to tolerate in exchange for a specific level of benefit (in this study; life expectancy and treatment response). MAB indicates the minimal amount of benefit participants require in order to tolerate for a specific level of risk.	<b>offs</b> , the ranking SW question will ask patients to trade-off between improvements regarding the attributes. In particular, patients need to trade-off between positive treatment outcomes, i.e., efficacy outcomes (improvements in life expectancy and treatment response) and improvements in negative outcomes (improvements of symptoms and side-effects). The second type of SW question asks patients to distribute 100 points between the same improvements, requiring them to trade-off between improvements regarding the same efficacy outcomes, symptoms, and side-effects.
<b>If the method can be used to estimate tradeoffs, can the method estimate tradeoffs among multiple attributes simultaneously (or are tradeoffs estimated pairwise)?</b>	DCE can estimate trade-offs among multiple attributes simultaneously.	SW can estimate trade-offs among multiple attributes simultaneously.
<b>Can the research method accommodate interactions between treatment characteristics?</b>	<b>Yes</b> ; the DCE design in study this study will allow to investigate interactions between attributes. In particular, it will be investigated whether there are interactions between life expectancy and issues (symptoms and side effects) that strongly affect patients' daily feelings and activities, such as pain and mobility problems.	<b>No</b> ; to our knowledge, SW is not equipped to investigate interactions between attributes.

**Concluding statement on choice of method:**

This study aims to identify which treatment outcomes matter most to MM patients and therefore should be included in MM drug development and evaluation. Several methods are available for quantifying treatment preferences, with multiple options ranging from ranking exercises to trade-off methods. The importance of knowing how method selection affects preference study results has been highlighted in PREFER WP2 studies, where stakeholders were concerned about a lack of guidance on method selection and research that validates and compares different preference methods.

Therefore, this study compares two preference methods: DCE and SW. The reason for selecting DCE and SW is that: i) DCE and SW are among the most recommended methods quantifying the weight of treatment attributes according to patients; which is an aim of this study; ii) a comparison between SW and DCE is possible since both of them allow obtaining the relative weight of attributes for each of the participants, and their use in the same study will serve as a validation of the preference study results; it is expected that, if both methods measure the same preferences, the choice behavior of a given patient in a given DCE question does not systematically deviate from the weights the patient assigned to the attribute in the SW questions and vice versa; and iii) a larger number of methods would increase the survey length too much.

## 6 Conclusion

The field of patient preference elicitation is dynamically evolving in terms of both development of methods and in the accumulation of experience in conducting patient preference studies. Currently, there are no generally accepted recommendations to guide applicants in how to design, conduct and analyze patient preference studies that are fit for purpose for decision-making.

IMI PREFER was initiated with the overall goal of addressing this gap by strengthening patient-centric decision-making throughout the life cycle of medicinal products/devices. This has been achieved by developing evidence-based recommendations to guide industry, regulators, HTA bodies, reimbursement agencies, academia and healthcare professionals on how and when patient-preference studies should be performed and how the results can be used to support and inform decision-making. To achieve this goal, PREFER took a structured approach by:

- assessing the needs and expectations of stakeholders
- assessing and classifying patient preference methods based on stakeholder input
- selecting patient preference methods suitable for use at different decision points during the MPLC
- identifying and prioritizing research questions based on expert knowledge and stakeholder input
- conducting case studies to address high priority research questions
- systematically searching for historical case studies
- analyzing case study results that will be supplemented with information from the historical case studies.

This document provides three main tools to facilitate the planning, design, conduct, reporting and review of patient preference studies for decision-making, namely:

1. a framework for patient preference studies in [section 3](#)
2. 'points to consider' for the methods selection in section 5, together with additional details of five key qualitative methods in [section 4](#).

### 6.1 PREFER framework for patient preference studies

The PREFER framework is the backbone of the document. It builds on existing work in the field and takes into account frameworks from other fields to provide detailed processes that can be applied when using any method for the elicitation of patient preferences. Like the estimands concept introduced in [ICH E9 \(R1\) addendum \(2019\)](#) and the [PICO framework](#) described by [Huang et al \(2006\)](#), the PREFER framework guides the user through a structured series of steps. The framework consists of three main building blocks to ensure there is clarity and high standards throughout the process, namely:

1. Defining the preference study purpose and objectives
2. Planning, design and conduct of the preference study
3. Applying the preference data to inform medical product decision-making

The framework should not be understood as a cookbook for the planning, conducting and reporting of patient preference studies but is expected to facilitate the discussion between applicants and reviewers on the value of any particular preference study.

## 6.2 Overview of commonly used and well understood methods

A large number of methods have been used for preference 'exploration' and 'elicitation', and a systematic appraisal of 33 methods was conducted by [Soekhai et al \(2019\)](#) in the context of PREFER. These methods vary in their performance, their usefulness for different stages of development, and in the body of evidence available in the context of decision making in the MPLC. Five of these methods were identified as having the potential to meet the need of decision makers for almost all stages of the MPLC.

For this document five methods were selected to further detail on the planning, conduct and analysis of PPS as outlined by the PREFER framework for preference studies. These methods represent a set of suitable methods for the context of use, cover a different type of methods and come with a relevant body of evidence. This choice should not be seen as a recommendation to only select from one of these five methods. On the contrary, further research is encouraged and the principles laid out in this document should facilitate the evaluation of further methods and, thereby, build the foundation for a structured discussion between applicants and decision-makers. The choice of the method should be made in the context of the research question so that it aligns with the study purpose and objective as outlined in [Section 3.3.2.3](#).

## 6.3 Points to consider for the methods selection

Choosing the appropriate method for eliciting patient preferences in any situation is context specific and depends on a number of factors ([Whichello et al., 2020](#)). As in other situations, a fit-for-purpose PPS may be achieved by more than one method. The 'points to consider' for methods selection provide a structured approach that builds on initial work published by [MDIC \(2015\)](#) and also provides a structured comparison of the five methods discussed in this document. This approach provides a balance between clear guidance where justified and sufficient room for judgement as appropriate for the specific situation. It should be noted that the use of the general structure is not limited to the five methods discussed but serves as a guide for selecting any method in a specific situation and facilitates the justification of the choice of method in this situation.



## 7 Questions and Applicant's positions

### Question 1

The intended objectives of the PREFER framework for patient preference studies are to:

- Inform a preference study research team on key considerations when designing, conducting and applying the results of a fit-for-purpose preference study
  - Guide decision-makers when assessing and using preference study results to inform decision-making
  - Support the discussion between industry, regulators, HTA bodies and payers about preference studies intended to inform medical product decision-making
1. Do EMA and EUnetHTA agree that these are the appropriate principal objectives for the framework?
  2. Do EMA and EUnetHTA agree that the framework achieves its intended objectives?

### Applicant's position

The current lack of framework inhibits the adoption of patients' preferences in industry, regulatory and HTA/payer decisions. PREFER believes that the PREFER framework would provide a structure that could assist in the planning, conduct and analysis of a preference study, and assist both industry, regulators and HTA bodies in the assessment of preference study data in decision-making. The availability of such a framework is expected to promote the wider use of preference studies.

The PREFER framework is intended to promote a common understanding across regulators, HTA bodies/payer and industry about the topic of patient preference studies and to address CHMP feedback on the initial briefing book, which has been updated to add further information on how results from preference studies can be used to inform medical product decision-making. The PREFER framework covers the issues expected to be relevant to any patient preference study – namely the study purpose and study objectives (section 3.2); the approach to planning, conducting and reporting a preference study (section 3.3) and the use of the preference results to inform a medical product decision (section 3.4). These issues are independent of the choice of method for the preference study: for the issues where further method-specific information is required, this content is covered in the methods section of the briefing book. As with other methods used in clinical development, the field of preference research continues to evolve; however, this framework could help establish baseline expectations and support discussions between industry, regulators and HTA bodies/payers about preference studies, and hence support the broader aim of increasing patients' input to medical product decision-making.

### Question 2

Do EMA and EUnetHTA agree that the 'Points to Consider' on method selection, together with the additional details of five key quantitative methods, when applied appropriately within the PREFER framework, can support generating patient preference evidence to inform decision-making throughout the MPLC?

### Applicant's position

The PREFER framework provides guidance for conducting and implementing a patient preference study that ensures the information generated by such a study are patient-centered, reliable, and useful to decision-makers. Multiple preference elicitation methods exist, each having unique characteristics that may make it more or less suitable to providing useful evidence in specific situations. Therefore, researchers and users of patient preference data need to consider multiple factors when assessing the appropriateness of a method in a particular circumstance. The descriptions of five common patient preference methods and the points to consider when selecting a patient preference elicitation method provide a structured approach to assessing which preference elicitation method may be most appropriate in a given situation, ensuring the quality and utility of the resulting patient preference information.

### Question 3

Do EMA and EUnetHTA agree that if preference study results inform a regulatory decision or a HTA, then:

- (for regulatory decisions) the corresponding data could be included in the drug label as applicable, and the manner in which the study informed the decision could be included in the public assessment report?
- (for HTA bodies) the manner in which the study informed the assessment could be included in the HTA report?

### Applicant's position

The applicant proposes (within [section 3.4.3](#) of the PREFER framework) that if preference study results inform a regulatory or HTA body decision, then the corresponding data could be included in the Clinical Overview, the EPAR and/or drug label, or in the HTA report. This approach would:

- provide more transparency about how patients' perspectives have informed decision-making
- (for preference study information included in labels) provide information to patients and prescribers that could assist their decision-making.

The inclusion of preference data in the Clinical Overview is consistent with [ICH M4E\(R2\)](#), and the inclusion of preference data in the EPAR would be aligned with the EMA Day 80 Assessment report template. The inclusion of preference data in the label would be aligned with the [FDA Patient Preference Information guidance \(2016\)](#), which states that preference data that supports FDA approval should be described in the device label, and advises that including preference data in decision summaries can be helpful to both healthcare professionals and patients who need to make tricky benefit-risk trade-off decisions.

A description of how preference data could be incorporated into the EPAR, an HTA, and the label (including examples of how such data could be described) is included in [section 3.4.3](#).

## List of References

- Abelson J. et al. Public and patient involvement in health technology assessment: a framework for action *Int. J. Technol. Assess. Health Care*, 32. 2016;32(4):256-264.
- Airoldi M., Morton, A.: Portfolio decision analysis for population health. In: Salo, A., Keisler, J., Morton, A. (eds.) *Portfolio Decision Analysis: Improved Methods for Resource Allocation*, pp. 359–381. Springer, New York (2011). [\(Link\)](#). Accessed on March 2d, 2021.
- Aizaki H., *Stated Preferences Methods Using R*. 2015.
- Al-Faruque. FDA weighs patients' risk tolerance in approving obesity device. *The Pink Sheet Daily*. January 20 2015. Available at: <https://www.pharmamedtechbi.com/publications/the-pink-sheet-daily/2015/1/20/fda-weighs-patients-risk-tolerance-in-approving-obesity-device>. (2015). Accessed January 6, 2019.
- Angelis A. and Kanavos P., *Multiple Criteria Decision Analysis (MCDA) for evaluating new medicines in Health Technology Assessment and beyond: The Advance Value Framework*. *Soc Sci Med*. 2017 Sep;188:137-156.
- Angelyn O. et al. Patient preferences for ketamine-based antidepressant treatments in treatment-resistant depression: Results from a clinical trial and panel *Neurology, Psychiatry and Brain Research*, Volume 37, 2020, Pages 67-78, ISSN 0941-9500,
- Appelt KC et al. The decision making individual differences inventory and guidelines for the study of individual differences in judgment and decision-making research. *Judgm Decis Mak*. 2011; 6(3):252-262.
- Baltussen R. et al. Multi-criteria decision analysis to support HTA agencies – benefits, limitations and the way forward. *Value in Health*. 2019
- Bansback N. et al, Does Introducing Imprecision around Probabilities for Benefit and Harm Influence the Way People Value Treatments? *Med Decis Making* 2016; 36:490-502
- Barnett, A. S., et al. Net clinical benefit of rivaroxaban compared with warfarin in atrial fibrillation: Results from ROCKET AF. *Int J Cardiol* 2018. 257: 78-83
- Bilvick Tai BW, Bae YH, Le QA. A Systematic Review of Health Economic Evaluation Studies Using the Patient's Perspective. *Value in health: the journal of the International Society for Pharmacoeconomics and Outcomes Research*. 2016; 19(6):903-8.
- Birch et al., Development and formative evaluation of patient research partner involvement in a multi-disciplinary European translational research project. *Research Involvement and Engagement* volume 6, Article number: 6 (2020).
- Bouvy J et al., Use of patient preference studies in HTA decision making - a NICE perspective. *Patient*. 2020 Apr; 13(2): 145-149.
- Bradley EH, Curry LA, Devers KJ. Qualitative data analysis for health services research: developing taxonomy, themes, and theory. *Health Serv Res*. 2007;42(4):1758–1772.

Bridges JF, Hauber AB, Marshall D, et al. Conjoint analysis applications in health – a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. *Value Health*. 2011. p403-13.

Bridges JF et al., Can patients diagnosed with schizophrenia complete choice-based conjoint analysis tasks? *Patient*. 2011;4(4):267-75.

Bridges, JF, Dong, L, Gallego, G et al. (2012). Prioritizing strategies for comprehensive liver cancer control in Asia: a conjoint analysis. *BMC Health Serv. Res.*, 12, 376.

Bridges, JFP, Mohamed, AF, Finnern, HW, Woehl, A and Hauber, AB (2012). Patients' preferences for treatment outcomes for advanced non-small cell lung cancer. *Lung Cancer*, 77, 224-231.

Broekhuizen H. et al., A review and classification of approaches for dealing with uncertainty in multi-criteria decision analysis for healthcare decisions. *Pharmacoeconomics*. 2015 May;33(5):445-55.

Broekhuizen H. et al., Estimating the value of medical treatments to patients using probabilistic multi criteria decision analysis. *BMC Med Inform Decis Mak*. 2015 Dec 2;15:102.

Broekhuizen H. et al., Weighing Clinical Evidence Using Patient Preferences: An Application of Probabilistic Multi-Criteria Decision Analysis. *Pharmacoeconomics*. 2017 Mar;35(3):259-269.

Broekhuizen H. et al., Assessing Lung Cancer Screening Programs under Uncertainty in a Heterogeneous Population. *Value Health*. 2018 Nov;21(11):1269-1277.

Brooker AS et al. Quantitative patient preference evidence for health technology assessment: A case study. *International Journal of Technology Assessment in Health Care*. 2013;29(3):290-300.

Brundage MD., Davidson JR. and Mackillop WJ., Trading treatment toxicity for survival in locally advanced non-small cell lung cancer. *J Clin Oncol*. 1997 Jan;15(1):330-40.

Burton N. et al., Best-worst scaling improves measurement of first impressions. *Cogn Res Princ Implic*. 2019 Sep 23;4(1):36.

Cheung KL. Et al., Using Best–Worst Scaling to Investigate Preferences in Health Care. July 2016 [Pharmacoeconomics](#) 34(12).

Cheung KL. Et al., Comparison of statistical analysis methods for object case best-worst scaling. *J Med Econ*. 2019 Jun;22(6):509-515.

Christiaens W, Kohn Let al. Models for citizen and patient involvement in health care policy - Part I: exploration of their feasibility and acceptability. *Health Services Research (HSR)*. Brussels: Belgian Health Care Knowledge Centre (KCE); 2012. Report No.: KCE Reports 195C. D/2012/10.273/105.

Council for International Organizations of Medical Sciences (CIOMS), Working Groups, <https://cioms.ch/working-groups/working-group-xi-patient-involvement/> Accessed March, 2d 2021.

Clark MD et al. (2014) Discrete choice experiments in health economics: a review of the literature. *Pharmacoeconomics*. 32(9):883-902.

Coast J et al., Valuing the ICECAP capability index for older people. *Soc Sci Med*. 2008 Sep;67(5):874-82.

Cook et al., Patient preference studies during early development - Aligning stakeholders to ensure development plans meet patient needs. *Front Med (Lausanne)*. 2019 Apr 24;6:82.

Cook et al., 2019, Impact of cough and mucus on COPD patients: primary insights from an exploratory study with an Online Patient Community, *International Journal of Chronic Obstructive Pulmonary Disease* 2019;14 1365–1376

Cozmuta R et al., Variability of the impact of adverse events on physicians' decision making. *BMC Med Inform Decis Mak.* 2014 Sep 25;14:86.

Craig B et al., Health Preference Research: An Overview. *Patient.* 2017 Aug;10(4):507-510.

Crocker JC et al. Is it worth it? Patient and public views on the impact of their involvement in health research and its assessment: a UK-based qualitative interview study. *Health Expectations.* 2016:1-10

Dales RE et al., Intubation and mechanical ventilation for COPD: development of an instrument to elicit patient preferences. *Chest.* 1999 Sep;116(3):792-800

Dalziel K. et al., Feasibility, Validity and Differences in Adolescent and Adult EQ-5D-Y Health State Valuation in Australia and Spain: An Application of Best-Worst Scaling. *Pharmacoeconomics.* 2020 May;38(5):499-513

Danner M et al., Integrating patients' views into health technology assessment: Analytic hierarchy process (AHP) as a method to elicit patient preferences. *Int J Technol Assess Health Care.* 2011 Oct;27(4):369-75.

De Bekker-Grob EW, Ryan M et al. Discrete choice experiments in health economics: a review of the literature. *Health Econ.;* (2012) 21(2):145-72.

De Bekker-Grob EW, Donkers B. et al. Sample size requirements for discrete choice experiments in healthcare: a practical guide. *Patient* (2015), 8, 373-384.

De Bekker-Grob EW, Swait JD et al. Are Healthcare Choices Predictable? The Impact of Discrete Choice Experiment Designs and Models. *Value Health.* 2019 Sep;22(9):1050-1062. doi: 10.1016/j.jval.2019.04.1924. Epub 2019 Jun 8.

De Bekker-Grob EW et al., Giving Patients' Preferences a Voice in Medical Treatment Life Cycle: The PREFER Public–Private Project. *Patient.* 2018 Jun;10(3):263-266.

De Bekker-Grob EW et al., Can healthcare choice be predicted using stated preference data?. *Soc Sci Med.* 2020 Feb;246:112736

Determann D et al. Designing unforced choice experiments to inform health care decision making: Implications of using opt-out, neither or status quo alternatives in discrete choice experiments. *Medical Decision making* (2019), 39(6), 681-692.

Devereaux PJ et al., Differences between perspectives of physicians and patients on anticoagulation in patients with atrial fibrillation: observational study. *BMJ.* 2001 Nov 24;323(7323):1218-22.

Diaby V et al., How to use multi-criteria decision analysis methods for reimbursement decision-making in healthcare: a step-by-step guide. *Expert Rev Pharmacoecon Outcomes Res.* 2014 Feb;14(1):81-99.

Dirksen C et al., The use of research evidence on patient preferences in health care decision-making: issues, controversies and moving forward. *Expert Rev Pharmacoecon Outcomes Res.* 2014 Dec;14(6):785-94.

Durbach I and Stewart T., Using expected values to simplify decision making under uncertainty. April 2009 [Omega](#) 37(2):312-330

Edwards W., 1977. How to use multi-attribute utility measurement for social decision-making. *IEEE Transactions on Systems, Man, and Cybernetics SMC-7:5*, 326–340.

Edwards W. and Barron F., SMARTS and SMARTER: Improved Simple Methods for Multi-attribute Utility Measurement. Volume 60, Issue 3, December 1994, Pages 306-325

EMA, ICH E9 Statistical principles for clinical trials. Note for guidance on statistical principles for clinical trials. September 1998.

EMA, The patient's voice in the evaluation of medicines. European Medicines Agency, Stakeholders and Communication Division; 2013. Report No.: EMA/607864/2013.

EMA, ICH guideline M4E(R2) - Common technical document for the registration of pharmaceuticals for human use – Efficacy. October 6, 2015. Report No.: EMA/CPMP/ICH/2887/1999.

EMA, Day 80 assessment report overview and D120 LOQ template\_en Rev 10.19.

EMA, EMA Regulatory Science to 2025. Strategic Reflection. 2020

EPF Patient Involvement in Health Technology Assessment in Europe: Results of the EPF Survey EPF 2013

European Commission Volume 2C Notice to Applicants – SmPC. Notice to applicant. September 2019.

Evans S.R. et al Desirability of Outcome Ranking (DOOR) and Response Adjusted for Duration of Antibiotic Risk (RADAR). *Clin Infect Dis.* 2015. **61**(5): 800-806

Evers P et al., The importance of early access to medicines for patients suffering from rare diseases. *Frontiers in Medicine.* 2016;13(2):5-8.

Facey K. et al., Patient Involvement in Health Technology Assessment. 2017

Fairchild A.O. What is clearance worth? Patients' stated risk tolerance for psoriasis treatments, 2017 *Journal of Dermatological Treatment*, 28:8, 709-715,

Felli JC et al., A multiattribute model for evaluating the benefit-risk profiles of treatment alternatives. *Med Decis Making.* Jan-Feb 2009;29(1): 104-15.

FDA, FDA News Release. FDA approves first-of-kind device to treat obesity. January 14, 2015.

FDA, CDRH, 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. 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; 2016.

FDA, Patient and provider views on clinical endpoints: a qualitative preference study of minimally invasive glaucoma surgical devices. January 2016.

- FDA, Comparing qualitative and quantitative approaches to eliciting patient preference: a case study on innovative upper limb prosthesis. 2016
- FDA, Approval letter for NXStage System One. Department of health and human services, FDA. August 24, 2017.
- FDA, Patient-Focused Drug Development: Collecting Comprehensive and Representative Input - Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders - DRAFT GUIDANCE. In: U.S. Department of Health and Human Services Food and Drug Administration CfDEaRC, Center for Biologics Evaluation and Research (CBER), editor. 2018.
- FDA, CDER, The voice of patient. A series of reports from the U.S. Food and Drug Administration's Patient-Focused Drug Development Initiative. Autism. January 2018.
- FDA, CBER FY 2018 Report from the Director. Center for Biologics Evaluation and Research (CBER), FDA. 2018.
- FDA, Summary of Safety and Effectiveness data (SSED) Tula System. FDA. November 25, 2019.
- FDA, Patient Preference Information (PPI) in Medical Device Decision-Making. FDA. 2019.
- FDA, Joint Meeting of the Psychopharmacologic Advisory Committee (PDAC) and the Drug Safety and Risk Management Advisory Committee (DsaRM). FDA, Center for drug evaluation and research. February 12, 2019.
- FDA, FDA News Release (2019). FDA approves system for the delivery of ear tubes under local anesthesia to treat ear infection. <https://www.fda.gov/news-events/press-announcements/fda-approves-system-delivery-ear-tubes-under-local-anesthesia-treat-ear-infection>. Accessed on March, 2d 2021.
- Finn A and Louviere JJ., Determining the appropriate response to evidence of public concern: The case of food safety. *J Public Policy Marketing*. 1992;11:12–25.
- Flynn, T. 2008. "Estimating Preferences for a Dermatology Consultation Using Best–Worst Scaling: Comparison of Various Methods of Analysis." *BMC Medical Research Methodology* 8 (1): 76.
- Flynn T. et al., Valuing citizen and patient preferences in health: recent developments in three types of best-worst scaling. *Expert Rev Pharmacoecon Outcomes Res*. 2010 Jun;10(3):259-67.
- Flynn T. et al., Scoring the Icecap-a capability instrument. Estimation of a UK general population tariff. *Health Econ*. 2015 Mar;24(3):258-69.
- Gale NK et al., Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Med Res Methodol*. 2013 Sep 18;13:117.
- Gallego, Gisselle, John FP Bridges, Terry Flynn, Barri M Blauvelt, and Louis W Niessen. 2012. "Using Best-Worst Scaling in Horizon Scanning for Hepatocellular Carcinoma Technologies." *International Journal of Technology Assessment in Health Care* 28 (3): 339–46.
- Genentech, Rituxan Hycela prescribing information. 2017
- Goetghebeur M et al., Bridging health technology assessment (HTA) and efficient health care decision making with multicriteria decision analysis (MCDA): applying the EVIDEM framework to medicines appraisal. *Med Decis Making*. Mar-Apr 2012;32(2):376-88

Gonzalez JM et al., Evaluating Risk Tolerance from a Systematic Review of Preferences: The Case of Patients with Psoriasis. *Patient*. 2018 Jun;11(3):285-300.

Greene, W., Hensher, D., & Rose, J. (2005). Using classical simulation- based estimators to estimate individual WTP values. In R. Scarpa & A. Alberini (Eds.), *Applications of simulation methods in environmental and resource economics* (pp. 17–34). Dordrecht, The Netherlands: Springer. (Chapter 2)

Gregory R., Failing L., Harstone M., Long G., McDaniels T., & Ohlson D. (2012). *Structured decision making: A practical guide to environmental management choices*. Oxford: Wiley- Blackwell.

Gutknecht M et al., A systematic review on methods used to evaluate patient preferences in psoriasis treatments. *J Eur Acad Dermatol Venereol*. 2016 Sep;30(9):1454-64.

Harrison M, et al. Preferences for treatments to prevent rheumatoid arthritis in Canada and the influence of shared decision-making. *Clin Rheumatol* (2020) 39, 2931–2941  
Hauber, A.B. et al. Benefits, Risk, and Uncertainty: Preferences of Antiretroviral-Naive African Americans for HIV Treatments. *AIDS patient care and STDs* (2009), Vol 23, Number 1.

Hauber AB. Et al., Quantifying Benefit–Risk Preferences for Medical Interventions: An Overview of a Growing Empirical Literature. *Appl Health Econ Health Policy*. 2013 Aug;11(4):319-29.

Hauber A.B, Fairchild A.O., Johnson FR. Quantifying Benefit–Risk Preferences for Medical Interventions: An Overview of a Growing Empirical Literature, *F. Appl Health Econ Health Policy* (2013) 11: 319.

Hauber AB. Et al., Patient preferences for treatments to delay bone metastases. *Prostate*. 2014 Nov;74(15):1488-97.

Hauber, A.B., Tunceli, K., Yang JC, et al. A Survey of Patient Preferences for Oral Antihyperglycemic Therapy in Patients with Type 2 Diabetes Mellitus. *Diabetes Ther* (2015) 6: 75.

Hauber, AB, Gonzalez JM, Groothuis-Oudshoorn, CGM, et al. Statistical Methods for the Analysis of Discrete Choice Experiments: A report of the ISPOR Conjoint Analysis Good Research Practices Task Force. *Value in Health* 2016, 19, 300-315.

Hauber AB et al., Hemodialysis patients' preferences for the management of anemia. July 2017, [BMC Nephrology](#) 18(1).

Hauber AB and Coulter J et al., Using the Threshold Technique to Elicit Patient Preferences: An Introduction to the Method and an Overview of Existing Empirical Applications. *Appl Health Econ Health Policy*. 2020 Feb;18(1):31-46.

Hensher DA et al., *Applied Choice Analysis: a Primer*. Cambridge University Press. 2005

Hess S. and Daly A., *Handbook of Choice Modelling*. Edward Elgar Publishing. 2014

Ho MP, Gonzalez JM et al. Incorporating patient-preference evidence into regulatory decision making. *Surgical endoscopy*. 2015, 29 (10):2984-93.

Ho MP et al., A Framework for Incorporating Patient Preferences Regarding Benefits and Risks into Regulatory Assessment of Medical Technologies. *Value Health*. Sep-Oct 2016;19(6):746-750.



Hockley K., et al. Study Protocol: Eliciting Patient Preferences on the Benefits and Risks of Treatments for Relapsing Remitting Multiple Sclerosis IMI 2014

Hollin IL et al. Patient-centered benefit-risk assessment in Duchenne Muscular Dystrophy. *Muscle & nerve* 2016

Hollin IL et al., Collecting patient preference information using a Clinical Data Research Network: demonstrating feasibility with idiopathic pulmonary fibrosis. *Patient Prefer Adherence*. 2019 May 16;13: 795-804.

Hollin IL et al., Do Patient Preferences Align With Value Frameworks? A Discrete-Choice Experiment of Patients With Breast Cancer. *MDM Policy Pract*. 2020 Jun 15;5(1)

Howitt D et al., *Introduction to Qualitative Research Methods in Psychology*, 3rd Edition. 2016.

Hsu J. C., et al. "Net clinical benefit of oral anticoagulants: a multiple criteria decision analysis." *PLoS One* 2015, 10(4)

Huls S.P.I. et al., What Is Next for Patient Preferences in Health Technology Assessment? A Systematic Review of the Challenges. *Value Health*. 2019 Nov;22(11):1318-1328.

Hummel MJM, et al. Using the analytic hierarchy process to elicit patient preferences: Prioritizing multiple outcome measures of antidepressant drug treatment. *Patient*. 2012; 5(4):225-37.

IQWiG, Choice-based Conjoint Analysis – pilot project to identify, weight, and prioritize multiple attributes in the indication “hepatitis C”. Institute for Quality and Efficiency in Health Care (IQWiG); 2014. Report No.: GA10-03.

Imaeda A, Bender D and Fraenkel L., What is most important to patients when deciding about colorectal screening? *J Gen Intern Med*. 2010 Jul;25(7):688-93.

Janssen EM et al., A Framework for Instrument Development of a Choice Experiment: An Application to Type 2 Diabetes. *Patient*. 2016 Oct;9(5):465-79.

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 Review of Pharmacoeconomics & Outcomes Research* (2017), 17 (6), 531-542.

Janssen EM, Hauber AB and Bridges JFP Conducting a Discrete-Choice Experiment Study Following Recommendations for Good Research Practices: An Application for Eliciting Patient Preferences for Diabetes Treatments. *Value in Health* 2018, 21, 59-68.

Janssen Research & Development 2019, Advisory Committee Briefing Document Esketamine Nasal Spray for Patients with Treatment-resistant Depression, viewed 25 July 2019

Janssens R, Russo S, van Overbeeke E, et al. ; Patient Preferences in the Medical Product Life Cycle: What do Stakeholders Think? Semi-Structured Qualitative Interviews in Europe and the USA; *The Patient - Patient-Centered Outcomes Research* 2019, 12(5), pp 513–526

Janssens R., Huys I., et al., Opportunities and challenges for the inclusion of patient preferences in the medical product life cycle: a systematic review. *BMC Med Inform Decis Mak*. 2019 Oct 4;19(1):189.

Jones LG, Hawkins GE, Brown SD. Using best–worst scaling to improve psychological service delivery: An innovative tool for psychologists in organized care settings. *Psychological Services*. 2015;12(1):20-27.

Johnson FR. Et al., Eliciting stated preferences for health-technology adoption criteria using paired comparisons and recommendation judgments. *Value Health*. Sep-Oct 2006;9(5):303-11.

Johnson FR., et al., Multiple sclerosis patients—benefit-risk preferences: Serious adverse event risks versus treatment efficacy. *J Neurol*. 2009 Apr;256(4):554-62.

Johnson FR, Lancsar E, Marshall D, et al. Constructing experimental designs for discrete-choice experiments: Report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. *Value Health* 2013, 16:3–13.

Johnson FR, Zhou M. Patient Preferences in Regulatory Benefit-Risk Assessments: A US Perspective. *Value in Health* 2016, 19(6):741-5.

Johnson FR, Beusterien K, Ozdemir S, Wilson L. Giving Patients a Meaningful Voice in United States Regulatory Decision Making: The Role for Health Preference Research. *Patient*. 2017;10(4):523-6.

Johnson FR, DiSantostefano RL, Yang JC, Reed SD, Streffer J, Levitan B, Something Is Better Than Nothing: The Value of Active Intervention in Stated Preferences for Treatments to Delay Onset of Alzheimer's Disease Symptoms, *Value Health*. 2019 Sep;22(9):1063-1069

Johnson, FR, Yang, JC, Reed, SD The Internal Validity of Discrete Choice Experiment Data: A Testing Tool for Quantitative Assessments *Value in Health* 2019, 22, 157-160.

Jonker M. et al., 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 Mar;28(3):350-363.

Kahneman A and Tversky D. The Framing of Decisions and the Psychology of Choice. *Science*, 30 January 1981, Vol 211:453-458.

Kaltoft MK et al., Dual Purpose, Dual Audience: MCDA-Based Tools Can Simultaneously Support Personal Health Decisions and Educate Persons and Clinicians. *Stud Health Technol Inform*. 2018;255:257-261.

KCE, Centre Federal d'Expertise des Soins de Sante, Position of KCE on patient involvement in health care policy research. Short report. 2019

Keeney RL and Raiffa H. Decisions with Multiple Objectives - Preferences and Value Trade-Offs. August 1979 [IEEE Transactions on Systems Man and Cybernetics](#) 9(7):403 – 403.

Keeney RL et al., Creativity in decision making with value-focused thinking. January 1994 [Sloan Management Review](#) 35(4):33-41

Keeney RL et al., Common Mistakes in Making Value Trade-Offs. December 2002; [Operations Research](#) 50(6):935-945.

Knox S. et al., What's Good and Bad About Contraceptive Products? A Best-Worst Attribute Experiment Comparing the Values of Women Consumers and GPs. October 2012 [PharmacoEconomics](#) 30(12)

- Lacey A. and Luff D., Qualitative data analysis. The NIHR Research Design Service for Yorkshire and the Humber. 2007.
- Lahdelma R and Salminen P. SMAA-2: Stochastic Multicriteria Acceptability Analysis for Group Decision Making. June 2001 [Operations Research](#) 49(3)
- Lancsar E and Louviere J. Conducting Discrete Choice Experiments to Inform Healthcare Decision Making. A User's Guide. *Pharmacoeconomics*, 2018, 26(8), 661-677.
- Lancsar, E, Fiebig, DG and Hole, AR Discrete Choice Experiments: A Guide to Model Specification, Estimation and Software. *Pharmacoeconomics* 2017, 35(7), 697-716.
- Leanne G. Jones, Guy E. Hawkins, Scott D. Brown. Using best–worst scaling to improve psychological service delivery: An innovative tool for psychologists in organized care settings. *Psychological Services*. 2015;12(1):20-27
- Leong, W., Hensher, D.A. 2012 Embedding decision heuristics in discrete choice models: a review. *Transport Reviews: A Transnational Transdisciplinary Journal* 32(3), 313–331.
- Leitjen FRM et al., The SELFIE framework for integrated care for multi-morbidity: Development and description. *Health Policy*. 2018 Jan;122(1):12-22.
- Levitan B, Hauber AB, Damiano MG, et al. The Ball is in Your Court: Agenda for Research to Advance the Science of Patient Preferences in the Regulatory Review of Medical Devices in the United States. *Patient* (2017 10(5): 531-536.
- Louviere J. et al., Stated Choice Methods: Analysis and Applications. Combining sources of preference data. 2000:227-251.
- Louviere, J. J., T. N. Flynn, and A. A. J. Marley. Best–Worst Scaling: Theory, Methods and Applications. 2015. *Best–Worst Scaling: Theory, Methods and Applications*. Cambridge, UK: Cambridge University Press.
- Lowe MM, Blaser DA, Cone L, Arcona S, Ko J, Sasane R, et al. Increasing Patient Involvement in Drug Development. *Value Health* 2016, 19(6):869-78
- Ma VY, Chan L, Carruthers KJ. Incidence, prevalence, costs, and impact on disability of common conditions requiring rehabilitation in the United States: stroke, spinal cord injury, traumatic brain injury, multiple sclerosis, osteoarthritis, rheumatoid arthritis, limb loss, and back pain. *Arch Phys Med Rehabil*. 2014 May;95(5):986-995.e1.
- Mansfield C et al., Patient preferences for treatment of metastatic melanoma. *Future Oncol*. 2019 Apr;15(11):1255-1268
- Marsh K et al., Assessing the Value of Healthcare Interventions Using Multi-Criteria Decision Analysis: A Review of the Literature. [Pharmacoeconomics](#) volume 32, pages345–365(2014)
- Marsh K et al., Incorporating Patient Preferences into Product Development and Value Communication: Why, When and How? *The Evidence Forum*. 2016:38-41.
- Marsh K, Caro JJ, Hamed A, Zaiser E. Amplifying Each Patient's Voice: A Systematic Review of Multi-criteria Decision Analyses Involving Patients. *Applied Health Economics and Health Policy*. 2017;15(2):155-62.

Marsh K et al., Incorporating Preferences and Priorities into MCDA: Selecting an Appropriate Scoring and Weighting Technique. [Multi-Criteria Decision Analysis to Support Healthcare Decisions](#) pp 47-66. 2017.

Marsh K, Caro JJ, Zaiser E, Heywood J, Hamed A. Patient-centered decision making: Lessons from multi-criteria decision analysis for quantifying patient preferences. *International journal of technology assessment in health care*. 2018, 34(1):105-10.

Marsh K et al., Patient-centered benefit-risk analysis of transcatheter aortic valve replacement. *F1000Res*. 2019 Apr 8;8:394.

Marshall D et al. Conjoint analysis applications in health – how are studies being designed and reported? An update on current practice in the published literature between 2005 and 2008. *Patient: Patient-Centered Outcomes Research*, 2010, 3(4), 249-256.

Marshall T. et al., Patient Preferences for Device-Aided Treatments Indicated for Advanced Parkinson Disease. *Value Health*. 2017 Dec;20(10):1383-1393

Marti J. et al., A best-worst scaling survey of adolescents' level of concern for health and non-health consequences of smoking. *Soc Sci Med*. 2012;75(1):87–97.

McFadden D. Quantitative methods for analyzing travel behaviour on individuals: some recent developments. In Hensher D, Stopher P. *Behavioral Travel Modelling* (1978) London: Croom Helm.

McFadden D and Train K. Mixed MNL models for discrete response. *Journal of applied econometrics*. 2000 Dec;15(5):447-470.

Medical Device Innovation Consortium. Medical Device Innovation Consortium (MDIC) Patient Centered Benefit-Risk Project Report: A Framework for Incorporating Information on Patient Preferences Regarding Benefit and Risk into Regulatory Assessments of New Medical Technology. Available at [https://mdic.org/wp-content/uploads/2018/05/MDIC\\_PCBR\\_Framework\\_Web.pdf](https://mdic.org/wp-content/uploads/2018/05/MDIC_PCBR_Framework_Web.pdf). Accessed February 9, 2021.

Medical Device Innovation Consortium (MDIC), MDICx Series. NxStage. Medical Device Innovation Consortium 2016

Medical Device Innovation Consortium (MDIC). Annual Report 2019. Available at 2019 Annual Report - MDIC. Accessed February 23, 2021.

Menon D et al. Involving patients in reducing decision uncertainties around orphan and ultra-orphan drugs: a rare opportunity? *Patient*, 2014, 8, pp. 29-39

Meyfroidt S, Hulscher M, De Cock D, Van der Elst K, Joly J, Westhovens R, et al. A maximum difference scaling survey of barriers to intensive combination treatment strategies with glucocorticoids in early rheumatoid arthritis. *Clin Rheumatol*. 2015;34(5):861–9.

Mohamed A et al., Patient preferences and linear scoring rules for patient reported outcomes. *Patient*. 2010 Dec 1;3(4):217-27.

Mohamed A et al. Physicians' stated trade-off preferences for chronic hepatitis B treatment outcomes in Germany, France, Spain, Turkey and Italy. *Eur J Gastroenterol Hepatol* 2012 24:419-426

Mohammadi T et al. Testing the External Validity of a Discrete Choice Experiment Method: An Application to Latent Tuberculosis Infection Treatment, *Value in Health* Volume 20, Issue 7, July–August 2017, Pages 969-975

Mol PG, Arnardottir AH et al. Understanding drug preferences, different perspectives. *British journal of clinical pharmacology*. 2015; 79(6):978-87.

Mott DJ. Whose preferences should be elicited for use in health-care decision-making? A case study using anticoagulant therapy. *Expert Rev Pharmacoecon Outcomes Res*. 2016;16(1):33-9.

Mott DJ. Incorporating Quantitative Patient Preference Data into Healthcare Decision Making Processes: Is HTA Falling Behind? *Patient*. 2018 Jun;11(3):249-252

Mühlbacher AC. Patient-centric HTA: different strokes for different folks. Expert review of pharmacoeconomics and outcomes research. 2015 Vol 15, Issue 4.

a) Mühlbacher AC and Kaczynski A. Making Good Decisions in Healthcare with Multi-Criteria Decision Analysis: The Use, Current Research and Future Development of MCDA. *Applied Health Economics and Health Policy*. 2016;14(1):29-40.

b) Mühlbacher AC, Johnson FR. Choice Experiments to Quantify Preferences for Health and Healthcare: State of the Practice. *Appl Health Econ Health Policy*. 2016 Jun;14(3):253-66

c) Mühlbacher AC, Juhnke C, Beyer AR, Garner S. Patient-Focused Benefit-Risk Analysis to Inform Regulatory Decisions: The European Union Perspective. *Value in Health*. 2016;19(6):734-40.

Mühlbacher AC, Sadler A. The Probabilistic Efficiency Frontier: A Framework for Cost-Effectiveness Analysis in Germany Put into Practice for Hepatitis C Treatment Options. *Value Health*, 2017;20(2):266-272

Mühlbacher et al. Preferences for antiviral therapy of chronic hepatitis C: a discrete choice experiment. *Eur J Health Econ*, 2017;18(2):155-165. Mühlbacher AC, Johnson FR. Giving Patients a Meaningful Voice in European Health Technology Assessments: The Role of Health Preference Research. *The patient*. 2017;10(4):527.

Mühlbacher AC & Marsh K et al., PP177 Health Preference Research In Europe: A Review Of Its Use. *The technology Assessment in Health Care*. 2019;35(1):1-1.

Myeloma UK project report June 2019: <https://www.myeloma.org.uk/wp-content/uploads/2019/07/NICE-Patient-Preferences-Report.pdf>

Mussen F et al., Retracted: A quantitative approach to benefit- risk assessment of medicines — part 2: the practical application of a new model. *Pharmacoepidemiol Drug Saf*. 2007 Oct;16(10):1180

NICE, NICE provides first scientific advice on patient preference study design. National Institut for Health and Care Excellence (NICE). 2019.

Nunley S. et al., A hospital-based study on caregiver preferences on acute seizure rescue medications in pediatric patients with epilepsy: Intranasal midazolam versus rectal diazepam. *Epilepsy Behav*. 2019 Mar;92:53-56.

Nutt DJ et al., Drug harms in the UK: A multi-criterion decision analysis. November 2010 [The Lancet](#) 376

Nutt DJ et al., Estimating the Harms of Nicotine-Containing Products Using the MCDA Approach  
Eur Addict Res 2014; 20:218–225

Oremus M. et al., Caregivers' willingness-to-pay for Alzheimer's disease medications in Canada.  
Dementia (London). 2015 Jan; 14(1):63-79

Ozdemir S et al., Hypothetical bias, cheap talk, and stated willingness to pay for health care. Journal  
of Health Economics, 2009; 28(4), 894– 901.

Parsons S, Starling B et al. What do pharmaceutical industry professionals in Europe believe about  
involving patients and the public in research and development of medicines? A qualitative interview  
study. BMJ Open. 2016; 6(1):e008928.

Patalano F et al., Gathering structured patient insight to drive the PRO strategy in COPD - Patient-  
centric drug development from theory to practice. Adv Ther. 2020 Jan; 37(1): 17-26.

Peay H. L. et al. (2019) 'Gene therapy as a potential therapeutic option for Duchenne muscular  
dystrophy: A qualitative preference study of patients and parents', PLoS One. 2019 May  
1; 14(5):e0213649.

Percy M.E. and Llewellyn-Thomas H. Assessing preferences about the DNR order: does it depend on  
how you ask? Med Decis Making. Jul-Sep 1995; 15(3):209-16. Pettus JH, Kushner JA et al. Adjunct  
Therapy in Type 1 Diabetes: A Survey to Uncover Unmet Needs and Patient Preferences Beyond  
HbA1c Measures, Diabetes Technol Ther. 2019 Jun; 21(6):336-343,

Pignatti F, Ashby D T et al. Structured Frameworks to Increase the Transparency of the Assessment  
of Benefits and Risks of Medicines: Current Status and Possible Future Directions. Clin Pharmacol  
Ther. 2015 Nov; 98(5):522-33

Pocock S.J. et al. The win ratio: a new approach to the analysis of composite endpoints in clinical  
trials based on clinical priorities." Eur Heart J. 2012 33(2): 176-182.

Pope C, Ziebland S et al. Qualitative research in health care: Analysing qualitative data. BMJ  
2000; 320:114–6

Postmus D, Mavris M et al. Incorporating patient preferences into drug development and regulatory  
decision making: Results from a quantitative pilot study with cancer patients, carers, and regulators.  
Clin Pharmacol Ther. 2016; 99(5):548-54.

Postmus D et al., Individual trade-offs between possible benefits and risks of cancer treatments:  
Results from a stated study with patients with a MM. Oncologist. 2018 Jan; 23(1): 44-51.

Potoglou D et al., Best-worst scaling vs. discrete choice experiments: an empirical comparison using  
social care data. Soc Sci Med. 2011 May; 72(10):1717-27.

Poulos C et al., Patient and physician preferences for multiple sclerosis treatments in Germany: A  
discrete-choice experiment study. Mult Scler J Exp Transl Clin. 2020 Mar 8; 6(1):2055217320910778

Rakotonarivo O.S. et al., A systematic review of the reliability and validity of discrete choice  
experiments in valuing non-market environmental goods. J Environ Manage. 2016 Dec 1; 183:98-  
109.

Raphael M.J. et al., The Value of Progression-Free Survival as a Treatment End Point Among Patients With Advanced Cancer - A Systematic Review and Qualitative Assessment of the Literature. *JAMA Oncol.* 2019 Sep 26.

Ratcliffe J. et al., An assessment of the construct validity of the CHU9D in the Australian adolescent general population. *Qual Life Res.* 2012 May; 21(4): 717-25.

Ratcliffe J. et al., Valuing the Child Health Utility 9D: Using profile case best worst scaling methods to develop a new adolescent specific scoring algorithm. *Soc Sci Med.* 2016 May; 157: 48-59.

Robinson A, Spencer A and Moffatt P., A framework for estimating health state utility values within a discrete choice experiment: modeling risky choices. *Med Decis Making.* 2015 Apr; 35(3): 341-50.

Rodríguez-Míguez E., Pinto-Prades J.L. and Mosquera-Nogueira J. Eliciting Health State Utilities Using Paired-Gamble Methods: The Role of the Starting Point. *Value Health.* 2019 Apr; 22(4): 446-452.

Ross M et al., A Best-Worst Scaling Experiment to Prioritize Caregiver Concerns About Attention-Deficit/Hyperactivity Disorder (ADHD) Medication for Children. [Psychiatr Serv. 2015 Feb 1; 66\(2\): 208-211.](#)

Rummel M et al, Prefmab: Final Analysis of Patient Preference for Subcutaneous Versus Intravenous Rituximab in Previously Untreated CD20+ Diffuse Large B-Cell Lymphoma and Follicular Lymphoma, *Blood* (2015) 126 (23): 3972

Rummel M, Kim TM, Aversa F, Brugger W, Capochiani E, Plenteda C, et al. Preference for subcutaneous or intravenous administration of rituximab among patients with untreated CD20+ diffuse large B-cell lymphoma or follicular lymphoma: results from a prospective, randomized, open-label, crossover study. *Ann Oncol.* 2017 Apr 1; 28(4): 836-842

Russo S et al. Understanding Patients' Preferences: A Systematic Review of Psychological Instruments Used in Patients' Preference and Decision Studies. *Value in Health.* 2019; 22(4): 491-501

Ryan M et al. Sensitivity of willingness to pay estimates to the level of attributes in discrete choice experiments. *Scottish Journal of Political Economy.* (2003); Vol.47, Issue 5.

Ryan M., Gerard K. et al., Using discrete choice experiments to value health care programmes: current practice and future research reflections. *Appl Health Econ Health Policy.* 2003; 2(1): 55-64.

Ryan M. et al., Rationalising the irrational: a think aloud study of discrete choice experiment responses. *Health Econ.* 2009 Mar; 18(3): 321-36.

Ryan M, Krucien N and Hermens F. The eyes have it: Using eye tracking to inform information processing strategies in multi-attributes choices. *Health Econ.* 2018 Apr; 27(4): 709-721.

Selig W. K. D. et al., Key considerations for developing & integrating patient perspectives in drug development: examination of the duchenne case study. *Biotechnol. Innovation Organ. Parent Project Muscular Dystrophy.* (2016)

Severin F et al., Eliciting preferences for priority setting in genetic testing: a pilot study comparing best-worst scaling and discrete-choice experiments. *Eur J Hum Genet.* 2013 Nov; 21(11): 1202-8

- Silverman S, Calderon A, Kaw K, et al. Patient weighting of osteoporosis medication attributes across racial and ethnic groups: a study of osteoporosis medication preferences using conjoint analysis. *Osteoporos Int.* 2013;24(7):2067–77.
- Singer DE et al. The net clinical benefit of warfarin anticoagulation in atrial fibrillation. *Ann Intern Med* (2009) 151(5): 297-305
- Shields GE et al., Utilising Patient and Public Involvement in Stated Preference Research in Health: Learning from the Existing Literature and a Case Study. *Patient.* 2020 Aug 4.
- Smith M et al., Patient Engagement at a Tipping Point—The Need for Cultural Change Across Patient, Sponsor, and Regulator Stakeholders: Insights From the DIA Conference, “Patient Engagement in Benefit Risk Assessment Throughout the Life Cycle of Medical Products”. *Ther Innov Regul Sci.* 2016 Sep;50(5):546-553.
- Soekhai V, De Bekker-Grob EW, Ellis AR, Vass CM, Discrete Choice Experiments in Health Economics: Past, Present and Future. *Pharmacoeconomics.* 2019 Feb;37(2):201-2.
- Soekhai V. et al., Methods for exploring and eliciting patient preferences in the medical product life cycle: a literature review. *Drug Discov Today.* 2019 Jul;24(7):1324-1331
- SriBhashyam S. et al., A benefit-risk analysis of pegvaliase for the treatment of phenylketonuria: A study of patients' preferences. *Mol Genet Metab Rep.* 2019 Aug 30;21:100507
- Stamuli E et al., Identifying the primary outcome for a randomised controlled trial in rheumatoid arthritis: the role of a discrete choice experiment. *J Foot Ankle Res.* 2017 Dec 15;10:57.
- Steures P et al., Patients' preferences in deciding between intrauterine insemination and expectant management. *Hum Reprod.* 2005 Mar;20(3):752-5
- Stewart K et al., Simplified approaches for multicriteria decision making under uncertainty. December 1995 [Journal of Multi-Criteria Decision Analysis](#) 4(4):246 - 258
- Stewart K et al., Preference for pharmaceutical formulation and treatment process attributes. *Patient Preference and Adherence.* 2016;10:1385-99.
- Sung L et al., Inpatient versus outpatient management of low-risk pediatric febrile neutropenia: measuring parents' and healthcare professionals' preferences. *J Clin Oncol.* 2004 Oct 1;22(19):3922-9.
- Szeinbach SL, Barnes JH, McGhan WF, et al. Using conjoint analysis to evaluate health state preferences. *Drug Inf J.* 1999;33(3):849–58.
- Tervonen T et al., A Survey on Stochastic Multicriteria Acceptability Analysis Methods. [Journal of Multi-Criteria Decision Analysis](#). 2008. 15(1- 2):1 - 14
- Tervonen T et al., A stochastic multicriteria model for evidence- based decision making in drug benefit- risk analysis. *Stat Med.* 2011 May 30;30(12):1419-28.
- Tervonen T et al., Applying Multiple Criteria Decision Analysis to Comparative Benefit-Risk Assessment: Choosing among Statins in Primary Prevention. *Medical Decision Making.* 2015. 35(7):859-871.



Tervonen T et al., MCDA swing weighting and discrete choice experiments for elicitation of patient benefit-risk preferences: a critical assessment. *Pharmacoepidemiol Drug Saf.* 2017 Dec;26(12):1483-1491.

Tervonen T et al., Quantifying Preferences in Drug Benefit-Risk Decisions. *Clin Pharmacol Ther.* 2019 Nov;106(5):955-959.

Tervonen T et al., From Individual to Population Preferences: Comparison of Discrete Choice and Dirichlet Models for Treatment Benefit-Risk Tradeoffs. *Med Decis Making.* 2019 Oct;39(7):879-885.

Thokala P et al. Multiple Criteria Decision Analysis for Health Care Decision Making-An Introduction: Report 1 of the ISPOR MCDA Emerging Good Practices Task Force. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2016; 19:1-13.

Torbica A et al., Eliciting preferences to inform patient-centred policies: the case of psoriasis. *Pharmacoeconomics.* 2014 Feb;32(2):209-23.

Train K. and Sonnier G., Mixed logit with bounded distributions of correlated partworths. *Applications of Simulation Methods in Environmental and Resource Economics.* 2005. 117-134

Van Amsterdam J et al., European rating of drug harms. *J Psychopharmacol.* 2015 Jun;29(6):655-60.

Van Houtven G, Johnson FR, Kilambi V, et al. Eliciting benefit-risk preferences and probability-weighted utility using choice-format conjoint analysis. *Medical Decision Making* 2011. 31 (3), 469-480

Van Houtven CH, Voils CI, Weinberger M. An organizing framework for informal caregiver interventions: detailing caregiving activities and caregiver and care recipient outcomes to optimize evaluation efforts. 2011 Nov 22;11:77

Van Overbeeke E, et al. Factors and situations influencing the value of patient preference studies along the medical product lifecycle: a literature review. *Drug discovery today.* 2019;24(1):57-68

Van Overbeeke E et al., Patient perspectives regarding gene therapy in haemophilia: Interviews from the PAVING study. *Haemophilia.* 2021 Jan;27(1):129-136.

Van Til, J.A., Ijzerman M.J. Why should regulators consider using patient preferences in benefit-risk assessment? *Pharmacoeconomics*, (2014) 32, pp. 1-4

Van Valkenhoef G et al., Multicriteria benefit-risk assessment using network meta-analysis. *J Clin Epidemiol.* 2012 Apr;65(4):394-403.

Van Valkenhoef G et al., Evidence synthesis assumes additivity on the scale of measurement: response to "Rank reversal in indirect comparisons" by Norton et al. *Value Health.* Mar-Apr 2013;16(2):449-51.

Vass, C. M et al. Using discrete choice experiments to inform the benefit-risk assessment of medicines: are we ready yet? *Pharmacoeconomics* (2017) 35(9) 859–866.

Vermersch P et al., Benefit-risk Assessment of Cladribine Using Multi-criteria Decision Analysis (MCDA) for Patients With Relapsing-remitting Multiple Sclerosis. *Clin Ther.* 2019 Feb;41(2):249-260

- Versteegh M.M. and Brouwer W.B.F., Patient and general public preferences for health states: A call to reconsider current guidelines. *Social Science & Medicine*. 2016;66-74
- Wagner F et al., Applying Reflective Multicriteria Decision Analysis (MCDA) to Patient–Clinician Shared Decision-Making on the Management of Gastroenteropancreatic Neuroendocrine Tumors (GEP-NET) in the Spanish Context. *Adv Ther*. 2018; 35(8): 1215–1231.
- Weber M. and Borcherding K. Behavioral influences on weight judgments in multiattribute decision making. *European Journal of Operational Research*, 1993, vol. 67, issue 1, 1-12
- Whitty JA and Oliveira Goncalves AS. A Systematic Review Comparing the Acceptability, Validity and Concordance of Discrete Choice Experiments and Best-Worst Scaling for Eliciting Preferences in Healthcare. *Patient*. 2018 Jun;11(3):301-317.
- Whichello C et al., Appraising patient preference methods for decision-making in the medical product lifecycle: an empirical comparison. *BMC Med Inform Decis Mak*. 2020 Jun 19;20(1):114.
- Wild D et al. ISPOR Task Force for Translation and Cultural Adaptation. *Value Health*. 2005 Mar-Apr;8(2):94-104
- Yang, J-C, Johnson, FR, Kilambi, V et al. Sample size and utility-difference precision in discrete-choice experiments: A meta-simulation approach. *The Journal of Choice Modelling* 2015, 16, 50-57
- Yu T et al., Outcome Preferences in Patients With Noninfectious Uveitis: Results of a Best–Worst Scaling Study. *Invest Ophthalmol Vis Sci*. 2015 Oct; 56(11): 6864–6872.
- Yuan Z et al., Yuan Z, Levitan B, Burton P, Poulos C, Hauber AB, Berlin JA. Relative importance of benefits and risks associated with antithrombotic therapies for acute coronary syndrome: patient and physician perspectives. *Curr Med Res Opin*. 2014 2014;30(9):1733-41.

## 8 Annexes

### 8.1 Annex I: Abstracts of all case studies

#### 8.1.1 Main case studies in PREFER

<b>Therapeutic Area</b>	<b>ON-GOING CASE STUDY SUMMARY: UPDATED AS OF JANUARY 31, 2021</b> <b>Rheumatoid Arthritis (RA)</b>
<b>Core Case Study, Academic Study, Industry Study</b>	Core Case Study
<b>Current status</b>	This case study will be finalized in April 2021
<b>Academic Lead, Industry Lead, Methods Lead, Clinical Lead</b>	Karim Raza, University of Birmingham, Case study lead Marie Falahee, University of Birmingham, Chief Investigator Gwenda Simons, University of Birmingham, Research fellow Larissa Valor-Méndez, Case study lead Germany Matthias Englbrecht, Freelance Data Scientist Jorien Veldwijk, Erasmus University, Academic co-lead and Methods Lead Rachael DiSantostefano, Janssen R&D, Industry co-lead
<b>Title of the Study</b>	Treatment preferences for preventive interventions for rheumatoid arthritis
<b>Background/ Study Rationale</b>	There is increasing research focus on intervention for RA at the earliest stages of disease development, including treatment to prevent RA in at-risk groups. Novel cellular therapies are in development, and the effectiveness of existing immunomodulatory agents to prevent RA in those at risk is under investigation. Quantitative evidence of likely uptake of preventive treatments, and preferences for benefits and risks of such treatments is limited.
<b>MPLC Decision-making</b>	Main stakeholder (industry, regulatory, HTA): Industry MPLC decision-point of interest: Post-approval for approved medicines. Consideration for new indications relating to an RA delay or prevention.

<p><b>Clinical Research Questions</b></p>	<ul style="list-style-type: none"> <li>• To establish the preferences of ‘at risk’ individuals and the general public about preventive treatments for RA.</li> <li>• To establish the maximum acceptable risk (MAR) of preventative treatments for RA. (e.g., what is the maximum acceptable risk for a reduction in the risk of developing RA?)</li> </ul>
<p><b>Methodological Research Questions</b></p>	<ul style="list-style-type: none"> <li>• To compare the application and results of two preference methods, a Discrete-Choice Experiments (DCE) and the Probabilistic Threshold Techniques (PTT)</li> </ul>
	<ul style="list-style-type: none"> <li>• To examine how psychological instruments might complement a preference study, regarding preference formation or preference heterogeneity</li> </ul>
	<ul style="list-style-type: none"> <li>• To examine heterogeneity of preferences and what might explain heterogeneity</li> </ul>
<p><b>Study Design</b></p>	<p>Overall: This study was divided into two phases: (i) qualitative study to identify attributes and levels; (ii) quantitative survey including a discrete choice experiment (DCE) task and probabilistic threshold technique (PTT).</p> <p>Qualitative Research: Attribute selection and presentation for the quantitative survey was informed by qualitative research in the form of focus groups/ interviews and ranking surveys, a systematic literature review, input from patient research partners and expert opinion. The actual survey design was also informed by patient research partners. The quantitative survey was pre-tested during qualitative interviews and revised.</p> <p>Quantitative Research: During survey design, the DCE attributes and levels were developed first. Then, the PTT tasks were adapted from the DCE, to ensure consistency. (The order in which respondents answered the DCE and PTT was randomly assigned).</p> <p>The quantitative full survey contained six sections:</p> <ul style="list-style-type: none"> <li>• Screening questions</li> <li>• Demographic questions (age, gender, employment status, education) questions related to family history of RA, smoking status.</li> <li>• Warmup questions followed by DCE or PPT choices (half of respondents received PPT then DCE; half with DCE then PPT in each sample); following each preference exercise, we asked questions about the ease of completing the exercise.</li> <li>• Literacy and numeracy assessment</li> <li>• Warmup questions followed by DCE or PPT choices</li> <li>• Question to assess perceived risk of RA; Belief about medicine and Illness Perception</li> </ul>

	<p>To inform the final Bayesian D-efficient experimental design and optimize statistical efficiency for the DCE, a survey pilot was conducted using the first 100 respondents of the general public in the UK. Prior information on the importance of the attributes was based on previous literature and best guesses, and outcomes of initial analysis (conditional logit) of pilot data for the main survey. For both the pilot and the final design, a total of 60 unique choice tasks were generated, which were divided over 4 blocks. Respondents were randomized to one of the four blocks. Dominant alternatives and some attribute and level combinations (e.g., iv drip daily), were excluded from the design. Additionally, interactions between effectiveness and change of a serious infection and between effectiveness and change of serious side effects were included in the design.</p> <p>Educational materials: Written explanation of RA, risk of RA and treatment attributes. Before respondents started with the DCE or PTT they saw an example choice task including a walkthrough of full question (showing the attributes, levels, alternatives, pop-up boxes for more information and an explanation on how to indicate their preferred choice).</p> <p>Psychosocial constructs: Single Item Literacy Screener (SILS), Subjective Numeracy Scale 3-item version (SNS- 3), Brief Illness Perception Questionnaire (B-IPQ), and the Beliefs about Medicines Questionnaire-General (BMQ-General).</p>
<p><b>Study Population</b></p>	<p>The samples included respondents from the general population and first-degree relatives (FDRs) across the UK, Germany and Romania. Within each country, general population samples were age- and gender- matched to correspond to a sample of FDRs from a prior study in the UK.</p> <ol style="list-style-type: none"> <li>1. aiming for 250 adults who are FDRs of an individual with RA (UK) [recruitment ongoing through February 2021]</li> <li>2. aiming for 50 adults who are FDRs of an individual with RA (Germany) [recruitment ongoing through February 2021]</li> <li>3. 1000 adults from the general population (UK)</li> <li>4. 1000 adults from the general population (Germany)</li> <li>5. 1000 adults from the general population (Romania)</li> </ol>
<p><b>Exposure &amp; Outcome</b></p>	<p>Exposure: Potential preventive treatments for RA that may lower the chance of developing RA in the upcoming two years.</p> <p>Outcomes: [i] relative preference weights for levels of treatment attributes; [ii] estimated risk equivalents (maximum acceptable risk (MAR) and minimal accepted benefit (MAB)) for changes in treatment attributes; [iii] potential treatment shares.</p>
<p><b>Study Setting</b></p>	<p>In the UK and Germany, FDRs were recruited to participate in the study via a parent or full sibling, with confirmed, clinical diagnosis of RA. The family member with RA was mailed information leaflets about the study including materials to pass along to their FDRs for consideration; or where we had email addresses of family members, they were contacted directly.</p> <p>In each country, the general populations without an RA diagnosis were recruited through established study panels.</p>

	<p>Respondents were told to assume a 60% chance of developing RA in the next 2 years without treatment as context for the subsequent preferences exercises, where treatments would lower the risk of an RA diagnosis.</p>
<p><b>Statistical Methods</b></p>	<p>Using a discrete choice experiment with a Bayesian D-efficient design, participants were offered a series of 15 choices between no treatment and two unlabeled hypothetical treatments to lower risk of RA development. Treatments were defined by six attributes with varying levels including: reduced chance of developing RA, how and how often treatment is taken, chances of mild and serious side effects, and chance of serious infection. One choice task with fixed levels described treatments representative of those under investigation for RA prevention (abatacept, hydroxychloroquine, atorvastatin and tolerogenic cell-based therapy).</p> <p>In the PTT, participants were offered a series of choices between no treatment and an unlabeled hypothetical treatments to lower risk of RA development. The no treatment option showed the base (60%) chance of developing RA but no risk of adverse events (mild side effects, serious infection, serious side effects). The treatment option showed a 20% chance of developing RA (risk reduction) and increased risks of adverse events. Participants were asked to indicate whether they would choose no treatment or treatment. In subsequent tasks the risk of one of the attributes was varied (increased or decreased) until participants changed their treatment choice. This exercise was repeated for all three types of adverse events.</p> <p>For the DCE, a logit-based analysis strategy was conducted to estimate preferences for attributes of RA prevention therapies including panel random parameter logit (RPL) models and latent class analyses (LCA). Final decisions on the modelling procedures for each research objective were based on model fit and clinical interpretation, with different models to address the different research questions in this case study.</p> <p>For the PTT, data were analysed using imputation and interval regression. The MAR values for benefits and risks were calculated and relative importance of attributes was determined, which allowed for comparison between DCE and PTT methods.</p> <p>Heterogeneity of preferences and the impact of participant characteristics (e.g. demographics, RA knowledge, psychological instruments) was investigated by applying appropriate statistical models including LCA for the DCE and/or subgroup analyses for the DCE and regression models including covariates for the PTT methods.</p> <p>For the DCE only, the potential treatment shares of currently existing preventive treatment were calculated. These results were examined against direct elicitation questions about whether or not participants would use those preventive treatments.</p> <p>All results described above were formally compared between the three countries and between FDRs [when available] and the general population where possible/appropriate.</p>

Results	PREFER Research Questions	How was it addressed in this preference study	What was found
	Comparison of Methods: DCE vs. PTT	Qualitative and quantitative comparison of results: MAR, relative importance of the attributes, proportion of respondents avoiding risks, and patient-reported ease of understanding/completion	<ul style="list-style-type: none"> <li>- In the general population, maximum acceptable risk values were numerically similar across methods for the values evaluated.</li> <li>- In the general population, the majority of the respondents found the DCE and PTT were easy or very easy to understand and answer. However, the DCE was found to be significantly easier to understand and answer than the PTT (<math>p &lt; 0.05</math>). Respondents who received the DCE first found all tasks (both DCE and PTT) easier to understand and compared to respondents who received the PTT first (<math>p &lt; 0.05</math>).</li> </ul>
	The role of psychological instruments in preference formation or preference heterogeneity	TBD.	Analysis to be completed by the end of April 2021.
	Factors that may explain heterogeneity of preferences	TBD.	Analysis to be completed by the end of April 2021.
<b>Conclusions</b>	<p><b>Clinical:</b> Effective preventive treatments for RA were acceptable to members of the general population told to assume up a 60% chance of developing RA in the next two years. In the general population, the relative importance of treatment attributes differed across countries (not shown). These findings could inform the design and treatments of RA prevention trials by considering acceptable benefit-risk tradeoffs observed in this study. [General population results only. More results will be added following completion of the other objectives by the end of April 2021]</p> <p><b>Methodological:</b> DCE and PPT methods produced similar modeling results for MAR and were relatively easy for respondents in our general population to understand and complete. However, DCE was easier to understand relative to PTT. [More following completion of the other objectives by the end of April 2021.]</p>		

<b>Limitations</b>	<p>A potential limitation of this study is external generalizability. Respondents in our general population samples may have preferences that differ from respondents in those countries as a whole. The FDR samples were convenience samples and may not generalize externally to all FDRs. Selection bias is a potential limitation of this study because respondents will be recruited through online panels and/or were required to use the internet. The preferences of those who participate in an online panel may be systematically different from those who do not. In addition, the preferences of those potential respondents who choose to complete the survey may be systematically different than those who do not, including those with FDRs who are recruited via their relatives with RA that were identified in a clinical setting.</p>
<b>Potential Use of Results in MPLC Decision-Making</b>	<p><b>Patients and/or caregivers:</b> For FDRs or the general population who may be faced with the decision on whether or not to choose RA preventive treatment, this study and preference studies like it provide a means for preferences to be elicited.</p> <p><b>Industry:</b> The results of this study would be helpful to industry in informing development decisions in RA delay or prevention. There are both approved medicines and medicines in development that might be acceptable for RA prevention indications. The preferences expressed in this survey provide guidance on what benefit-risk trade-offs may be acceptable, and to whom, in future treatments that might be developed to reduce the risk of RA.</p> <p>The comparison of DCE and PTT approaches and evaluation of samples across countries provide some information about the robustness and reliability of preferences to inform decisions. These results may contribute to regulatory guidance on use of preferences or specific methods in decision making. In addition, these results may inform reviews of RA prevention evidence packages for medication approval or reimbursement decisions. Lastly, the comparison of preferences between FDRs and a general population may be of interest to payers, who make health allocation decisions with the general population in mind.</p>



<b>Therapeutic Area</b>	<b>Lung Cancer</b>
<b>Core Case Study, Academic Study, Industry Study</b>	Core Case Study
<b>Current status</b>	This case study will be finalized in April 2021
<b>Academic Lead, Industry Lead, Methods Lead, Clinical Lead</b>	<p>Academic lead: Serena Oliveri (serena.oliveri@ieo.it)</p> <p>Industry co-lead: Meredith Smith (Meredith.Smith@alexion.com)</p> <p>Methods Lead: Jorien Veldwijk (<a href="mailto:veldwijk@eshpm.eur.nl">veldwijk@eshpm.eur.nl</a>)</p> <p>Clinical co-lead: Gabriella Pravettoni (gabriella.pravettoni@ieo.it)</p> <p>Clinical co-lead: Marina Garassino (Marina.Garassino@istitutotumori.mi.it)</p>
<b>Title of the Study</b>	PREFER Project Lung Cancer Case Study
<b>Background/ Study Rationale</b>	<p>Lung cancer is the most common malignancy in men and the third most common in women. The high prevalence of late stage diagnosis is one of the reasons why lung cancer has such a low survival rate, with a 5-year overall survival rate of 18.1% in all lung cancer stages and 4.5% in metastatic stage.</p> <p>In recent years, there has been a shift in the treatment paradigm for Non-Small Cell Lung Cancer (NSCLC). Different compounds targeting the receptors Programmed Death 1 (PD-1) and PD-Ligand 1 (PD-L1), involved in the regulation of anti-tumor immune response, have shown superior efficacy over standard treatments across various indications. The combination of Immuno-Oncotherapy (IO) and chemotherapy (CT) has been approved by both the FDA and EMA. The combination of chemo-immunotherapy was found to be superior to CT alone in terms of both PFS and OS. Chemo-immunotherapy obtained higher Overall Response Rate and longer duration of response than CT.</p> <p>While the clinical benefit of combined chemo-immunotherapy is clear for patients with PD-L1 low/negative NSCLC for which the only alternative is standard chemotherapy, data are less clear for those with PD-L1 positive NSCLC who can choose between IO and chemo-immunotherapy combination. No direct comparison exists between chemo-immunotherapy and IO alone in this subgroup of patients. This has led to the rise of two new alternative potential treatment algorithms for NSCLC in PD-L1 positive patients, with no data to guide clinicians and patients' decisions between these alternatives. Due to the lack of a clear best choice between the two alternative standards of care, experts recommend that physicians discuss treatment options with patients and individualize treatment decisions based on patients' preferences. Patients and clinicians will face a</p>

	<p>choice between a more aggressive treatment with a greater impact on their quality of life versus an alternative that may be less effective but that carries fewer side effects.</p> <p>In such instances of clinical equipoise, patients' preferences allow clinicians to quantify how much risk is acceptable for the patient to obtain an uncertain benefit.</p>
<p><b>MPLC Decision-making</b></p>	<p><b>Main stakeholder (industry, regulatory, HTA): Regulatory, Industry</b></p> <p><b>MPLC decision-point of interest:</b> Point of marketing authorization application; post-marketing for reimbursement decisions. Knowing patient preferences for different treatment options' attributes could be valuable for clinicians when determining the therapeutic plan, but it can also be precious valuable for marketing authorization, reimbursement decisions, and regulators providing information on patients' perception of relevant benefit and risks related to different treatments options and the level of risk they are willing to tolerate when making decisions between alternatives.</p>
<p><b>Clinical Research Questions</b></p>	<p>The primary clinical objective of this study is to assess the tradeoff between benefit and risks related to treatment alternatives that patients are willing to accept in line with their preferences for relevant attributes related to Immuno-Oncotherapy treatment alternatives.</p> <p>The secondary clinical objective is to evaluate the maximum acceptable risk (MAR)/Minimum acceptable benefit (MAB) that patients would accept for treatment alternatives (e.g., How much risk or toxicity is acceptable for a specific increase in survival time?/What is the minimum increase in survival time that patient deem worth for an increased risk?)</p>
<p><b>Methodological Research Questions</b></p>	<p>The primary methodological objective of this study is to evaluate to what extent assessing preferences for treatment alternatives in NSCLC with a more demanding instrument, such as a discrete choice experiment (DCE), will provide higher quality information if compared to a less expensive method like the swing weighting task (SW).</p> <p>Another secondary objective is to assess patients' characteristics (clinical and demographic information, psychological characteristics, stage of the disease) that may explain heterogeneity in preferences.</p> <p>A final methodological objective of this study is to evaluate the use of an educational tool to increase patients' understanding of treatment options and related attributes, and how to complete 2 different preference elicitation tasks (i.e., discrete choice experiment and swing weighting).</p>
<p><b>Study Design</b></p>	<p>Overall:</p> <p>1. Qualitative study to identify patient-relevant characteristics of lung cancer treatment.</p> <p>The study design for the qualitative phase included the following steps:</p> <ul style="list-style-type: none"> <li>Literature review to identify treatment characteristics that are most relevant to lung cancer patients to be discussed in the focus groups;</li> </ul>

	<ul style="list-style-type: none"> <li>• 4 focus groups (2 in Italy, 2 in Belgium) to identify attributes which patients consider most relevant, followed by thematic inductive analysis of the transcripts to extract overarching themes from the attributes;</li> <li>• 2 additional focus groups (1 in Italy, 1 in Belgium) conducted using the Nominal Group Technique (NGT) to rank and prioritize the final set of attributes;</li> </ul> <p>2. Quantitative study to evaluate the maximum acceptable risk (MAR)/Minimum acceptable benefit (MAB) and to compare two different preference elicitation tasks (DCE vs SW).</p> <p>The study is cross-sectional and consists of a one-time online quantitative preference survey administered to 504 NSCLC patients. The preference survey will be administered to patients accessing one of the participating clinical sites (2 in Italy, 1 in Belgium). Participants will be divided depending on the stage of their diagnosis. Patients in early stages will participate to the study after a surgical treatment, during the follow-up clinical program. Patients in later stages will complete the survey while receiving their cancer treatments.</p> <p>Initially the survey had to be completed on a tablet, whereas the procedure has been revised due to the current pandemic situation. All the patients are contacted and enrolled for this study by phone, and invited to complete the survey on line. The survey includes the following components:</p> <ol style="list-style-type: none"> <li>1. An interactive digital educational module that will provide information about the disease, the treatment options and the preference elicitation tasks;</li> <li>2. Discrete Choice Experiment and Swing Weighting preference elicitation tasks (order will be alternated to control for its effect);</li> <li>3. Psychological instruments that assess health literacy and numeracy, health locus of control;</li> <li>4. Patients' acceptability and evaluation of the educational tool, Discrete Choice Experiment and Swing Weighting tasks;</li> <li>5. Patients' quality of life.</li> <li>6. Open questions about general themes that emerged during the focus group discussions and suggestions for the survey.</li> </ol> <p>Educational materials: Digital tool was developed to provide interactive education and training on: a) significant attributes related to Lung Cancer disease and therapies; b) how to complete two different preference elicitation tasks: a discrete choice experiment (DCE) survey and a swing weighting (SW) exercise.</p> <p>Psychosocial constructs: Health literacy, locus of control.</p>
<b>Study Population</b>	Patients with stages I-IV who have been diagnosed with non-small cell lung cancer (NSCLC)

<b>Exposure &amp; Outcome</b>			
<b>Study Setting</b>	Hospital		
<b>Statistical Methods</b>	Descriptive statistics, including respondent demographics, outcomes from psychological instruments, and Likert scales will be reported. Preference weights and MAR/MAB, will be estimated using appropriate modeling techniques for DCE and SW. Results will be stratified by country. Results may be pooled if deemed appropriate based on the Swait & Louvriere test. Additional analyses will be conducted to characterize respondents and to explore preference heterogeneity.		
<b>Results</b>	<b>PREFER Research Questions</b>	<b>How was it addressed in this preference study</b>	<b>What was found</b>
	<p>What attributes of lung cancer treatments are important to NSCLC patients?</p> <p>[Qualitative study]</p>	<p>Data were collected in Italy and Belgium using a combination of methodologies (i) 4 focus group discussions and thematic analysis of transcripts, and (ii) 2 additional focus group discussion with nominal group techniques to refine the list of attributes.</p>	<p>Patients identified attributes that were important to them and these were consistent across countries. Three overarching themes emerged:</p> <ol style="list-style-type: none"> <li>1. Positive effects or expected gains of treatment associated with greater life expectancy, decrease in cancer growth, cancer remission, and maintenance of daily functioning</li> <li>2. Negative effects or adverse events related to therapy affecting daily functioning. These spanned common (e.g. severe skin problems, nausea, infections, hair loss, infusion reaction, gravity of edema, fatigue, weight gain) and less common (e.g. probability of renal failure, having pain, hearing impairment, cognitive limitations, difficulty in breathing) side effects. Some cultural differences emerged, whereby Italians reported vomiting, diarrhea, and sexual impotence, while Belgians complained weeping eyes.</li> <li>3. Uncertainty regarding the duration of either positive and negative effects of treatment and need for clearer communication.</li> </ol>

			<p>Final list of most important attributes included in the quantitative phase:</p> <ol style="list-style-type: none"> <li>1) Probability of survival 5 years from the start of cancer treatment;</li> <li>2) How the treatment is administered to you;</li> <li>3) Probability of feeling extremely tired;</li> <li>4) Probability of long-lasting skin problems;</li> <li>5) Severity of hair loss.</li> </ol>
	<p>To what extent assessing preferences for treatment alternatives in NSCLC with a more demanding instrument, i.e., a discrete choice experiment (DCE), will provide higher quality information if compared to a less expensive method, i.e., the swing weighting task (SW)?</p> <p>[Quantitative study]</p>	<p>Treatment attributes identified in the qualitative phase will be evaluated in a new sample of patients using two different types of elicitation methods:</p> <ol style="list-style-type: none"> <li>1. Discrete choice experiment (DCE) – where patients evaluate scenarios that feature different levels of each attribute, and select the preferred combination of attributes</li> <li>2. Swing weighting task (SW) – where patients rank attributes based on the importance of their improvement from the worse to the best level, and assign weights to express the importance of each improvement.</li> </ol> <p>Results obtained using the two methods will be compared both qualitatively and quantitatively.</p>	<p>Not yet available</p>

		<p>Some of the comparisons will involve validity checks, completion time, and drop-out, as well as direct comparisons of the relative importance of each attribute.</p> <p>The online survey will also assess the intuitiveness of each tool (e.g. ., "How easy or difficult was it for you to understand the questions?" on a 5-point Likert scale from 'very easy' to 'very difficult')</p>	
	<p>What is the maximum acceptable risk (MAR)/Minimum acceptable benefit (MAB) that patients would accept for treatment alternatives?</p> <p>[Quantitative study]</p>		<p>Not yet available</p>
	<p>To what extent do patients' characteristics (clinical and demographic information, psychological characteristics, stage of the disease) explain heterogeneity in preferences?</p> <p>[Quantitative study]</p>	<p>Participants will be asked questions regarding socio-demographic characteristics and clinical aspects on a 5-levels scale (5-level EQ-5D version)</p> <p>Psychological constructs will be measured using the following instruments:</p> <p>1. Health literacy and numeracy (Chew's Set of Brief Screening</p>	<p>Not yet available</p>

		<p>Questions, The Newest Vital Sign)</p> <p>2. Health locus of control (Multidimensional Health Locus of Control Scale - Form C)</p>	
	<p>To what extent did participants find an educational tool to be acceptable and helpful to increase their understanding of treatment options and related attributes, and how to complete 2 different preference elicitation tasks (i.e., discrete choice experiment and swing weighting)?</p> <p>[Quantitative study]</p>	<p>The online survey will contain questions regarding the usefulness of the education tool (e.g. "Did the educational material help you understand the features included in the choices?"), which must be rated on a 5-point Likert scale.</p>	<p>Not yet available</p>
<b>Conclusions</b>	<p>The qualitative study revealed that overall patients preferences can be grouped in i) positive effects of expected gains of treatment, ii) negative effects or adverse events related to therapy that negatively impact patients' daily functioning and iii) uncertainty regarding the duration/type of positive and negative effects of treatment. These findings demonstrate the value of qualitative measures to investigate patients trade-offs in treatment preferences, and will inform the quantitative survey.</p>		
<b>Limitations</b>	<p>Limitations of the qualitative study:</p> <ol style="list-style-type: none"> <li>1. Representativeness of the patient sample and possibility of a bias introduced by clinicians when selecting participants for the study</li> <li>2. Lack of quantitative measures of attributes and themes (this limitation is addressed in the PREFER quantitative study)</li> </ol> <p>Quantitative study</p>		
<b>Potential Use of Results in MPLC Decision-Making</b>	<p>In 2019, EMA approved combination immunotherapy + chemotherapy for first-line treatment of metastatic squamous NSCLC. NSCLC patients with PD-L1 expression &gt;50% can now "choose" between:</p> <ul style="list-style-type: none"> <li>- Immunotherapy alone</li> </ul>		



	<p>- Immunotherapy + Chemotherapy</p> <p>Efficacy and patients' compliance for these two therapeutic alternatives depend on: a) Disease aggressiveness; b) Patient's tolerance of higher toxicity.</p> <p>Metastatic NSCLC, stage IV- under discussion whether this will be expanded to include Stage III; Clinical drug development; Optimize &amp; prioritize assets.</p>
--	---



<b>Therapeutic Area</b>	<b>Neuromuscular Disorders (Rare Diseases)</b>
<b>Core Case Study, Academic Study, Industry Study</b>	Academic Study (Newcastle University)
<b>Current status</b>	This case study will be finalized in April 2021
<b>Academic Lead, Industry Lead, Methods Lead, Clinical Lead</b>	Clinical Lead: Grainne Gorman (Newcastle University) Academic Lead: Ardine de Wit (University Medical Center Utrecht – Utrecht University) Industry Lead: Cathy Anne Pinto (Merck) Methods Lead: Esther de Bekker-Grob (Erasmus University Rotterdam)
<b>Title of the Study</b>	Quantifying patient preferences in neuromuscular disorders: a case study of the IMI PREFER Project
<b>Background/ Study Rationale</b>	<p>Neuromuscular diseases (NMD) represent uncommon, serious and debilitating (i.e. muscle weakness) conditions; all progressive with poor prognosis and with limited or no treatment options available. For this case study, we have selected two neuromuscular disorders (i.e. myotonic dystrophy type 1 [DM1] and mitochondrial myopathies [MM]) that despite having different physio-pathological pathways both manifest in a clinically similar manner. In both diseases, different body tissues and organs may be affected in function or during development resulting in a multisystem and heterogeneous phenotype. General cognitive deficits have been described in over 60 to 70% of patients and the prevalence and severity depends on the age at onset of the disease, with patients with symptoms established earlier in life more prone to this phenotype. There is no specific cure for either disease, or standard of care offers few options for managing symptoms, although there are emerging potential treatments. Addressing patient preferences for potential treatments in development could be a sensitive action in this situation as these will inform stakeholders about gaps in knowledges (e.g. underestimated medical needs and risk tolerances) or would highlight treatment expectations that could differ between patients and those currently established by health-care practitioners. However, the presence of cognitive deficits and the reliance on caregivers for relevant decision making, makes caregivers an additional party of interest when addressing NMD treatment needs.</p> <p>In summary, the key points that can categorize this study as a patient preference study in a preference sensitive situation are:</p> <ul style="list-style-type: none"> <li>-Lack (or absence) of treatment options</li> </ul>

	<p>-Patients may value treatment attributes differently (between each other and when compared to other stakeholders)</p> <p>-Due to the rareness and heterogeneity of these disease groups, the perspectives of the population may not be yet well understood</p>
<p><b>MPLC Decision-making</b></p>	<p>Main stakeholder (industry, regulatory, HTA): Industry and Regulators</p> <p>MPLC decision-point of interest:</p> <p>Industry: potential value of new developments, decisions on relevant endpoints for future clinical trials</p> <p>Regulators: benefit risk decision-making processes of patients</p>
<p><b>Clinical Research Objectives</b></p>	<p>Primary:</p> <p>To elicit and quantify patient preferences, including benefit to risk trade-offs (e.g. relative importance, minimum acceptable benefit (MAB), maximum acceptable risk (MAR)) for future NMD treatments.</p> <p>Secondary:</p> <ul style="list-style-type: none"> <li>- To describe and compare preferences among the different participant subgroups (i.e. disease type, disease phenotypes, patients and caregivers): <ul style="list-style-type: none"> <li>◦ To assess how generalizable preferences (i.e. relative importance, MAB and MAR) are from one specific disease to a different disease but with similar clinical characteristics (i.e. DM1 and MD);</li> <li>◦ To understand the degree to which patients' preferences and caregiver preferences align with each other;</li> <li>◦ To identify clinically meaningful subgroups based on the association between specific preferences and specific demographic and clinical characteristics.</li> </ul> </li> <li>- To describe heterogeneity of responses resulting from different preference elicitation methods for patients and caregivers based on demographics and medical history (e.g. patients disease severity; prior treatment history; and, genetic status of caregivers)</li> <li>- To demonstrate the feasibility and acceptability of assessing PPI in a population that may have varying levels of cognitive limitations.</li> </ul>
<p><b>Methodological Research Objectives</b></p>	<p>Primary:</p> <p>To describe and compare results obtained from three different preference assessment methods (DCE, BWS type 2, Q-methodology);</p>

	<p>To conduct intra-methods comparability analyses</p> <p>To analyze relatively simpler versus more complex preference elicitation methods.</p> <ul style="list-style-type: none"> <li>- To describe and compare responses (e.g. preference results, heterogeneity in responses, compliance and level of understanding) from three different preference elicitation methods applied within the same disease population.</li> <li>- To describe and compare responses reliability and preference variability between two related diseases with similar clinical characteristics (i.e. DM1 and MD patients).</li> <li>- To describe and compare preferences of two different types of stakeholder groups (i.e. comparing patient preferences with caregivers' judgments regarding patient preferences).</li> <li>- To describe associations between responses from a preference elicitation method and results obtained from psychosocial constructs assessments.</li> </ul>
<b>Study Design</b>	<p>Overall: Online survey with a cross-over design to compare methods and with a cross-sectional design to compare subgroups. The study design is depicted in the graph below:</p>

	<table border="1"> <thead> <tr> <th></th> <th>Group 1: Patients with early disease onset (diagnosis/symptoms &lt;20 years old)</th> <th>Group 2: Patients with late disease onset (diagnosis/symptoms ≥20years old)</th> <th>Group 3: Caregivers</th> </tr> </thead> <tbody> <tr> <td><b>BWS, Type 2</b></td> <td colspan="3">[Diagram: Three blue stars in a row with horizontal double-headed arrows between them, indicating between-group comparisons.]</td> </tr> <tr> <td><b>Q-Methodology</b></td> <td colspan="3">[Diagram: Three red stars in a row with vertical double-headed arrows between them, indicating within-group comparisons.]</td> </tr> <tr> <td><b>DCE</b></td> <td colspan="3">[Diagram: Two green stars in a row with horizontal double-headed arrows between them, indicating between-group comparisons.]</td> </tr> </tbody> </table> <p>Legend:    Between group comparisons or same methods   Within group comparison of results using different methods</p>		Group 1: Patients with early disease onset (diagnosis/symptoms <20 years old)	Group 2: Patients with late disease onset (diagnosis/symptoms ≥20years old)	Group 3: Caregivers	<b>BWS, Type 2</b>	[Diagram: Three blue stars in a row with horizontal double-headed arrows between them, indicating between-group comparisons.]			<b>Q-Methodology</b>	[Diagram: Three red stars in a row with vertical double-headed arrows between them, indicating within-group comparisons.]			<b>DCE</b>	[Diagram: Two green stars in a row with horizontal double-headed arrows between them, indicating between-group comparisons.]			
	Group 1: Patients with early disease onset (diagnosis/symptoms <20 years old)	Group 2: Patients with late disease onset (diagnosis/symptoms ≥20years old)	Group 3: Caregivers															
<b>BWS, Type 2</b>	[Diagram: Three blue stars in a row with horizontal double-headed arrows between them, indicating between-group comparisons.]																	
<b>Q-Methodology</b>	[Diagram: Three red stars in a row with vertical double-headed arrows between them, indicating within-group comparisons.]																	
<b>DCE</b>	[Diagram: Two green stars in a row with horizontal double-headed arrows between them, indicating between-group comparisons.]																	
<b>Study Population</b>	<p>DM1 and MD patients and caregivers 18 years old or older volunteering for the study and self-reporting their personal (or their patients) diagnosis.</p> <p>Participants were stratified into three different groups, as followed:</p>																	

The attributes included in BWS and DCE are: cognition, muscle strength, energy and endurance, balance, (risk of) blurry vision, (risk of) liver damage.

Educational materials: A set of 'story line videos' based on two components: 1) Introduction to explain why patients are participating; and, 2) Instruction manuals (one for each specific elicitation method). Each with voice-over English-reading and lasting about 5 minutes.

Psychosocial constructs: Chew's Set of Brief Screening Questions and Subjective Numeracy Scale

PRO: The ACTIVLIM questionnaire is a Rasch model assessment tool that measures limitations in the execution of daily life activities that a patient might face.

	<p>Group 1 (Patients): self-reported as diagnosed with DM1 or MD patients with early onset of disease (established diagnosis or first reported symptoms before 20 years of age).</p> <p>Group 2 (Patients): self-reported as diagnosed with DM1 or MD patients with late onset (established diagnosis or first reported symptoms on or after 20 years of age)</p> <p>Group 3 (Caregivers): caregiver defined as a spouse, partner, parent, legal guardian, close relative, or other adult close to the family living either in the same house or in contact with the DM1 or MD patient in a caregiver relationship at least 4 times/week for at least one hour or more per day.</p>
<b>Exposure &amp; Outcome</b>	<p>Participants will need to complete an online survey that will be administered at two different time points with a 2-week period (+-7 days) in between. The initial section of the survey will include (additional to one of the different preference elicitation methods) questions to assess participants demographical and clinical characteristics.</p>
<b>Study Setting</b>	<p>Online survey. Patients that are member of patients associations or patient registries will be invited to participate in the survey. Patient organizations from different countries will collaborate with the NMD team as recruitment platforms: UK, Canada, US, Australia &amp; New Zealand. In addition, direct invitations to clinical patient populations registered on the UK Myotonic Dystrophy Patient Registry and involved on the Newcastle Mitochondrial Disease Patients Cohort.</p>
<b>Statistical Methods</b>	<p>BWS2 data will be analysed in several ways. First, counts analysis will be performed both on sample and on respondent level. On a sample level, for each attribute an aggregated best-minus-worst score will be determined. Based on these aggregated scores, the attribute levels will be ranked. Second, mean relative importance scores will be calculated for each attribute and the top three (for worst and best) will be used to determine surrogates of MAB and MAR.</p> <p>In our study, WLS will be used as relatively simple regression model to explore the best-worst data. With WLS, which can be considered as an extension of the well-known ordinary least squares (OLS) regression method, a standard regression model is estimated while using the choice frequencies of the attribute levels as weights in the WLS regression. MNL models will be used to gain insights into the importance of attribute levels without taking scale and preference heterogeneity into account. More sophisticated models like for example MXL and/or LCM will also be used to account for observed preference heterogeneity. Estimates from these models will be used to calculate the relative importance of attribute levels, both for subgroups or the entire sample.</p> <p>DCE analysis:</p> <p>A multinomial logit model (MNL) with error term heteroscedasticity (or scale variation) and observed preference heterogeneity or a scaled Latent Class Model (LCM) to analyse the choice observations will be employed to analyse DCE data. MAB and MAR estimates will be based on the estimated DCE coefficients. Confidence intervals intervals based on the individual specific MAR and MAB will be estimated.</p>

Results	PREFER Research Questions	How was it addressed in this preference study	What was found
	2d1. How similar are the results of simpler/faster/cheaper methods vs. more rigorous/in depth/expensive methods	By Comparing DCE vs BWS and BWS vs Q-Methodology	DCE was not found to be more difficult to complete or more difficult to understand than BWS. Results of BWS and DCE differed with regard to relative importance scores for the 6 attributes. More results will be added later.
	4a3. Are the results of these format questions consistent in different patient populations and disease areas?	By Comparing results from the MD group with the DM1 group	Results will be added later
	9b. How generalizable are the preferences from one specific population in a disease to different populations in that or related diseases?	By comparing results from MD patients to DM1 patients By comparing results from patients to those from caregivers By comparing the milder phenotype (i.e. Group 2 patients) to the more affected phenotype (i.e. Group 1 patients)	Results will be added later
	10b. Can the measurement of psychosocial constructs including PREFER Class I constructs provide insight?	By analyzing the impact of health literacy and numeracy scores to test compliance and responses of dominant test	Results will be added later
	11. To what degree do preference results in a disease vary based on: characteristics of patients,	By comparing preferences from patients to those of the caregivers (in the BWS method applied to the whole sample).	Results will be added later

	<p>stakeholders, information level of patients, disease experience or severity and region.</p>	<p>By analyzing preferences in the BWS based (regression model) on clinical characteristics such as time since disease onset and disease severity (i.e. results from patient-reported outcome).</p> <p>(If feasible) by comparing preferences of responses from participants of different regions (i.e. US, Canada, UK, Australia and New Zealand).</p>	
<p><b>Conclusions</b></p>	<p>Will be added upon finalization of the study.</p>		
<p><b>Limitations</b></p>	<p>Limited sample size.            Different recruitment methods (clinical population and members of patient associations)            Diagnosis and group assignment (group 1 or group 2) not medically confirmed            Sample heterogeneity – different levels of disease severity, heterogeneous phenotypes (different symptoms might be present in different degrees), and caregivers may represent a different spectrum of the patient population that patients participating themselves (e.g. paediatric population and very severe patients would only be represented by caregivers).</p>		
<p><b>Potential Use of Results in MPLC Decision-Making</b></p>	<p>These results will inform stakeholders investing (or interested in investing) at early stages of the drug development process, both to inform potential treatment profiles but also highlighting priority endpoints when designing clinical trials.</p>		

### 8.1.2 Prospective additional industry case studies

<b>Therapeutic Area</b>	<b>Chronic obstructive pulmonary disease (COPD)</b>
<b>Core Case Study, Academic Case Study, Industry Case Study</b>	<b>Industry Case Study</b>
<b>Current status</b>	This case study is finalized.
<b>Academic Lead, Industry Lead, Methods Lead, Clinical Lead</b>	<b><u>Industry Leads</u>: Nigel S. Cook, Florian Gutzwiller</b>
<b>Title of the Study</b>	<b>Patient Preferences in Benefit-Risk Assessments during the Drug Life Cycle</b>
<b>Background/ Study Rationale</b>	<p>The measurement of lung function (FEV1) and exacerbations are traditional measures in clinical trial of chronic obstructive pulmonary disease (COPD). With the exception of shortness of breath, the measurement of symptoms is much less established. However, previous research evaluating patient perspectives in COPD has shown that besides breathlessness and exacerbations, persistent cough, mucus production, sleep disturbance and urinary incontinence are also of great concern for patients. Because patients are impacted daily by excess mucus production, (chronic) cough (which usually occur together) and shortness of breath in addition to downstream impacts on sleep, physical activity, and as urinary leakage (caused by frequent coughing), COPD imposes a substantial burden on patients' lives.</p> <p>To evaluate the preference of COPD patients, Novartis designed a systematic mixed-methods research program. A targeted literature review, social media listening (SML) study and a qualitative online bulletin board (OBB) study informed the design of this quantitative patient preference study using a discrete choice experiment (DCE). The findings from the literature review, SML and OBB studies, alongside evidence from the clinical guidelines and other preference literature/best practice guidelines have been used to inform the identification and development of attributes and levels for the COPD PPS. This PPS sought to quantify the relative importance that patients place on alleviation of excess mucus production, (chronic) cough and shortness of breath in addition to downstream impacts on sleep, physical activity, and as urinary leakage, including exacerbations.</p>
<b>MPLC Decision-making</b>	<b><u>Main stakeholder (industry, regulatory, HTA)</u>: The primary stakeholders for this study are regulators, industry and HTA.</b>



<p><b>Clinical Research Questions</b></p>	<p>Study aim: To quantify the preferences of COPD patients regarding symptoms, the impact on their QOL, and evaluate whether preferences vary with certain respondent characteristics.</p> <p>Primary objective: The level of importance which COPD patients place on symptoms such as cough and mucus secretion and consequences thereof, versus more traditional endpoints such as breathlessness and exacerbations. During the conduct of the study the overall strategy changed and therefore the analysis plan was adapted to analyze preference for improvement of excess mucus production, (chronic) cough and shortness of breath in addition to downstream impacts on sleep, physical activity, and as urinary leakage compared to improvement of exacerbations alone.</p> <p>Secondary objective: To provide predicted choice probabilities on trade-offs COPD patients are willing to make to achieve desired COPD disease states.</p> <p>Exploratory objectives: To assess how stated preferences may vary according to patient symptomatic burden and disease severity; socio-demographic characteristics and the level of patient 'activation'.</p>
<p><b>Methodological Research Questions</b></p>	<ul style="list-style-type: none"> <li>• 4b1: What tests assess whether patients can perform a given set of cognitive tasks?</li> <li>• 10b1: Can psychosocial constructs including PREFER Class I constructs and Class II constructs provide insight regarding an individual's preferences?</li> <li>• 10c2: Help to show external validity, particularly in conjunction with PRO or related instruments specific to the disease area?</li> <li>• 11: To what degree do preference results in a disease vary with Characteristics of patients, Stakeholder, (11d) Disease experience or severity, Treatment experience, (11f) Region, (11g) Culture, Healthy vs heaving disease, Time of diagnosis, Longitudinal sampling, Patients recruited through different channels?</li> <li>• 15b2: What is the impact of including attributes related to the disease/quality of life/price/... on the preferences elicited?</li> <li>• 15b3: When in the drug development process should preferences be measured for informing MA/payer/industry decisions?</li> <li>• 15c2: When along the MPLC should a patient preference study be conducted to inform regulators or HTA?</li> </ul>

<p><b>Study Design</b></p>	<p>Qualitative phase: As preparation for this PPS in COPD, a targeted literature review, a SML study (<a href="http://doi.org/10.1183/23120541.00128-2018">http://doi.org/10.1183/23120541.00128-2018</a>) and a qualitative OBB study (<a href="http://doi.org/10.2147/COPD.S202580">http://doi.org/10.2147/COPD.S202580</a>) were previously conducted. The synthesized findings from the studies, alongside evidence from clinical guidelines and other preference literature/best practice guidelines were used to inform the identification and development of attributes and levels for the COPD PPS.</p> <p>Quantitative phase: The findings from the qualitative research (described in section 2.1), alongside evidence from clinical guidelines and other preference literature/best practice guidelines, were used to inform patient inclusion and exclusion criteria, the identification and development of an attribute and levels (A&amp;L) grid for DCE in this COPD PPS. These input were reviewed by the project team, clinical experts and patient group representatives from 5 countries (British Lung Foundation, UK; COPD Foundation, US; La Fondation du Soufflé, France; Lung Foundation Australia, Australia and J-Breath, Japan) to finalize the design of the PPS comprising three phases. NICE were consulted for scientific advice on the study design and statistical analysis (NICE provides first scientific advice on PPS design). The study design was also informed by the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) taskforce guidelines and FDA guidance documents for</p>
<p><b>Study Population</b></p>	<p>Patients were included in the study if they were aged <math>\geq 40</math> years at the time of survey with diagnosis of COPD by a healthcare practitioner, had self-reported moderate-to-severe COPD defined as having had <math>\geq 1</math> COPD flare-up/exacerbation in the last year (requiring additional treatment with oral corticosteroids and/or antibiotics to manage their chest symptoms, or had to visit the hospital/emergency room due to COPD) along with daily symptoms of cough and excess mucus production. Patients were excluded if they had participated in</p>
<p><b>Study Setting</b></p>	<p>Primary recruitment was conducted via patient support groups and participation was voluntary. Patient groups in the UK, USA, France, and Australia leveraged their patient's network for the patient recruitment. Supplementary recruitment via patient panels was introduced once patient group recruitment was saturated, with participation compensated at fair market value to encourage patient involvement. In Japan, patients were recruited only through patient panels.</p>

<p><b>Statistical Methods</b></p>	<p>Data collected from Phase 2 quantitative survey were analysed using IBM SPSS data collection survey reporter software and P-values reported where required. Hierarchical Bayesian analysis with effects-coding parameterization (subject to acceptable model fitting as confirmed by multinomial logistic regression (MNL) was undertaken on the preference data (alternative disease state) to robustly estimate (using Gibbs sampling) the relative value each respondent places on an attribute level (i.e., the part-worth utilities). Interactions between attributes were tested in terms of their contribution to the model, this was measured in terms of significant change in model fit. In order to represent the effect of current treatments (e.g. standard of care: bronchodilator and/or inhaled corticosteroids [ICS]) and future treatments (for alleviating cough and mucus production), sensitivity analysis was conducted. A simulator tool was developed (built in Excel 2016) to predict “preference shares” (expressed as a percentage) for input profiles. Profile scenarios were simulated in which the health state across different attributes were altered from the average patient (“DCF index”) to</p>		
<p><b>Results</b></p>	<p><b>PREFER Clinical and Methodological Research Questions</b></p>	<p><b>How was it addressed in this preference study</b></p>	<p><b>What was found</b></p>
	<p>10b1 Can psychosocial constructs including PREFER Class I constructs and Class II constructs provide insight regarding an individual’s preferences, and</p> <p>4b1 What tests assess whether patients can perform a given set of cognitive tasks?</p>	<p>We included a series of survey experience questions to determine things such as the ease of understanding, ease of answering, level of interest to the patient.</p>	<p>Results indicate the study was understandable and that they did not have a problem to answer and complete the survey.</p> <p>The qualitative debriefing interviews prior to finalization of the study design also highlighted areas where respondents had difficulty to understand/answer, allowing us to address these points in the final survey questionnaire and</p>
	<p>10c2 - Help to show external validity, particularly in conjunction with PRO or related instruments specific to the disease area?</p>	<p>With both general &amp; disease specific PRO instruments.</p>	<p>PRO questionnaires were used including the CASA-Q, and CAT to assess the severity and impact of COPD in drug development studies for COPD, PAM-Q for attitudinal and EQ-5D [3L] - overall health status</p>

	<p>11 - To what degree do preference results in a disease vary with Characteristics of patients, Stakeholder, Information level of patients, Disease experience or severity, Treatment experience, Region, Culture, Healthy vs heaving disease, Time of diagnosis, Longitudinal sampling, Patients recruited through different channels?</p>	<p>COPD severity was assessed using four variables: 1) self-perceived severity; 2) number of exacerbations in past 12 months that required hospitalization; 3) COPD Assessment Test (CAT) total score and 4) Cough and Sputum Assessment Questionnaire (CASA-Q) total and domain scores. Patient Activation was assessed with the PAM-Q</p>	<p>Overall, preferences were consistent across the three severity groups. Similarly, patient preference results across activation levels were consistent with the overall findings with no evidence of an association of preferences with activation level.</p>
	<p>11d To what degree do preference results in a disease vary with Disease experience or severity?</p>	<p>Patient preferences were analysed per attribute according to disease severity</p>	<p>Patient preferences were consistent across the severity groups.</p>
	<p>11f To what degree do preference results in a disease vary with Region? and 11g To what degree do preference results in a disease vary with Culture?</p>	<p>The purpose of the study was to elicit overall preferences for symptoms, therefore inter-country comparisons were not in our focus. For the purpose of this IMI case study, some results can be given:</p>	<p>Our study spanned 5 countries. Across countries patients placed almost equal importance on shortness of breath, sleep quality and urinary incontinence. Some differences in relative importance did emerge across the countries: shortness of breath was the most important attribute for patients from Australia and France, and sleep quality for patients from Japan. As we did not specifically categorize the participants by culture, we can only reply to this</p>

	<p>15b2 - What is the impact of including attributes related to the disease/quality of life/price/... on the preferences elicited?</p>	<p>Prior to doing the DCE exercise, the study respondents performed a profile matching exercise where they were asked using the attribute and level grid to select the level for each attribute which most closely matched their current disease state.</p>	<p>This information served to categorize the clinical status of the patients and also is being used to generate a 'DCE Index' for each patient, which will be used for correlation analysis with EQ-5D scores. Additionally, this exercise served to familiarize patients with the full context of the DCE, attributes and levels, prior to beginning the choice task.</p>
	<p>15b3 - When in the drug development process should preferences be measured for informing MA/payer/industry decisions, and</p> <p>15c2 - When along the MPLC should a patient preference study be conducted to inform regulators or HTA?</p>	<p>n/a</p>	<p>Gaining patient insights by collecting information on their disease experiences and perspectives should be an ongoing process throughout the product development lifecycle and starts in very early development phases. Regulatory authorities also acknowledge the importance of early patient engagement and appreciate and early engagement with authorities on methods to collect patients' insights. Collectively these initiatives should lead to</p>

	<p>17a - What is the optimal approach towards recruitment of participants? What incentives are needed to recruit patients into preference studies?</p>	<p>Our PPS conducted primary recruitment via patient support groups and participation was voluntary.</p>	<p>Patient groups in the UK, US, France, and Australia leveraged their patients' network for the patient recruitment. Supplementary recruitment via patient panels was introduced once patient group recruitment was saturated, with participation compensated at fair market value to encourage patient</p>
<p><b>Conclusions</b></p>	<p>The study showed patient preference studies are a useful way to assess which aspects of a disease matter most to patients. Analysis of changes in preference weights (utilities) for equivalent improvements in cough and mucus combined, are higher (more valued) than those for shortness of breath alone. As a result, if these improvements are included in a health-state preference simulation (with two profiles: A) cough and mucus improved and B) shortness of breath improved) there is a clear preference for the cough and mucus-improved profile. When comparing two profiles A) daily symptoms improved and B) exacerbations improved, there is a clear preference for the daily symptoms improved profile. This study also showed that gathering patients' insights via a structured and systematic approach is feasible early in the product lifecycle and may benefit</p>		

<b>Limitations</b>	<p>There are several important limitations to this study, starting with the range of options offered for each attribute. The worst outcome offered for an attribute is “hospitalisation” for the exacerbation attribute. This is possibly introducing a bias:</p> <ul style="list-style-type: none"><li>a) Temporal effect<ul style="list-style-type: none"><li>a. Hospitalisations due to exacerbations (severe exacerbations) are much less frequent than exacerbations treated with additional medication (moderate exacerbations). This means that over time there is more benefit accrued by treating exacerbations with medications than with hospitalisations. This temporal effect is only implicitly taken into account, in that it is assumed that when patients performed the DCE, they accounted for these temporal differences. The lack of explicit adjustment gives the exacerbation attribute possibly more weight.</li><li>b. The temporal effect is even larger, when frequency of hospitalisations due to exacerbations are compared to daily impact of e.g. cough and sputum symptoms.</li></ul></li><li>b) Range of options offered<ul style="list-style-type: none"><li>a. A hospitalisation is the worst possible outcome across all attributes. However, it was only offered as an option for the exacerbation attribute. The results suggest a hierarchy of patient preference to first avoid severe acute events leading to hospitalisation. It is not surprising therefore that patients the highest single weight to this attribute level compared to all others.</li><li>b. It is also questionable, whether true independence can be assumed for exacerbations, as they represent a worsening of symptoms and are not really a distinct separate entity.</li></ul></li></ul> <p>Nevertheless, the study was designed based on the assumption that the attributes are operating independently and the research has confirmed that the attributes have been considered independent in the context of its meaning for the conjoint. The study used an orthogonal design to ensure attributes and levels were being tested independently from one another – the important contributions of the individual features can be isolated from the rest of the (possibly confounding) effects presented simultaneously as part of the DCE. But this does not mean that there are no interrelationships and/or influences of one or more attributes on another.</p> <p>It is acknowledged that online surveys have certain limitations, especially in an aging population, however,</p>
--------------------	---

	<p>Another limitation of this study was the diagnosis of the COPD as eligibility criteria, which was patient self-reported. Since patients were recruited through patient groups and patient panels (not physicians) there will be no healthcare professional confirmed diagnosis, eligibility will be determined by patient-reported confirmation that they have received a diagnosis of moderate to severe COPD from a healthcare professional. This does pose the risk of inclusion of patients without a correct COPD diagnosis. To mitigate this, patients were recruited only via closed patient group membership or patient panels and the screening questions ensured patients regularly experience the most common symptoms related to COPD (cough and mucus); however this does assume honest answering of the questions posed, which must be recognised as a limitation to the study.</p> <p>In this study patient recruitment was via COPD/respiratory patient support groups or COPD patient panels, this should ensure correct inclusion into the study, however patients who are registered with support groups or panels for research purposes, may also be more engaged with their disease management, which could influence their preferences. Some caution is, therefore, warranted in extrapolating these results to the broader COPD patient population or beyond those countries who were part of the study.</p>
<b>Potential Use of Results in MPLC Decision-Making (Optional)</b>	



<b>Therapeutic Area</b>	<b>ON-GOING CASE STUDY SUMMARY: UPDATED AS OF February 9th, 2021</b> <b>Haemophilia Gene Therapy</b>
<b>Core Case Study, Academic Study, Industry Study</b>	<b>Industry Case Study</b>
<b>Current status</b>	<b>This case study will be finalized in April 2021</b>
<b>Academic Lead, Industry Lead, Methods Lead, Clinical Lead</b>	<p><b>Industry</b></p> <ol style="list-style-type: none"> <li>1. Dr. Ione Woollacott, Study Lead and Author, Pfizer UK</li> <li>2. Jim Thomson, Study Contributor, Pfizer UK</li> <li>3. Dr. Jack Brownrigg, Study Reviewer &amp; Approver, Pfizer UK</li> <li>4. Dr. Ian Winburn, Study Contributor, Pfizer Global</li> <li>5. Dr. Eline van Overbeeke, Collaborator, Pfizer NL &amp; (formerly of) University of Leuven</li> </ol> <p><b>Methods</b></p> <ul style="list-style-type: none"> <li>• Matthew Cawson, Principal Investigator, HCD Economics</li> <li>• George Morgan, Data Analysis, HCD Economics</li> <li>• Dr. Antony Martin, Study Design, (formerly of) HCD Economics</li> </ul> <p><b>Clinical</b></p> <ol style="list-style-type: none"> <li>1. Professor Pratima Chowdary, Study Contributor (Clinical), Royal Free Hospital</li> <li>2. Nicola Dunn, Study Contributor (Clinical), Independent Psychologist</li> </ol> <p><b>Academic</b></p> <ul style="list-style-type: none"> <li>• Sissel Michelsen, Collaborator, (formerly of) University of Leuven</li> <li>• Professor Isabelle Huys, Collaborator, University of Leuven</li> </ul>
<b>Title of the Study</b>	<b>Examining the Preferences of People with Haemophilia for Gene Therapy</b>
<b>Background/ Study Rationale</b>	People with haemophilia (PWH) experience various degrees of bleeding, depending on residual coagulation factor levels. Bleeding occurs most commonly in joints, soft tissue, and muscles, causing short-term symptoms (acute bleeding and acute pain) and long-term-complications (chronic pain, haemophilia arthropathy, and disability). Acute and chronic complications may dramatically reduce the health-related quality of life (HRQoL) of PWH.

	<p>Recent gene therapy trials in both hemophilia A and B have reported promising results, which suggest that gene therapy could yield a long-term functional “cure” by changing or replacing the missing/abnormal genes, leading to sustained in vivo Factor VIII or Factor IX production. Gene therapy could be life changing for PWH, relieving them of the daily burden and limitations they currently experience. Currently, Phase 3 trials of gene therapy for haemophilia are in progress, and these novel therapies have the potential to change the landscape of haemophilia management. However, more research is needed to understand the preferences of PWH for gene therapy. Several studies have explored preferences of PWH, physicians and pharmacists for different haemophilia treatments using discrete choice analyses. However, these have not addressed preferences as gene therapy becomes available.</p> <p>A recent study, the Patient preferences to Assess Value IN Gene therapies (PAVING) study, conducted at the University of Leuven in Belgium, has examined how features of standard therapy and gene therapy for haemophilia influence patients’ choices between these therapies. This qualitative study used semi-structured interviews to identify attributes of gene therapy and standard of care that are important to patients. This was conducted in order to enable better understanding of trade-offs that patients may make when choosing between gene therapy and current standard of care.</p> <p>Further information on patient preferences for gene therapy in the United Kingdom (UK) would provide a clear view of the patient voice and could identify which specific attributes of gene therapy are important to PWH in the UK population. These data will be informative in Health Technology Appraisal (HTA) or other submissions to the National Institute for Health and Care Excellence (NICE) for review of gene therapies for PWH, and for informing future directions, including ongoing clinical trials and patient care. Assessment of haemophilia treatment preferences in PWH is also important as this may help to inform the most appropriate prescribed treatment regimen, which may translate into increased treatment satisfaction and adherence.</p> <p>This study therefore aimed to assess how people with moderate and severe haemophilia (A and B) in the UK value different attributes of treatment, when considering whether to undergo gene therapy.</p>
<p><b>MPLC Decision-making</b></p>	<p><b>Main stakeholder (industry, regulatory, HTA):</b> Industry is leading the study, but impact will also be on regulatory and HTA. Industry, regulatory and HTA bodies can use information from this study to enable understanding of how PWH in the UK value and weigh up different aspects of gene therapy for haemophilia.</p> <p><b>MPLC decision-point of interest:</b> Regulatory approval and HTA/other decision-making for gene therapy for haemophilia</p>
<p><b>Clinical Research Questions</b></p>	<ul style="list-style-type: none"> <li>• <b>Primary objective:</b> <ul style="list-style-type: none"> <li>○ To determine the preferences of PWH for haemophilia treatments including <b>gene therapy</b>.</li> </ul> </li> <li>• <b>Secondary objectives:</b> <ul style="list-style-type: none"> <li>○ To define the relative importance and the value (preference) of various treatment attributes (e.g. differences in the frequency and modality of therapy).</li> </ul> </li> </ul>

	<ul style="list-style-type: none"> <li>○ To examine how individuals trade-off between attributes and between levels within these attributes.</li> <li>○ To identify predictors from the collected variables (e.g. demographic or clinical) that may explain differential preferences for attributes of treatment.</li> </ul>
<p><b>Methodological Research Questions</b></p>	<ul style="list-style-type: none"> <li>• This study will contribute to the Innovative Medicines Initiative (IMI) workstream and the broad objectives of this workstream, including some of the IMI Patient Preferences in Benefit and Risk Assessments during the Treatment Life Cycle (IMI PREFER) research objectives, several of which are also specified in Chapter 4 of the “Further PREFER strategy and tactics” document: <ul style="list-style-type: none"> <li>○ To investigate how patient preferences can be used in decision making</li> <li>○ To understand how patient preference studies can be conducted in rare diseases</li> <li>○ To investigate differences in preferences with varying levels of disease severity (11d)</li> <li>○ To investigate how to ensure patient preference information from preference studies is applicable/tailored to HTA/payer decision making (15b)</li> </ul> </li> </ul>
<p><b>Study Design</b></p>	<p>This study is being conducted by HCD Economics, and is sponsored by Pfizer, with input from collaborators at the University of Leuven. The UK Haemophilia Society is involved for participant recruitment. The main part of the study is the phase involving the discrete choice experiment (DCE), which is being used to elicit the preferences of PWH in the UK for gene therapy. This phase launched in December 2020 and is ongoing.</p> <p>However, this study consisted of 4 phases:</p> <p><b>Phase 1 – profile development:</b> this was a qualitative phase, involving a targeted literature review and then interviews with a group of multiple stakeholders, including patient advocates and clinicians, to identify haemophilia treatment attributes which could be used in the DCE for examining preferences for gene therapy. This phase is now complete.</p> <p>The literature review aimed to compile information on characteristics of the therapies for PWH, the attributes and levels used in previous DCE studies in haemophilia, and clinical trials of gene therapies currently underway in haemophilia. This information was used to construct the first proposal of attributes and levels, which were then explored further by gaining opinions from multiple stakeholders to assess the importance of various treatment attributes and levels.</p> <p>Semi-structured interviews were conducted with multiple patient and clinician stakeholders to identify and rank treatment attributes and levels. These were conducted with UK patient representatives (n=14) and clinical experts (n=6: 2 haematologists, 2 nurse specialists, 1 physiotherapist and 1 haemophilia psychologist). Screener questions and interview guides were designed based on: (i) the interview guide used in the IMI PREFER PAVING study and (ii) a short survey which include validated comprehension questions. Interviews were based on a structured interview guide that allows for open discussion with the participants. Interviewees answered open-ended questions and were asked to name treatment attributes which they considered important; these spontaneously mentioned attributes were then included in that individual’s subsequent ranking exercise. Individuals were then asked to rank, by importance (in their opinion) their top 6 attributes in</p>

making a decision about opting for haemophilia treatments and also for gene therapy from a predefined list of treatment attributes. Framework analysis was used to analyse the qualitative data (NVivo12) and points were allocated based upon rank of treatment attributes and sum totals calculated across interviews to indicate importance.

The key themes identified in the interviews included 'the effect of gene therapy on daily life' and the importance of 'informed decision making'. The following attributes about haemophilia treatments in general were ranked by clinician and patient stakeholders as the top 5 most important: 'effect on factor level' (79 points), 'uncertainty regarding long-term risks' (57 points), 'impact on daily life' (41 points), 'frequency of monitoring' (33 points), and 'impact on ability to participate in physical activity' (29 points). The top 5 categories ranked as the most important when considering gene therapy as a treatment option included: 'benefits' (135 points), 'quality of life' (86 points), 'risks' (85 points), 'administration' (61 points) and 'follow up' (33 points).

**Phase 2 – profile validation:** based on findings from Phase 3, five attributes and levels were further defined and validated for inclusion in the DCE, and the final DCE scenarios were produced. Variables for inclusion in the participant questionnaire were also refined. An orthogonal DCE design was chosen to define the number of scenarios to include in the study and the sample size requirements were adjusted dependent on the number of selected attributes and levels. This phase is now complete.

During this phase, five key attributes with different levels were developed and validated. These were reviewed by experts in health literacy from Northwestern University and 2 UK patient advocates with haemophilia, to ensure they were appropriate, and understandable from an English health literacy perspective. Following further adaptation of the attributes and levels during subsequent detailed consultations with expert stakeholders (a UK haemophilia consultant and a UK psychologist), a further 2 UK patient advocates also provided feedback. This led to further revision of the levels' wording to improve understanding and reduce cognitive burden.

The final five attributes selected for the DCE were:

- Therapeutic option
- Treatment effectiveness
- Safety concerns
- Hospital attendances and self-management
- Role limitations

During this process a decision was made to include an additional exploratory analysis in the DCE design, using levels from one domain of the Short- Form-Six- Dimension version 2 (SF-6Dv2) utility instrument (alongside collection of the full SF-6Dv2 responses within the questionnaire). The aim was to utilise an innovative technique to demonstrate quality adjusted life year (QALY) equivalence. This approach uses a domain from a utility instrument (SF-6Dv2) in place of one of the DCE

attributes: 'Role limitations'. The calculation element of this approach will use partial values between levels in the utility domain included in the DCE, in combination with utility tariffs (in this case the SF-6Dv2), to estimate the utility value associated with each level. In combination with the baseline utility, collected in the administration of the full SF-6Dv2 during the participant questionnaire, the utility gain achieved by a participant moving from worst to best in the DCE attribute levels can then be estimated.

A pilot study using the full study online platform (educational video, questionnaire and DCE) was then conducted with seven patient advocates in the UK, who provided feedback, including their estimates of the various levels of risk presented within the 'Safety concerns' attribute. Adjustments to the study, including further improving the wording of levels in the DCE, were made to meet any changes required.

**Phase 3 – study conduct:** this involves delivering the study (including the DCE) in an electronic format, administered to PWH in the UK online, to elicit participant preferences. This phase launched on 11<sup>th</sup> December 2020 and study participation is ongoing. Participants who meet the inclusion criteria (see Study Population) enter into the study via a secure web-link, read a participant information document and complete an online informed consent form.

Participants first watch a short, online educational video about gene therapy, which was developed by a Belgian company, MindBytes, in collaboration with the University of Leuven. This video was validated by the University of Leuven in consultation with 3 Belgian consultant haematologists who specialise in haemophilia, and 10 Belgian PWH. It also received input from Pfizer UK Medical Affairs and Health and Value colleagues and a Pfizer Global Medical Affairs colleague. It was reviewed by health literacy experts from Northwestern University and 2 UK patient advocates with haemophilia from an English health literacy perspective. This video will not collect any data and is for informing participants only.

After watching the video, participants then complete an online questionnaire to answer baseline demographic, clinical, knowledge, comprehension and satisfaction questions. Data from variables in this questionnaire will be used to examine predictors of preferences for the various treatment attributes. Data are also gathered on quality of life (utilities), captured by participants completing the SF-6Dv2. Participants then complete the DCE, where for each of 14 hypothetical scenario sets they will choose 1 of 2 scenarios, deciding between levels presented for each of the 5 attributes listed above. Descriptives for demographic, clinical and quality of life information will be performed, and analyses as detailed in 'Statistical Methods' will be conducted to estimate the value of each of the treatment attributes and levels, and to identify groups of PWH who may value treatment attributes differently.

**Phase 4 – formulating recommendations:** this will involve interpretation of the findings from the qualitative and DCE study phases and dissemination of this information. The results of this study will be comprehensively summarised in a final report. The findings will be published as research papers in peer-reviewed journals and presented as abstracts and posters at suitable conferences. An abstract of the results of Phase 1 has already been presented at ISPOR 2020 and a

	<p>draft manuscript of the results of Phases 1 and 2 is currently undergoing internal Pfizer review. The aim is for results from Phase 3 to be submitted as a manuscript by mid-2020.</p>
<p><b>Study Population</b></p>	<p>Participants must meet all of the following inclusion criteria to be eligible for inclusion in the study:</p> <ul style="list-style-type: none"> <li>• Aged 18 years or older</li> <li>• Able to read written English</li> <li>• Live in the United Kingdom</li> <li>• Diagnosed with moderate or severe haemophilia A or B</li> <li>• Have provided informed consent to participate in the study</li> </ul> <p>A minimum total of 120 people with moderate or severe hemophilia (A and B) in the UK will be recruited, aiming for recruitment of a minimum of 25 participants with haemophilia B to ensure adequate representation of this subgroup. This calculation was based on the ratio of people with haemophilia A to haemophilia B in the UK population, as per the UK National Haemophilia Database Bleeding Disorder Statistics for 2018-2019, produced by the UK Haemophilia Centre Doctors' Organisation (UKHCDO). In addition, following the completion of the phases for the development of possible attributes and levels (Phases 1 and 2), including the pilot study, a further exploration of minimum sample size requirements was performed based on sample size calculations and the number of attributes and levels. Estimates for possible parameter (coefficient) values were made to address any additional parameter requirements.</p>
<p><b>Exposure &amp; Outcome</b></p>	<p><b>Exposure:</b> This is a non-interventional study, consisting of participants watching a brief educational video on gene therapy in haemophilia, then completing a questionnaire to elicit demographic, clinical, knowledge and comprehension information, and the SF-6Dv2 to determine quality of life scores, then completing a DCE to elicit patient preferences for gene therapy in haemophilia.</p> <p><b>Outcomes:</b> The variables captured in the questionnaire, the SF-6Dv2 scores, and the participant responses in the DCE scenarios, are the data for analysis. The main endpoints of interest are:</p> <ul style="list-style-type: none"> <li>• Magnitude and direction of attribute-level coefficients (standard error [SE] &amp; 95% confidence intervals)</li> <li>• Relative importance of each level in each attribute</li> <li>• Marginal rate of substitution among attributes</li> </ul>
<p><b>Study Setting</b></p>	<p>This study involves primary data collection without any study sites. All data from the study are being collected through an electronic questionnaire and DCE completed online by PWH in the UK. Recruitment of these participants is being managed by the UK Haemophilia Society. The online link to the study platform (encompassing the participant information sheet, informed consent form, the video, the questionnaire and SF-6Dv2, and the DCE) is being distributed to PWH in the UK by</p>

	<p>the Haemophilia Society, who contact their members via email and also on social media platforms (invitation-only Facebook and WhatsApp groups operated by the Haemophilia Society).</p>
<p><b>Statistical Methods</b></p>	<p>All data processing and analyses will be performed with STATA 16.0 statistical software (STAT Corp., College Station, TX). The baseline demographics collected in this study will be summarised, alongside any clinical and quality of life information collected in the participant questionnaire and SF-6Dv2. The results captured from the DCE will be used to generate covariates for each level of each attribute in the DCE design. These attribute-level covariates will be used to assess the endpoints of interest. Potential adjustments may be made in these analyses using the baseline demographics or clinical outcomes information, dependent on sample size.</p> <p>The following objectives will be addressed in the statistical analyses:</p> <p><b>Objective 1:</b> To assess the direction and magnitude of the attribute-level coefficients to infer the preferences of PWH.</p> <ul style="list-style-type: none"> <li>• Primary Analysis</li> <li>• Endpoint: Magnitude and direction of attribute level coefficients</li> <li>• Population: Full Sample</li> <li>• Statistical Methods: Conditional Logit Model, Likelihood ratio chi-square test</li> <li>• Covariates Used: Attribute level information</li> <li>• Method of adjustment for missing values: Remove participants with missing data</li> <li>• Role of analysis: Understand the preferences of PWH for the attribute-levels identified in qualitative stage</li> </ul> <p><b>Objective 2:</b> Measure the relative importance of each attribute over the range of levels included in the experiment.</p> <ol style="list-style-type: none"> <li>1. Secondary Analysis</li> <li>2. Endpoint: Relative importance of each level within each attribute</li> <li>3. Population: Full Sample</li> <li>4. Statistical Methods: Conditional Logit Model, Conditional relative importance</li> <li>5. Covariates Used: Attribute level information</li> <li>6. Method of adjustment for missing values: Remove participants with missing data</li> <li>7. Role of analysis: Understand the importance within each attribute for PWH</li> </ol> <p>To determine the relative importance of an attribute, the difference between the attribute level with the highest preference weight and the level with the lowest preference weight will be calculated. This difference represents the maximum change in conjoint utility achievable with any attribute, given the levels chosen for the attributes in the study. The conditional relative importance of an attribute also describes the relative importance of each attribute relative to all other attributes included in the study, conditional on the range of levels of the attribute. The standard errors and the 95% confidence interval for these</p>

differences will be calculated using the delta method. Conditional relative importance estimates are usually rescaled so that the attribute with the largest conditional relative importance is set to 10 and the conditional importance of each of the other attributes is rescaled relative to that attribute.

**Objective 3:** Evaluate the marginal rate of substitution among attributes, therefore the rate at which respondents are willing to trade off among the attributes

- i. Secondary Analysis
- ii. Endpoint: Marginal rate of substitution
- iii. Population: Full Sample
- iv. Statistical Methods: Conditional Logit Model, Marginal Rate of Substitution
- v. Covariates Used: Attribute level information
- vi. Method of adjustment for missing values: Remove participants with missing data
- vii. Role of analysis: Understand the trade-off between attributes for PWH

The maximum acceptable risk (MAR) will be calculated which refers to the mean maximum level of treatment-related risk patients are willing to accept for a given improvement in benefit outcomes. It is calculated as the change in risk of a given adverse event (safety concerns) that would exactly offset the perceived benefit of a given improvement in benefit (therapeutic options, treatment effectiveness, hospital attendances and self-management, role limitations). The standard errors and the 95% confidence interval for them can be calculated using the delta method.

Analysis for the DCE will be performed using a range of regression approaches. Goodness of model fit will be determined using the Bayesian information criterion and the Akaike information criterion, where lower values indicate a better fit. Conditional logit regressions, often known as multinomial logit model, will be used to analyse responses from the DCE as it allows us to relate the probability of choice among two or more alternatives to the characteristics of the attribute levels defining those alternatives. The result of this analysis generates a preference weight which represents the relative contribution of the attribute level to the utility that respondents assign to an alternative. For conditional logit analysis, the functional form is specified as:

$$U_{isj} = \beta x_{isj} + \epsilon_{isj}$$

which represents the utility of option j in choice set s for survey respondent i, where  $x_{isj}$  is a vector of dummy variables representing the levels of the health state presented in option j,  $\beta$  is a vector of utility weights associated with each level and  $\epsilon_{isj}$  is the error term. Mixed logit model will be implemented to evaluate preference heterogeneity among respondents:

$$U_{isj} = (\beta + \eta_i) x_{isj} + \epsilon_{isj}$$



	<p>Where <math>\beta</math> represents population mean preferences and <math>\eta_i</math> is the individual deviation around those mean preferences.</p> <p>The mixed logit model, or random-parameters logit (RPL) model, is a logit model for which the parameters are assumed to vary from one individual to another. It is therefore a model that takes the heterogeneity of the population into account. The RPL model relates respondents' treatment choices to the attribute levels of each treatment profile in the choice questions. The RPL model mitigates potential estimation bias in the mean preference weight estimates because of unobserved preference heterogeneity among respondents by estimating a distribution around each mean preference parameter and accounts for the fact that each respondent made multiple treatment choices over a series (panel) of choice questions. In the RPL models, preference heterogeneity was assumed to follow a continuous normal distribution.</p> <p>Subgroup analyses will be conducted using the RPL model in order to examine systematic differences in preferences between subgroups. Subgroup analyses may or may not provide much information about the correlation between preference heterogeneity and observed characteristics of the respondents. To further explore preference heterogeneity and differences in preferences among respondents with different characteristics, latent class modeling can be considered. Here the researcher assumes a discrete, rather than a continuous, mixing distribution to describe preference heterogeneity among respondents in the sample.</p> <p>A cluster analysis, a form of latent class analysis, will be carried out to identify groups of PWH who may value treatment attributes differently. Due to limitations with the conditional logit model such as scale and preference heterogeneity, other exploratory modelling methods may be explored that overcome these possible limitations. However, this will not be the pre-specified primary outcome due to relatively small sample size and ability to converge.</p> <p>The goodness of fit of regression models will be assessed using the likelihood ratio chi-square test to determine whether including attribute-level variables significantly improves the model fit compared with a model without any attribute-level variables and indicates whether one or more of the preference weights are expected to be different from 0.</p>		
<b>Results</b>	<b>PREFER Research Questions</b>	<b>How was it addressed in this preference study</b>	<b>What was found</b>
	To investigate how patient preferences can be used in decision making	Conduct of a DCE with planned analysis of the relative contributions of different attributes and levels to patient decision making in PWH	The study is still recruiting participants and analysis has not yet commenced so results are not yet available (to be confirmed, TBC)

	<p>To understand how patient preference studies can be conducted in rare diseases</p>	<p>Conduct of a DCE in PWH; haemophilia is a rare disease</p>	<p>TBC</p>
	<p>To investigate differences in preferences with varying levels of disease severity (11d)</p>	<p>Planned analyses for patient preferences elicited in the DCE:</p> <ol style="list-style-type: none"> <li>1. Subgroup analyses using the RPL model to examine systematic differences in preferences between subgroups (including haemophilia severity subgroups).</li> <li>2. Cluster analysis to identify groups of PWH (e.g. with differing severities) who may value treatment attributes differently.</li> <li>3. Latent class modelling to explore preference heterogeneity and differences in preferences among respondents with different characteristics (including severity).</li> </ol>	<p>TBC</p>
	<p>To investigate how to ensure patient preference information from preference studies is applicable/tailored to HTA/payer decision making (15b)</p>	<p>Planned presentation of patient preference study design at the Pfizer haemophilia B gene therapy NICE scientific advice meeting (11 February 2021). This will aim to elicit feedback from NICE on how patient preference studies will be valued by NICE and how results of patient preference studies such as this one should be included in, and appraised by, future HTA or other approval submissions.</p>	<p>TBC</p>
<p><b>Conclusions</b></p>	<p><b>Clinical:</b> So far, results are only available from the initial, qualitative phases of the study (Phases 1 and 2). An abstract of the results of Phase 1 has already been presented at ISPOR 2020 and a draft manuscript of the results of Phases 1 and 2 is currently undergoing internal Pfizer review.</p> <p>In Phase 1, the key themes identified in the interviews included ‘the effect of gene therapy on daily life’ and the importance of ‘informed decision making’. The following attributes about haemophilia treatments in general were ranked by clinician and patient stakeholders as the top 5 most important: ‘effect on factor level’ (79 points), ‘uncertainty regarding long-term risks’ (57 points), ‘impact on daily life’ (41 points), ‘frequency of monitoring’ (33 points), and ‘impact on ability to participate in</p>		

	<p>physical activity' (29 points). The top 5 categories ranked as the most important when considering gene therapy as a treatment option included: 'benefits' (135 points), 'quality of life' (86 points), 'risks' (85 points), 'administration' (61 points) and 'follow up' (33 points).</p> <p>These results demonstrate that treatment benefits, treatment risks and quality-of-life impact are important factors to inform treatment selection and decision-making when patients are considering whether to proceed with gene therapy for haemophilia. These factors may also be useful to highlight to regulators and HTA bodies (such as NICE) in future discussions and submissions about gene therapy for PWH, particularly in the UK.</p> <p><b>Methodological:</b> Phases 1 and 2 were essential in defining the five attributes to be used in the DCE (Therapeutic option; Treatment effectiveness; Safety concerns; Hospital attendances and self-management; Role limitations). As the DCE study phase is ongoing, more results will be added following completion of this study and data analysis, by the end of April 2021.</p>
<p><b>Limitations</b></p>	<p>As choice data in DCEs are collected by using health profiles based on hypothetical alternatives, some possible attribute-level combinations could be implausible or illogical. Participants may have difficulty evaluating such illogical combinations, which could increase the potential for hypothetical bias, unobserved, heterogeneous interpretations by participants; or lower response efficiency. The design approach of the DCE was therefore amended to specify combinations that should not appear in the DCE. In addition, poorly defined attributes and levels would provide biased and potentially meaningless interpretations of choices and preferences. As such, the various study phases tried to ensure the reliability and validity of choosing attributes for the DCE.</p> <p>There are many methodological approaches that can be implemented to inform the choice of DCE attributes, including literature reviews, expert opinion, meta-ethnography and use of previous studies. Qualitative research involving stakeholder semi-structured interviews in Phase 1 was important in the development of relevant attributes for the DCE. However, qualitative research tends to be explorative and expansive, whereas with designing DCE a reductive process is necessary. As such, it was not possible to include all potential attributes which could be relevant, but rather infer which attributes are preferred.</p> <p>A DCE is a stated preference study and some patients may find the task cognitively challenging. The number of possible attributes presented was limited to include only the most important attributes for PWH based on the targeted literature review and the qualitative research undertaken. The DCE was also reviewed by 2 different sets of patient advocates to gain feedback during the attribute/level refinement process. The initial attributes and levels arising from the qualitative study phase were validated in an independent manner by Northwestern University, who also consulted 2 patient advocates in the UK to ensure that the health literacy of these attributes and levels was appropriate for participants. Following further adaptation of the attributes and levels during subsequent detailed consultations with further expert stakeholders (a UK haemophilia consultant and a UK psychologist), a further 2 patient advocates in the UK provided feedback. This led to revision of the levels' wording to improve understanding and reduce cognitive burden. The DCE was then tested via a pilot study in a group of 7 patient</p>

	<p>advocates in the UK. Feedback from the pilot study was used to improve further the wording of some of the attribute levels, before finally launching the study in the UK PWH population in December 2020.</p> <p>This study includes an exploratory analysis, aiming to infer QALY equivalence with the use of a domain from the SF-6Dv2 included as one of the attributes within the DCE (Role limitations). This exploratory analysis has, to date, not been reported in a peer reviewed publication, nor in a published HTA, therefore the validity of this approach is still to be established. This limitation does not impact the interpretation of attribute 5 (role limitations) within the overall DCE study.</p>
<b>Potential Use of Results in MPLC Decision-Making</b>	<p><b>Patients and/or caregivers:</b> For PWH and their caregivers, who may be faced with the decision on whether or not to choose gene therapy for their haemophilia, this study provides a means for preferences to be elicited and key attributes that determine choice to be explored.</p> <p><b>Industry:</b> The results of this study will be helpful to industry in informing development decisions about gene therapy for haemophilia. The preferences expressed in the DCE may provide guidance on what benefit-risk trade-offs may be acceptable, and to whom, including what kind of patient-related factors (demographic or clinical) may influence these decisions.</p> <p>The DCE results may contribute to regulatory guidance on use of preferences or specific methods in decision making. In addition, these results may help to inform reviews of gene therapy in HTA or other submissions, for approval or reimbursement decisions.</p>

<b>Therapeutic Area</b>	<b>Myocardial Infarction</b>
<b>Core Case Study, Academic Case Study, Industry Case Study</b>	<b>Industry Case Study</b>
<b>Current status</b>	<b>This case study is finalized</b>
<b>Academic Lead, Industry Lead, Methods Lead, Clinical Lead</b>	<b><u>Industry Lead:</u> Cathy Anne Pinto; <u>Academic Lead:</u> N/A; <u>Methods Lead:</u> Tommi Tervonen (Evidera); <u>Clinical Lead:</u> internal therapeutic area consultation only</b>
<b>Title of the Study</b>	Comparing Patient Preferences for Antithrombotic Treatments for Patients Following an Acute Coronary Syndrome and Patients with Chronic Stable Disease
<b>Background/ Study Rationale</b>	Past studies have assessed patient preferences for antithrombotic treatments in the secondary prevention of cardiovascular disease (CVD) in terms of their benefit and risk characteristics, such as reduction of mortality probability and increases in the probability of non-fatal bleedings. These preferences may differ based on patient characteristics. Currently, no data exists on how patient preferences differ based on disease stage (acute or chronic), and little is known regarding other clinical and demographic factors that may cause preference heterogeneity among patients who have previously experienced a myocardial infarction (MI).
<b>MPLC Decision-making</b>	<b><u>Main stakeholder (industry, regulatory, HTA):</u></b> industry
	<b><u>MPLC decision point of interest:</u></b> clinical development (optimize and prioritize assets)
<b>Clinical Research Questions</b>	How generalizable preferences are from one stage of disease to another?
	Is there preference heterogeneity based on other demographic and clinical characteristics?
<b>Methodological Research Questions</b>	<u>Question 11j.</u> To what degree do preference results in a disease vary with longitudinal sampling: General population, treatment naïve, shortly after treatment, after chronic treatment, etc.?
	<u>Question 11a.</u> To what degree do preference results in a disease vary with characteristics of patients?
	<u>Question 2d1.</u> How similar are the results from different methods applied with the same set of attributes on the same population? How similar are the results of simpler / faster / cheaper methods vs. more rigorous / in depth / expensive methods?

	<u>Question 10b2</u> . Can the measurement of psychological constructs including health literacy and numeracy provide insight regarding individual preferences?
<b>Study Design</b>	This study was divided into three phases: (i) qualitative pilot; (ii) quantitative pilot; and (iii) main survey. The primary instrument of preference elicitation in the survey was a discrete choice experiment (DCE) task, and the secondary instrument was a best-worst scaling type 1 (BWS-1) task.
	<u>Educational materials</u> : Written materials only
	<u>Psychosocial constructs</u> : <ul style="list-style-type: none"> <li>• Health literacy: Chew Health literacy screening questions</li> <li>• Numeracy: based on Lipkus expanded numeracy scale items</li> </ul>
<b>Study Population</b>	<p>This study involved adults (<math>\geq 18</math> years of age) who have experienced their most recent MI, either in the past year (<math>\leq 365</math> days, acute stratum) or more than a year (<math>&gt; 365</math> days, chronic stratum) ago.</p> <p>Patients meeting the following criteria were enrolled in the study:</p> <ul style="list-style-type: none"> <li>• At least 18 years of age</li> <li>• Previously hospitalized with MI: <ul style="list-style-type: none"> <li>o within a year of enrolment (acute MI group)</li> <li>o more than one year prior to enrolment (chronic MI group)</li> </ul> </li> <li>• Able to read and understand English</li> <li>• Willing and able to complete an online survey</li> <li>• Willing and able to provide (electronic) consent to participate in study</li> <li>• Resident of England</li> <li>• For qualitative pilot only: <ul style="list-style-type: none"> <li>o Willing and able to participate in a telephone interview, and be audio recorded</li> </ul> </li> </ul> <p>Patients who met the following criteria were excluded:</p> <ul style="list-style-type: none"> <li>• Have a cognitive impairment, hearing difficulty, visual impairment, acute psychopathology, or insufficient knowledge of English that—in the opinion of the investigator/interviewer—could interfere with a patient’s ability to provide written consent and complete an interview or survey.</li> <li>• Are pharmaceutical employees and those employed in a position where they have a direct role in treating patients with an MI.</li> </ul> <p>In this study, 10 patients with chronic MI participated in the qualitative pilot while an additional 40 patients with chronic MI participated in the quantitative pilot. In the main survey, a total of 335 patients were recruited, of which n=180 were patients with chronic MI and n=155 were patients with acute MI.</p>
<b>Exposure &amp; Outcome</b>	This study assessed the preferences of MI patients for antithrombotic treatment attributes using cross-sectional survey methodology. As such, there was no assignment of patients to a therapeutic strategy or use of any diagnostic or monitoring process for participation in or during the study. The primary outcome obtained from the DCE and BWS was

	the estimated marginal utilities of the treatment attributes and levels. Subsequently, the relative attribute importance (RAI) was computed from the estimated marginal utilities as the secondary outcome measure to facilitate comparison between DCE and BWS. In addition, estimated marginal utilities from the DCE were used to calculate the maximum acceptable risk (MAR) of cardiovascular death that patients were willing to accept for a reduction of other risks.		
<b>Study Setting</b>	Patients with chronic MI were recruited from online patient panels for the qualitative and quantitative pilot phase between February-May 2018 (qualitative pilot) and July 2018 (quantitative pilot) respectively. For the main survey, patients with chronic MI were recruited from online patient panels; while patients with acute MI were recruited by clinicians and research staff at five participating UK National Health Service (NHS) clinical sites. The NHS clinical sites involved in the study were: Kent Community Health NHS Trust, University Hospitals Plymouth NHS Trust, Chesterfield Royal Hospital NHS Foundation Trust and Imperial College Healthcare NHS Trust. Recruitment for patients with chronic MI for the main survey occurred in February 2019 and recruitment of patients with acute MI from the NHS sites took place between February 2019 and May 2020. Ethical approval for this study was received from the research ethics committee (REC) and health research authority (HRA) on 25 January 2018. Findings from the pilot studies were used to update the survey instrument. An approval for the amendments was sought from the REC/HRA, with approval for the main survey data collection received on 10 December 2018.		
<b>Statistical Methods</b>	Descriptive statistics were used to summarize patients' sociodemographic and clinical characteristics. DCE responses collected from the survey were analyzed using discrete choice models based on random utility maximization (RUM) theory. Multinomial logit (MNL) was estimated for the combined sample (both acute and chronic MI patients) and for the acute and chronic patient subgroups. In addition, mixed logit (MXL), latent class (LC), and heteroscedastic logit (HMNL) models were estimated separately for the combined sample. Responses from the BWS were analyzed using MNL model.		
<b>Results</b>	<b>PREFER Clinical and Methodological Research Questions</b>	<b>How was it addressed in this preference study</b>	<b>What was found</b>
	<b>Clinical Questions:</b>		
	Overlap with methodologic questions 11j and 11a.	see below	see below
	<b>Methodological Questions:</b>		
	<u>Question 11j</u> . To what degree do preference results in a disease vary	In order to assess the effect of stage of MI in patients' preferences for anti-thrombotics, this study administered a DCE survey to two independent group of patients who differed in their duration since most	In order to test the presence of any systematic difference in preferences between patients with acute and chronic MI, interaction effects between MI duration

	<p>with longitudinal sampling: General population, treatment naïve, shortly after treatment, after chronic treatment, etc.?</p>	<p>recent MI (i.e. acute MI patients defined as those with their most recent MI within the year (<math>\leq 365</math> days) prior to study enrolment and chronic MI patients defined as those enrolled more than one year (<math>&gt; 365</math> days) after experiencing their most recent MI).</p>	<p>(acute vs chronic) and the attributes were included in the MNL model. This study demonstrated that patients' preferences for anti-thrombotic treatment are similar between patients with acute and chronic. Patients with acute and chronic MI valued reduction in risk of CV death (<math>\beta_{Acute}=0.156</math> vs. <math>\beta_{Chronic}=0.156-0.021=0.135</math>, <math>p\text{-value}=0.319</math>), heart attack (<math>\beta_{Acute}=0.078</math> vs. <math>\beta_{Chronic}=0.078+0.009=0.087</math>, <math>p\text{-value}=0.552</math>), stroke (<math>\beta_{Acute}=0.071</math> vs. <math>\beta_{Chronic}=0.071-0.020=0.051</math>, <math>p\text{-value}=0.660</math>), bleeding within the skull (<math>\beta_{Acute}=0.222</math> vs. <math>\beta_{Chronic}=0.222-0.039=0.183</math>, <math>p\text{-value}=0.205</math>), and other severe bleeding (<math>\beta_{Acute}=0.142</math> vs. <math>\beta_{Chronic}=0.142+0.020=0.162</math>, <math>p\text{-value}=0.463</math>) similarly.</p>								
	<p><u>Question 11a</u>. To what degree do preference results in a disease vary with characteristics of patients?</p>	<p>The effect of individual characteristics on patients' preference for anti-thrombotic was explored using subgroup analysis for the following characteristics:</p> <table border="1" data-bbox="882 967 1543 1431"> <thead> <tr> <th>Characteristics</th> <th>Category</th> </tr> </thead> <tbody> <tr> <td>Disease stage</td> <td> <ul style="list-style-type: none"> <li>chronic,</li> <li>acute 0–6 months since M or acute 6–12 months since MI</li> </ul> </td> </tr> <tr> <td>Bleeding risk factors (low body weight, prior use of antithrombotic medications, prior CVD)</td> <td> <ul style="list-style-type: none"> <li>no risk factor</li> <li><math>\geq 1</math> risk factors</li> </ul> </td> </tr> <tr> <td>Prior health outcomes of interest, other than MI (stroke, ICH, other</td> <td> <ul style="list-style-type: none"> <li>no prior outcome</li> <li><math>\geq 1</math> prior outcomes</li> </ul> </td> </tr> </tbody> </table>	Characteristics	Category	Disease stage	<ul style="list-style-type: none"> <li>chronic,</li> <li>acute 0–6 months since M or acute 6–12 months since MI</li> </ul>	Bleeding risk factors (low body weight, prior use of antithrombotic medications, prior CVD)	<ul style="list-style-type: none"> <li>no risk factor</li> <li><math>\geq 1</math> risk factors</li> </ul>	Prior health outcomes of interest, other than MI (stroke, ICH, other	<ul style="list-style-type: none"> <li>no prior outcome</li> <li><math>\geq 1</math> prior outcomes</li> </ul>	<p>Patients who were 65 years old and above valued reduction in risk of heart attack more than patients who are below 65 years old (<math>\beta_{Above\ 65}=0.066+0.032=0.097</math> vs. <math>\beta_{Below\ 65}=0.066</math>, <math>p\text{-value}=0.035</math>). Meanwhile, patients without any bleeding risk factor valued reduction in risk of cardiovascular death (<math>\beta_{No\ bleeding\ risk\ factor}=0.184</math> vs. <math>\beta_{Bleeding\ risk\ factor\ \geq 1}=0.184+(-0.060)=0.125</math>, <math>p\text{-value}=0.006</math>) and heart attack (<math>\beta_{No\ bleeding\ risk\ factor}=0.114</math> vs. <math>\beta_{Bleeding\ risk\ factor\ \geq 1}=0.114+(-0.047)=0.067</math>, <math>p\text{-value}=0.005</math>) more than patients who have at least one bleeding risk factor. On the other hand, patients who are at high risk of developing future ischaemic event (as quantified by TIMI risk score of <math>\geq 3</math>) valued risk of cardiovascular death less than patients who are at low risk (TIMI</p>
Characteristics	Category										
Disease stage	<ul style="list-style-type: none"> <li>chronic,</li> <li>acute 0–6 months since M or acute 6–12 months since MI</li> </ul>										
Bleeding risk factors (low body weight, prior use of antithrombotic medications, prior CVD)	<ul style="list-style-type: none"> <li>no risk factor</li> <li><math>\geq 1</math> risk factors</li> </ul>										
Prior health outcomes of interest, other than MI (stroke, ICH, other	<ul style="list-style-type: none"> <li>no prior outcome</li> <li><math>\geq 1</math> prior outcomes</li> </ul>										



		major bleeding/ hemorrhage)		<p>risk score <math>\leq 1</math>) (<math>\beta_{\text{TIMI score} \geq 3} = 0.149 + (-0.073) = 0.076</math> vs. <math>\beta_{\text{TIMI score} \leq 1} = 0.149</math>, p-value=0.011).</p>
		More than one MI in the past	<ul style="list-style-type: none"> <li>• no</li> <li>• yes</li> </ul>	
		Prior family member or close friend with health outcomes of interest	<ul style="list-style-type: none"> <li>• no</li> <li>• yes</li> </ul>	
		Age	<ul style="list-style-type: none"> <li>• &lt;65 years</li> <li>• <math>\geq 65</math> years</li> </ul>	
		Gender	<ul style="list-style-type: none"> <li>• male</li> <li>• female</li> </ul>	
			<ul style="list-style-type: none"> <li>• Low (TIMI score <math>\leq 1</math>)</li> <li>• Medium (TIMI score = 2)</li> <li>• High (TIMI score <math>\geq 3</math>)</li> </ul>	
	<p><u>Question 2d1</u>. How similar are the results from different methods applied with the same set of attributes on the same population? How similar are the results of simpler / faster / cheaper</p>	<p>In order to explore how different preference elicitation method may impact on the reporting of preferences, this study compared the use of DCE and BWS-1 to elicit patients' preference for anti-thrombotic.</p>	<p>The estimated coefficients from the MNL models for the DCE and BWS cannot be compared directly due to differences in scale. As such the estimated coefficients were used to compute the relative importance of attributes.</p> <p>In the DCE, the most important attribute for patients was reducing the risk of cardiovascular death (52.4%; 95% CI 48.9 - 55.9), followed by risk of heart attack (18.2%; 95% CI 15.6 - 20.8], risk of</p>	

	<p>methods vs. more rigorous / in depth / expensive methods?</p>		<p>bleeding within the skull (14.9%; 95% CI 12.7 – 17.0), risk of other form of severe bleeding (11.2%; 95% CI 9.3 – 13.0) and stroke (3.3%, 95% CI 1.1 – 5.4).</p> <p>In the BWS, patients cared most about avoiding cardiovascular death (<math>\beta=2.943</math> (SE 0.141)) followed by stroke (<math>\beta=0.163</math> (SE 0.68)), heart attack (<math>\beta=-0.489</math> (SE 0.068)), bleeding within the skull (<math>\beta=-0.492</math> (SE 0.063)) and other form of severe bleeding (<math>\beta=-2.125</math> (SE 0.085)).</p> <p>Comparison between DCE and BWS-1 showed difference in how patients valued the importance of the treatment attributes/ outcomes particular for prevention of stroke. This difference is likely to be due to the different underlying decision-making process between DCE and BWS-1 in this study. In the BWS-1, patients were asked to indicate the most and least important outcome to avoid without information on the risks of developing the outcomes. While in the DCE, they were asked to indicate their preferred treatment choice based on varying level of risks for each attribute.</p>
	<p><u>Question 10b2.</u> Can the measurement of psychological constructs including health literacy and numeracy provide insight regarding</p>	<p>In this study, the effect of patients' subjective health literacy and numeracy skills on their preferences for anti-thrombotic treatments were explored using latent class model. Following identification of different classes/groups of preferences with the latent class model, the probability of belonging to the different classes was analyzed as a function of the patients' observable characteristics as well as their subjective health literacy and numeracy skills in a MNL model.</p>	<p>Based on the DCE responses, the latent class model identified two distinct classes of preferences. Class 1 contains approximately 28% of the sample population. Patients in class 1 did not have any significant preferences for changes in risk of cardiovascular death and heart attack. Their choices were mainly driven by changes in risk of bleeding within the skull, other severe bleeding and stroke.</p>

	individual preferences?		<p>Meanwhile, approximately 72% of the sample population falls into class 2. Their choices were mainly driven by changes in risk of cardiovascular death.</p> <p>Individuals characteristics that influence class membership were explored using a MNL regression model. Only presence of bleeding risk factors and health numeracy skill was were found to significantly contribute to class allocation. Patients with at least one bleeding risk factor were more likely to fall into Class 1 compared to patients without any bleeding risk factors. In addition, patients who have inadequate level of health numeracy skills were more likely to fall into Class 1 compared to patients with adequate level of health numeracy skills.</p>
<b>Conclusions</b>	<p>This study demonstrated that patients' preferences for anti-thrombotic treatment are similar between patients with acute and chronic MI. Consistent with the findings in existing literature, there was a large variability in patients' preferences for anti-thrombotic treatment across patients which can be explained in part by their sociodemographic and clinical characteristics. This study also showed that different elicitation method may affect reporting of preferences due to different underlying decision-making process. Future research comparing DCE and BWS methods is warranted.</p>		
<b>Limitations</b>	<p>The study compared preferences between patients with acute and chronic MI using data obtained from two independent groups of patients and, with this approach, may not be able to account for the variation in individual-specific effect and assess if the same individual would report the same preference over time. Also, the study included key treatment attributes often used in regulatory B-R decision-making and, although patients were asked during the qualitative pilot to identify any additional treatment attributes that were important, no separate qualitative research was conducted to identify and formulate attributes of importance to patients in advance of this study.</p>		
<b>Potential Use of Results in MPLC Decision-Making</b>	<p>Understanding preference heterogeneity by disease stage is of particular relevance for antithrombotic drugs as they are investigated and approved for initial administration at different stages of the disease process, including acute (at the point of hospitalization) and chronic phases of disease, with the probability of future events increasing dramatically within the first several months of the acute event and plateauing in more chronic phases. Knowledge of whether acute and chronic patients have different preferences (i.e., willingness to accept higher probability of side effects in exchange for higher efficacy) can inform different decision-making processes during the medial product lifecycle. For e.g., if a pharmaceutical company simultaneously brings two products through early phases of development including one that is administered</p>		

	<p>during the acute period of hospitalization and a second during more chronic phases of disease, then differences regarding tolerability of side effects can impact go-no go decisions for one or both products as each drug enters different phases of drug development. During prioritization of assets, prescribing preferences are often assessed, but patient preferences are not a common perspective taken into consideration during these earlier phases of drug development. Challenging portfolio trade-off decisions must be made if the investment needed to initiate later phases of development affects other projects in the portfolio competing for resources. Approximately two thirds of the drug candidates entering Phase 3 will successfully advance to regulatory submission in about two and a half years. Given drugs are approved by regulators for use during more chronic and acute phases of disease, understanding differences in tolerability of patients could also inform benefit-risk decision-making processes to determine marketing authorization.</p>
--	---

<b>Therapeutic Area</b>	<b>Chronic pain associated with osteoarthritis (OA) and/or chronic low back pain (CLBP)</b>
<b>Core Case Study, Academic Study, Industry Study</b>	<b>Industry Case Study</b>
<b>Current status</b>	<b>This case study is finalized</b>
<b>Academic Lead, Industry Lead, Methods Lead, Clinical Lead</b>	<b>Industry Leads: Leo Russo (Pfizer), Kristin Bullok (Lilly), and Brett Hauber (Pfizer)</b>
<b>Title of the Study</b>	<b>Patient preference for osteoarthritis pain and chronic low back pain treatment in the United States and United Kingdom</b>
<b>Background/ Study Rationale</b>	Pfizer and Eli Lilly are developing tanezumab, a novel treatment for chronic pain. Tanezumab, a nonopioid analgesic administered subcutaneously every 8 weeks, is being investigated in difficult-to-treat patients with moderate-to-severe knee and hip OA or CLBP with inadequate treatment response —or intolerance—to prior treatment with three classes of pain treatment. If approved, tanezumab will offer a novel approach to the treatment of chronic pain because it is a long-acting, injectable treatment that differs from currently available treatments in its mechanism of action, duration of effect, and mode of administration. Tanezumab is not associated with adverse events (AEs) such as the physical dependence seen with opioids or heart attack risk seen with NSAIDs. However, the unique mechanism of action of NGF inhibition may result in tanezumab causing musculoskeletal and neurological AEs. Understanding the preferences of patients with moderate-to-severe OA or CLBP for chronic pain treatment attributes that are important to them and that differentiate tanezumab from other chronic pain treatments will help establish the place in therapy for tanezumab in these conditions. Separate preference studies were conducted in the US and United Kingdom (UK) using the same survey instrument.
<b>MPLC Decision-making</b>	<b>Main stakeholder (industry, regulatory, HTA): Regulatory</b> <b>MPLC decision-point of interest: Market Authorization</b>
<b>Clinical Research Questions</b>	The overall objective of this study was to quantify patients' preferences for attributes of pharmaceutical treatments for chronic moderate-to-severe musculoskeletal pain associated with OA and/or CLBP that are both relevant to patients and that differentiate tanezumab from alternative classes of analgesics (including NSAIDs and opioids), and other NGF-inhibitor products, and to quantify both the relative importance of each of these treatment attributes to patients and the tradeoffs patients are willing to make among these attributes.  A secondary objective of this study was to explore heterogeneity in preferences for nine subgroups. Preferences were analyzed for prespecified subgroups of interest to identify systematic differences in the preferences of different types of respondents.

<p><b>Methodological Research Questions</b></p>	<p>How similar are the results from different methods applied with the same set of attributes on the same population? (IMI-PREFER Question 2d)</p>
	<p>How generalizable are preferences from one specific population in a disease to different populations in that disease or related diseases? (IMI-PREFER Question 9b)</p>
	<p>What tests assess whether patients can perform a given set of cognitive tasks? (IMI PREFER Question 4b1)</p>
	<ul style="list-style-type: none"> <li>• Are there certain properties of a preference problem that indicate a certain method works better than others? (IMI-PREFER Question 5b1)</li> </ul>
	<p>Can the measurement of psychosocial constructs including PREFER class I constructs (health literacy and numeracy, patient activation, and health locus of control) and class II constructs (self-efficacy, treatment-related beliefs, illness perception, and risk propensity) provide insight regarding preference heterogeneity (i.e., how much of the heterogeneity of preference studies can be explained by different results in particular psychological instruments)? (IMI-PREFER Question 10b3)</p>
	<p>To what degree do preference results in a disease vary with “Characteristics of patients?” (Question 11a)</p>
<p><b>Study Design</b></p>	<p>Overall: The study involved the development, administration, and analysis of a stated-preference survey instrument for patients with moderate-to-severe OA only, moderate-to-severe CLBP only, and concurrent moderate-to-severe OA and CLBP in the UK to elicit preferences for treatment attributes that were previously identified as being important to patients and that potentially differentiate tanezumab from other analgesics. The survey included a discrete-choice experiment (DCE) to elicit patients’ preferences for a primary set of treatment attributes and an object-case best-worst scaling (BWS) exercise to elicit patients’ assessment of the relative importance of additional relevant treatment attributes that could not be included in the DCE because of the limit on the number of attributes in the DCE. The attributes in the DCE and BWS were the same for patients with OA only, CLBP only, and concurrent OA and CLBP.</p> <p>The attributes included in the DCE were symptom control, incremental treatment-related risk of severe rapidly progressive joint problems requiring total joint replacement, which was intended to capture the risk of rapidly progressive osteoarthritis (RPOA) and was described to patients using the label “risk of severe joint problems,” risk of heart attack, risk of physical dependency, mode and frequency of administration, and personal (out-of-pocket) cost per month. The three harm attributes in the DCE were chosen to capture the most severe event for each of the three treatment classes of interest (i.e., opioids, NSAIDs, and NGF inhibitors). The attributes included in the BWS were risk of moderate-to-severe constipation while taking a medicine, risk of feeling foggy and drowsy while taking a medicine, risk of having a bleeding stomach ulcer when first starting a medicine, risk of mild-to-moderate nausea and vomiting while taking a medicine, risk each year of having a heart attack because of a medicine, risk each year of severe joint problems because of a medicine, risk each year of becoming physically dependent on a prescription pain medicine, risk of having a tingling or burning sensation in the fingers or toes while taking a medicine, risk of mild-to-moderate swelling in the ankles and feet while taking a medicine, and risk each year of having a moderate stroke because of a medicine.</p>

	<p>In addition to the DCE and BWS questions, the survey included written descriptions of each attribute and level included in the DCE, questions designed to assess patients' comprehension of the attributes and choice tasks, questions about respondents' experiences using treatments for moderate-to-severe OA pain and CLBP, an assessment of respondent locus of control using the Multidimensional Health Locus of Control (MHLC) Scale – Form C (Wallston et al., 1994), and questions regarding respondents' demographic characteristics. The survey instruments differed between conditions (i.e., OA pain only, CLBP only, and concurrent OA and CLBP) only where necessary in the questions developed to elicit disease and treatment experience.</p>
	<p>Educational materials: None</p>
	<p>Psychosocial constructs: Multidimensional Health Locus of Control (MHLC) Scale – Form C (Wallston et al., 1994),</p>
<p><b>Study Population</b></p>	<p>Respondents were residents of the UK and were aged 18 years or older with a self-reported physician diagnosis of hip or knee OA only, CLBP only, or concurrent OA and CLBP, each diagnosed at least 3 months ago; with self-reported moderate-to-severe pain in the hip, knee, and/or lower back, each defined as a self-assessed rating of 5 or greater on average in the past week on an 11-point numeric pain scale ranging from 0 (no pain) to 10 (worst possible pain); and who had taken or tried (1) three or more classes of pain treatment in the past 2 years; (2) two prior classes of pain treatment, either excluding NSAIDs due to NSAID contraindication or excluding opioids due to the respondent's unwillingness to take opioids; or (3) one prior class of pain treatment excluding NSAIDs due to NSAID contraindication and excluding opioids due to the respondents' unwillingness to take opioids. For respondents with concurrent OA pain and CLBP, a self-assessed pain rating of 5 or greater (of 10) was required for either OA pain or CLBP for the respondent to be eligible. Respondents were able to read and understand English and provide informed consent.</p> <p>The study goal was to have approximately one-third of respondents with pain due to OA only, approximately one-third of respondents with CLBP only, and approximately one-third of respondents with concurrent OA pain and CLBP. A second goal was that approximately half of the respondents were currently taking or had tried within the past 2 years an opioid to treat OA pain and/or CLBP. Respondents were recruited from an existing Internet panel.</p>
<p><b>Exposure &amp; Outcome</b></p>	<p>The variables captured in the online survey included demographic and disease- and treatment-experience variables, assessments of respondent comprehension of the treatment attributes and DCE questions, items included in the MHLC Scale – Form C, and responses to the DCE and BWS questions.</p>
<p><b>Study Setting</b></p>	<p>This study involved primary data collection using a self-administered online survey instrument. The data were collected using a nationwide patient panel from 21 February 2019 to 15 April 2019.</p>
<p><b>Statistical Methods</b></p>	<p>Descriptive variables and items included in the MHLC Scale – Form C were analyzed using STATA 15 (Stata Corp, College Station, TX). Random-parameters logit (RPL) models were used to analyze the DCE and BWS data from the sample. Exploratory LC analysis was also conducted on the DCE data. Data from the DCE and BWS questions were analyzed independently. Each RPL model of the DCE data yielded a set of relative preference weights for the attribute levels included in the DCE survey. The results of the RPL model of the DCE data were used to calculate the conditional relative importance of</p>

	<p>each attribute included in the DCE and the marginal rates of substitution between pairs of attributes. Each RPL model of the BWS data yielded a set of relative importance estimates for the attributes included in the BWS exercise. Analyses of the DCE and BWS data were conducted using STATA 15 (Stata Corp, College Station, TX). All quantitative estimates were presented with 95% confidence intervals (CIs).</p> <p>Sub-group comparisons were made using a joint test of significance that tested all interaction terms capturing differences between subgroups for all attribute levels. The difference tested is whether all terms are jointly significantly different from zero, commonly referred to as “statistically significantly systematically different”.</p>		
Results	PREFER Research Questions	How was it addressed in this preference study	What was found
	<p>How similar are the results from different methods applied with the same set of attributes on the same population? (IMI-PREFER Question 2d)</p>	<p>The pain preference study included both a discrete-choice experiment (DCE) and a Case 1 best-worst scaling (BWS) exercise. The DCE was the primary preference elicitation method in this study. The BWS exercise was included in this study to allow us to elicit preference information on a number of adverse events related to pain treatments that could not be accommodated in the DCE because of the effective limit on the number of attributes that we can include in the DCE.</p> <p>As noted above, the BWS exercise was not included to answer the same question as the DCE. That is, the attributes in the DCE and the attributes in the BWS were not the same. Because the full set of attributes differs between the DCE and BWS exercises, we were not able to directly compare the two methods as suggested in the research question.</p> <p>However, to provide a link between the attributes in the DCE and the attributes in the BWS, three risk attributes were included in the both exercises. Specifically, the highest level of each risk attribute included in the DCE was also included in</p>	<p><b>US study:</b></p> <p>Rank Order: The rank ordering of the conditional relative importance of the risks included in both the DCE and the BWS was consistent between the methods: 25% risk of physical dependence was the most important of these three risks. 0.5% risk of heart attack was the second most important of these risks. 4% risk of severe joint damage was the least important of these risks.</p> <p>Statistically Significant Differences in Importance:</p> <p>In the DCE, the conditional relative importance of a 25% risk of physical dependence was statistically significantly greater than the conditional relative importance of a 0.5% risk of heart attack; however, the conditional relative importance of a 0.5% risk of heart attack was not statistically significantly greater than the conditional relative importance of a 4% risk of severe joint damage. In contrast, in the BWS, the relative importance of a 25% risk of physical dependence was not statistically significantly greater than the conditional relative importance of a 0.5% risk of heart attack, but the conditional relative importance of a 0.5% risk of heart attack was statistically significantly greater</p>



		<p>the BWS exercise (4% risk of severe joint damage, 25% risk of physical dependence, and 0.5% risk of heart attack).</p> <p>Therefore, we were able to compare the rank ordering of these three attributes between the two exercises and assess qualitatively the consistency of the conditional relative importance estimates for these attributes between the two exercises.</p>	<p>than the conditional relative importance of a 4% risk of severe joint damage.</p> <p>Ratios of Relative Importance:</p> <p>In addition, the ratios among these three risks differed substantially between the two exercises. Specifically, in the DCE (BWS) a 25% risk of physical dependence was 4.3 (1.1) times as important as a 0.5% risk of heart attack, a 25% risk of physical dependence was 6.3(1.5) times as important as a 4% risk of severe joint damage, and a 0.5% risk of heart attack was 1.5 (1.4) times as important as a 4% risk of severe joint damage.</p> <p><b>UK study:</b></p> <p>Rank Order: The rank ordering of the conditional relative importance of the risks included in both the DCE and the BWS was inconsistent between the methods. In the DCE, 25% risk of physical dependence was the most important of these three risks. 0.5% risk of heart attack was the second most important of these risks. 4% risk of severe joint damage was the least important of these risks. In the BWS, 0.5% risk of heart attack was the most important of these three risks. 25% risk of physical dependence was the second most important of these risks. 4% risk of severe joint damage was the least important of these risks.</p> <p>Statistically Significant Differences in Importance:</p> <p>As noted above, the rank order of the conditional relative importance of a 25% risk of physical dependence and a 0.5% risk of heart attack differed between the DCE and the BWS. In both exercises, the conditional relative importance estimates for 25% risk of dependence and 0.5%</p>
--	--	---	---

			<p>risk of heart attack were statistically significantly different. However, in the DCE, the conditional relative importance of a 0.5% risk of heart attack (ranked 2nd) and the conditional relative importance of a 4% risk of severe joint problems (ranked 3rd) were not statistically significantly different. In contrast, in the BWS, the conditional relative importance of a 25% risk of dependence (ranked 2nd) was statistically significantly greater than the conditional relative importance of a 4% risk of severe joint damage.</p> <p>Ratios of Relative Importance:</p> <p>As noted above, the rank order of the conditional relative importance of a 25% risk of physical dependence and a 0.5% risk of heart attack differed between the DCE and the BWS. Therefore, the ratios of the conditional relative importance estimated of these two differed between the two exercises. Specifically, in the DCE (BWS), a 25% risk of physical dependence was 7.4 (1.3) times as important as a 4% risk of severe joint damage and a 0.5% risk of heart attack was 1.6 (1.6) times as important as a 4% risk of severe joint problems.</p>
	<p>What tests assess whether patients can perform a given set of cognitive tasks? (IMI-PREFER Question 4b1)</p>	<p>The survey included 7 questions to assess respondents' comprehension in three specific areas: comprehension of the attribute descriptions, comprehension of the risk presentation, and comprehension of the DCE questions. We calculated the frequency with which respondents answered each comprehension question incorrectly and tested for whether respondents who answered three or more comprehension questions incorrectly had</p>	<p>We found statistically significantly different preference estimates between respondents who answered 3 or more comprehension questions incorrectly and respondents who answered less than 3 comprehension questions incorrectly in both the UK and the US.</p> <p>In the US, the preference estimates and conditional relative importance estimates for those respondents who answered 3 or more comprehension questions incorrectly were very similar to those for respondents who answered less than 3 comprehension questions incorrectly.</p>

		<p>statistically significantly systematically different preference estimates than those who answer less than three comprehension questions incorrectly.</p> <p>It is important to note that is a respondent answered a comprehension question incorrectly, the correct answer to the comprehension question was provided prior to the respondent proceeding with the survey. Therefore, an incorrect response to a comprehension question does not necessarily indicate that the respondent did not understand the corresponding concept when completing the remainder of the survey.</p>	<p>The primary differences between the two subgroups was that respondents who answered 3 or more comprehension questions incorrectly appear to have placed greater weight on pain control and those who answered less than 3 comprehension questions incorrectly appear to have placed slightly less weight on the risk of severe joint damage.</p> <p>In the UK, the preference estimates and conditional relative importance estimates for those respondents who answered 3 or more comprehension questions incorrectly were noticeably different than for respondents who answered less than 3 comprehension questions incorrectly. Respondents who answered fewer than 3 comprehension questions incorrectly placed noticeably greater weight on pain control and appear to be less averse to heart attack risk than respondents who answered 3 or more comprehension questions correctly.</p>
	<p>Are there certain properties of a preference problem that indicate a certain method works better than others? (IMI-PREFER Question 5b1)</p>	<p>This question was not be addressed directly in this study. However, we had a strong rationale for choosing the methods employed in this study and based this choice on the ability of the method to address the research questions.</p>	<p>We effectively approached this study with two distinct types of research questions in mind. Our primary objective was to estimate the tradeoffs that patients were willing to make between the benefits, risks, and mode and frequency of administration of treatment. The treatments under consideration all are (or are intended to be) indicated for the same conditions and have the same primary objectives – to reduce pain and improve function. However, the three types of treatment under consideration have substantially different mechanisms of action and each has different adverse events that are of concern to patients, physicians, and regulators. Therefore, the willingness of patients to tolerate these different risks in order to</p>

			<p>achieve the same outcome is important. In addition, these treatments vary substantially in their mode and frequency of administration and it is unknown whether patients would be willing to accept additional risk or decreases in expected benefit in order to have the mode and frequency of administration they prefer. Therefore, the research question is one in which decision makers are essentially asked to trade off multiple attributes simultaneously. Therefore, we believe that this question is best addressed using a discrete-choice experiment.</p> <p>The second objective of this research was to understand where multiple other potential risks fit into patient decision making relative to the major or key risks that were included in the DCE. Because there were many potential risks associated with each of these types of products, conducting a DCE that included all of these risks likely would not have been feasible. Therefore, we chose to use a method that could accommodate a larger number of risks and provide estimates of relative importance, namely a Case 1 best-worst scaling (BWS) exercise.</p>
	<p>How generalizable are preferences from one specific population in a disease to different populations in that disease or related diseases? (IMI-PREFER Question 9b)</p>	<p>The sample included patients with related diseases—osteoarthritis only, chronic low back pain, and comorbid osteoarthritis and chronic low back pain. The sample was stratified to include nearly equal representation across all three categories to facilitate subgroup analyses. The results for each of these subgroups were compared. In addition, the sample included patients from the United States and the United Kingdom. Respondents from these different countries were</p>	<p>In each country, we used the same survey to elicit patient preferences for all three conditions. We believe that this was appropriate because the goals of treatment are the same for all conditions as are the available treatment options.</p> <p>In the US, subgroup analysis of the DCE data did not detect statistically significantly systematic differences in preferences among the three conditions. Latent class analysis showed a marginally insignificant (<math>P = 0.055</math>) difference in preferences for patients with chronic low back</p>

		<p>recruited separately using the same inclusion criteria.</p>	<p>pain only. Subgroup analysis of the BWS data did not detect statistically significantly systematic differences in preferences among the three conditions.</p> <p>In the UK, subgroup analysis of the DCE data did not detect statistically significantly systematic differences in preferences among the three conditions. Latent class analysis showed a marginally insignificant (P = 0.065) difference in preferences for patients with chronic low back pain only. Subgroup analysis of the BWS data showed statistically significantly systematic differences in preferences among the three conditions.</p> <p>The results of the DCE suggest that treatment preferences are mostly generalizable across conditions when the goal of treatment and the treatment options under consideration are the same; however, there is weak evidence that there may be systematic differences in preferences based on condition that were not captured in the DCE. In contrast, evidence from the UK suggests that preferences elicited using the Case 1 BWS are not generalizable across conditions.</p>
	<p>Can the measurement of psychosocial constructs including IMI-PREFER Class I constructs (health literacy and numeracy, patient activation, and health locus of control) and Class II constructs (self-efficacy, treatment-related beliefs, illness perception, and risk propensity) provide insight</p>	<p>The survey instrument includes the Multidimensional Health Locus of Control (MHLC) Scale – Form C. Subgroup analysis was conducted in which mutually exclusive subgroup pairs are defined in three ways:</p> <ul style="list-style-type: none"> <li>• Respondents who are classified as high “internal” locus of control and patients who are classified as low “internal” locus of control (the median score will be used to split the sample in those two subgroups)</li> </ul>	<p>In the US, subgroup analysis of the DCE data did not detect statistically significantly systematic differences in preferences based on locus of control. Latent class analysis showed no difference in preferences between patients with different levels of locus of control for any of the three types of locus of control. Subgroup analysis of the BWS data showed statistically significantly systematic differences in preferences between patients with high and low</p>

	<p>regarding preference heterogeneity (i.e., how much of the heterogeneity of preference studies can be explained by different results in particular psychological instruments)? (IMI-PREFER Question 10b3)</p>	<ul style="list-style-type: none"> <li>• Respondents who are classified as high “change” locus of control and patients who are classified as low “change” locus of control (the median score will be used to split the sample in those two subgroups)</li> <li>• Respondents who are classified as high “powerful others” locus of control and patients who are classified as low “powerful others” locus of control (the median score will be used to split the sample in those two subgroups).</li> </ul>	<p>levels of locus of control for all three types of locus of control.</p> <p>In the UK, subgroup analysis of the DCE data did not detect statistically significantly systematic differences in preferences based on locus of control; however, the difference in preferences between those with high and low “powerful others” locus of control was only marginally insignificant (P = 0.063). Latent class analysis showed no difference in preferences between patients with different levels of locus of control for any of the three types of locus of control. Subgroup analysis of the BWS data showed statistically significantly systematic differences in preferences between patients with high and low levels of locus of control for all three types of locus of control.</p>
	<p>To what degree do preference results in a disease vary with “characteristics of patients?” (IMI-PREFER Question 11a)</p>	<p>This study will include the following subgroup analyses in addition to the subgroup analyses described above:</p> <ul style="list-style-type: none"> <li>• Respondents with opioid experience in the last 2 years and those without opioid experience in the last 2 years</li> <li>• Patients younger or older than median age in the sample or patients living with chronic pain for more time and less time than the median time in the sample</li> <li>• Respondents with moderate baseline pain and respondents with severe baseline pain in the past week as self-assessed on an 11-point numeric pain scale ranging from 0 (no pain) to 10 (worst possible pain).</li> </ul>	<p>In the US DCE, only patient performance on the comprehension questions was associated with statistically significant systematic differences in preferences in subgroup analysis. In the latent class analysis, prior opioid experience was the only additional patient characteristic that had a statistically significant effect on class membership. In the subgroup analysis of the BWS data, multiple patient characteristics were found to have statistically significant systematic effects in preferences including opioid experience, age, time since diagnosis, and baseline level of pain.</p> <p>In the UK DCE, patient age and performance on the comprehension questions were associated with statistically significant systematic differences in preferences in subgroup analysis. Also, in subgroup analysis of the DCE data, systematic differences in preferences based on</p>

			<p>time since diagnosis were only marginally insignificant (<math>P = 0.053</math>). In the latent class analysis, patient age was also a statistically significant predictor of class membership. In the subgroup analysis of the BWS data, multiple patient characteristics were found to have statistically significant systematic effects in preferences including age, time since diagnosis, and baseline level of pain.</p>
<p><b>Conclusions</b></p>	<p>In general, respondents in the US &amp; UK on average preferred better symptom control, avoiding risk of treatment related physical dependency, and taking oral pills daily. Respondents were much less concerned with avoiding an incremental treatment-related risk of heart attack or severe rapidly progressive joint problems requiring total joint replacement. Patients were willing to accept some level of each type of risk to improve pain-symptom control. There is evidence of preference heterogeneity within the sample. These results suggest that patients, on average, prefer alternative pharmacological treatments for pain to opioids and that patients, on average, would view NGF-inhibitor products as acceptable alternatives to both NSAIDs and opioids, despite the risks associated with NGF-inhibitor medications.</p>		
<p><b>Limitations</b></p>	<p>Respondents of the online survey were recruited through a panel used by SSI. Participants are recruited to be panel members via partnerships with trusted loyalty programs, as well as through banner ads, pop-ups and messages on websites, television advertising, and offline recruiting (e.g., telephone recruitment of targeted populations). Patients who choose to be members of such panels may have preferences that differ from those of the overall population of patients meeting the eligibility criteria. In addition, respondents who chose to respond to the recruitment invitation may have preferences that differ from those who chose not to respond to the recruitment invitation. This may have resulted in potential volunteer bias, leading to an underestimate or overestimate of respondent satisfaction and preferences. Research has shown that results from online stated-preference studies are, in general, not statistically different from those results elicited through face-to-face interviews (Nielsen, 2011; Marta-Pedroso et al., 2007).</p> <p>The diagnosis of hip or knee OA and/or CLBP and chronic moderate-to-severe pain in the hip, knee, or lower back, as well as patient demographics and treatment history, were self-reported. The online survey assessed eligibility based on self-report to a series of screening questions.</p> <p>The study sample included diversity in terms of gender, education, age, and condition. The study sample also included diversity in terms of duration of chronic pain and prior treatment experience; however, there was a lack of diversity with regard to race and ethnicity. The majority of respondents in this sample self-identified as white or Caucasian.</p> <p>Potential information and selection bias introduced by study design may be the result of the recruitment method chosen. Our respondents may be younger and more educated than the actual patient population. Generalizability of the results may be a concern because of selection bias.</p>		

**Potential Use of  
Results in MPLC  
Decision-Making**

The tradeoffs between benefits and risks may inform regulatory (i.e., market authorization) decisions.



### 8.1.3 Prospective additional academic case studies

<b>Therapeutic Area</b>	Diabetes
<b>Core Case Study, Academic Case Study, Industry Case Study</b>	Academic Case Study
<b>Current status</b>	Data collection for this case study is complete, the final study report will be available in April 2021
<b>Academic Lead, Industry Lead, Methods Lead, Clinical Lead</b>	<u>Academic Lead</u> : Ardine de Wit, Jorien Veldwijk, Esther de Bekker <u>Methods Lead</u> : Jorien Veldwijk, Esther de Bekker
<b>Title of the Study</b>	Preferences of diabetic patients for different glucose monitoring devices
<b>Background/ Study Rationale</b>	Recent developments in glucose monitoring devices have led to preference-sensitive decisions where patients have a large variety of choices in regard to function, features, cost, and other factors to consider when choosing which glucose monitor is best for them. This study aims to determine the preferences and trade-offs of diabetes patients when selecting a device for monitoring their glucose, and to determine whether these outcomes differ by type of preference elicitation method used, the kind of educational tool they are presented with, the way patients are recruited, and patient characteristics, or experiences
<b>MPLC Decision-making</b>	<u>Main stakeholder (industry, regulatory, HTA)</u> : The primary stakeholders for this type of study are industry and HTA. Industry stakeholders can use this information to target development of new devices that meet market needs. HTA stakeholders can use this information to better calculate predicted uptake rates.
	<u>MPLC decision point of interest</u> : Pre-discovery and Post approval
<b>Clinical Research Questions</b>	Clinical research question #1 : Which attributes of blood glucose monitoring devices patients consider when deciding on their preferred device? What are the relative importance of these different attributes in choosing which devices to use?
	Clinical research question #2: What are the minimum acceptable benefits needed to justify increased costs?

<p><b>Methodological Research Questions</b></p>	<p>Methodological research question #1(2d): How do preference values vary between different preference assessment methods with varying complexity when applied to the same population sample?</p> <p>Methodological research question #2 (7d): Are preferences for devices are different (i.e. overall preference accuracy and variability of preferences) when patients are educated regarding blood glucose monitoring using an interactive educational tool vs traditional textual information?</p> <p>Methodological research question #4 (11a): Can preference heterogeneity be explained based on type of diabetes, geographic, or socioeconomic factors (e.g. country of origin, age, sex)?</p>
<p><b>Study Design</b></p>	<p>Qualitative pilot: In step 1, a scoping literature review was conducted in PubMed to identify relevant attributes of glucose-monitoring devices and develop an interview guide. In step 2, semi-structured interviews were conducted with Type 1 and 2 diabetes patients (n=19), clinicians (n=5), patient organization representatives (n=2), and pharmaceutical industry representatives involved in glucose monitoring device development (n=4). The resulting list of 12 attributes identified in steps 1 and 2 were rated and reduced according to relevance, completeness, non-redundancy, operationality, and preferential independency.</p> <p>Quantitative pilot: The draft questionnaire was pre-tested during six ‘think-aloud’ tests, checking for comprehensibility and clarity. It was then revised and pilot tested in a sample of 99 participants in order to retrieve priors to inform the design of the final DCE. After the pilot test of the DCE, the DCE design with three alternatives per choice task was substituted by a “best-best” or a so-called ‘dual response’ DCE</p> <p>Main survey and main preference elicitation survey: Participants were randomized to receive either text or animated information prior to completing the survey as well as whether to see the DCE or SW task first after the information was presented.</p> <p><u>Educational materials:</u> A script was developed based on input from clinicians and patient representatives. The script consisted of two parts: general information about diabetes and glucose monitoring, and information about the attributes. The general information section was based on research showing that many diabetes patients were never properly educated about the basics of diabetes self-management, the importance of keeping glucose under control, and how to use glucose monitoring to help this. An animated version of the text was created with a voice over of the text.</p> <p><u>Psychosocial constructs:</u> Three psychosocial constructs were assessed. These were subjective numeracy (SNS-3), health literacy (Chew’s Health Literacy screening questions), and patient activation (Patient Activation Measure; PAM). In addition to this, diabetes self-management was assessed (Diabetes Self-Management Questionnaire- Revised, DSMQ-R)</p>

<b>Study Population</b>	Adult (age $\geq$ 18) self-reporting diagnosis diabetes with no restriction on type who reside in the Netherlands (N=459 panel respondents) and Poland (N=522 panel respondents); able to give informed consent; able to verbally communicate by themselves; able to read, speak, and understand Dutch or Polish; have access to a computer with internet connection		
<b>Exposure &amp; Outcome</b>	DCE: Bayesian D-efficient designed DCE was developed, consisting of three blocks of 12 choice tasks. Each choice task was presented in a dual response design in which participants were first asked to choose between two alternative devices and then to choose between the alternative chosen or a status quo opt-out. Each alternative was described using seven attributes and 22 levels. Participants were given two 'warm-up' DCE choice-tasks before the main exercise in order to ensure comprehension. Please describe the survey methods (including attributes and levels) and outcomes (e.g. relative attribute importance, maximum acceptable risk)		
<b>Study Setting</b>	Effects coded mixed-logit models were used to analyze the DCE outcomes. SW was analyzed using relative preference scores and rank order centroid analysis. Online survey completed by patient panels respondents and patients recruited through clinical networks, patient representative organization, and social media		
<b>Statistical Methods</b>	Comparison of self-reported feedback from participants indicating how easy the methods were to understand and answer. Please describe statistical methods (max 3 or 4 sentences)		
<b>Results</b>	<b>PREFER Clinical and Methodological Research Questions</b>	<b>How was it addressed in this preference study</b>	<b>What was found</b>
	<b>Clinical Questions:</b>		
	Which attributes of blood glucose monitoring devices patients consider when deciding on their preferred device?	Effects coded mixed-logit models	For both Dutch and Polish patients, the most important attributes were monthly costs, precision compared to fingerpricking, and probability of skin irritation, respectively.
	What are the minimum acceptable benefits needed to justify increased costs?	Effects coded mixed-logit models	Overall, Dutch patients were willing-to-pay more for each attribute. However, uptake rates for the most preferred device were higher for Polish patients, particularly if they were over 50 years old or used fingerpricking exclusively.

Methodological Questions:		
How do preference values vary between different preference assessment methods with varying complexity when applied to the same population sample?	Self-reported feedback from participants indicating how easy the method was to understand and answer; comparison of attribute (relative) importance	the DCE was reported by participants to be both easier to understand and to complete, compared to the SW. For the DCE, the proportion of attribute importance is very different for all the attributes (Figure 3). Contrastingly, all attributes in the SW, received between 12-17% of the designated importance. The DCE had a 14.9-fold difference between the most and least important attribute, while the SW had a 1.4-fold difference.
<p>Are preferences for devices are different (i.e. overall preference accuracy and variability of preferences) when patients are educated regarding blood glucose monitoring using an interactive educational tool vs traditional textual information?</p> <p>Can preference heterogeneity be explained based on type of diabetes, geographic, or socioeconomic factors (e.g. country of origin, age, sex)?</p>	<p>Comparison of Effects coded mixed-logit model outcomes; comparison of time to complete survey</p> <p>Comparison of Effects coded mixed-logit model outcomes and WTP estimates</p>	<p>No statistically significant difference was found between the two educational tool types within each country. Slight differences were found in parameter estimates, but these differences did not result in different outcomes after applying the estimates to expected uptake rates. The amount of time that participants spent in the choice tasks indicates that educational material was not highly utilized.</p> <p>Significant statistical differences were found between the two countries in regards to parameter estimates. These differences did not result in different outcomes after applying the estimates to expected uptake rates. The most important attributes of a glucose-monitoring device are similar for both patients in the Netherlands and Poland; The primary difference between the two countries was in how the fingerprick status quo was valued. Further research should explore how differing national reimbursement procedures could affect patients'</p>

			choice of devices and willingness-to-pay out of pocket.
<b>Conclusions</b>	<p>Similar patterns of preferences were seen within the two countries. The primary difference was how the finger-prick was valued. Data regarding attribute ranking was similar for both DCEs and SW instruments. DCEs were considered easier to use by participants and resulted in more detailed information for stakeholders to use. Patients showed similar preferences between countries, but differences in uptake rates were found. The limited time spent in the educational tools raises questions about whether they were actually utilized or only seen as a hurdle to get to the choice tasks.</p>		
<b>Limitations</b>	<p>The data is limited to preferences reported by panel participants. Results of a preference study done in participants recruited via clinical partners may result in different findings. This study was unable to recruit via these channels due to reprioritize resources by clinical partners as a response to increased demand associated with COVID-19. Differences between the two countries regarding healthcare availability and healthcare costs may make the patient groups incomparable. Panel respondents self-reported as having diabetes with no way to verify this. Additionally, the educational levels of the two countries were different as the Polish population was more highly educated than the Dutch population.</p>		
<b>Potential Use of Results in MPLC Decision-Making (Optional)</b>	<p>Option to describe potential use of these results in MPLC decision-making</p>		

<b>Therapeutic Area</b>	<b>Rheumatoid arthritis, second-line treatment</b>
<b>Core Case Study, Academic Case Study, Industry Case Study</b>	Academic case study
<b>Current status</b>	This case study is finalized
<b>Academic Lead, Industry Lead, Methods Lead, Clinical Lead</b>	<u>Academic Lead</u> : Ulrik Kihlbom, Karin Schölin Bywall <u>Methods Lead</u> : Jorien Veldwijk
<b>Title of the Study</b>	<b>Assessing patients with rheumatoid arthritis preferences for second-line treatment using an educational tool: a discrete choice experiment in a regulatory decision-making context in Sweden</b>
<b>Background/ Study Rationale</b>	Patient preference studies needs to be tailored to the regulatory context in order to provide useful information in an understandable and accurate form to inform regulatory marketing authorisation decisions. The inclusion of patient preferences in regulatory marketing authorisation decisions is currently hindered by the lack in guidance on when and how to assess patient preferences and how to use this information. More best practice patient preference studies are critically needed to support the development of such guidance.
<b>MPLC Decision-making</b>	<u>Main stakeholder (industry, regulatory, HTA)</u> : This study may support decisions in the regulatory marketing authorisation process by providing information on what is most important for patients with rheumatoid arthritis and compares the effect of an educational tool to traditional written information on the outcomes of a DCE study. <u>MPLC decision point of interest</u> : Patient preference information assessed by this study may provide important additional information in decisions relating to a renewal of an existing drug or for an approval of a new drug (submission and validation, scientific opinion, commission decision).
<b>Clinical Research Questions</b>	To elicit RA patient preferences (including risk-benefit trade-offs and relative importance of attributes) for biologics and JAK-inhibitors when changing treatment. To what extent do patient characteristics explain preference heterogeneity amongst patients with rheumatoid arthritis in Sweden? To estimate minimum acceptable benefit of patients with rheumatoid arthritis for changing to biologics and JAK-inhibitors

<b>Methodological Research Questions</b>	To what extent does the application of an educational tool (vs standard written information) influence patient preferences as well as the outcomes of a discrete choice experiment? 1.16 (7d)																			
	To what extent do psychological constructs (health literacy and numeracy) provide insight regarding the accuracy and consistency of patient preferences? 1.19, (10b1)																			
<b>Study Design</b>	The attributes for the discrete choice experiment were developed via a scoping literature overview, validation meetings with experts, and focus groups with patients. A preliminary list of attributes and attribute levels drafted from the scoping literature overview was revised based on feedback from the validation meetings and the focus group meetings. Seven attributes were included in the discrete choice experiment: mode of administration, frequency of use, probability of mild short-term side effects, probability of side effects changing appearance, probability of psychological side effects, probability of severe side effects, and treatment effectiveness.																			
	<u>Educational materials:</u> Respondents received the same training content in one of two forms: either as a written (plain) text or as an educational tool that used graphics, pictograms, icon arrays, voice-over, and click-on functions. The educational tool was developed and illustrated with assistance from MindBytes, ( <a href="http://www.mindbytes.be">http://www.mindbytes.be</a> ).																			
	<u>Psychosocial constructs:</u> Measures of health literacy, and subjective numeracy were included.																			
<b>Study Population</b>	The following inclusion criteria were used: diagnosis of rheumatoid arthritis; between 18–80 years old; able to understand and answer the questions; and able to read and understand Swedish without aid.																			
<b>Exposure &amp; Outcome</b>	The survey started with information about rheumatoid arthritis and available treatment options before entering the DCE. The last section of the survey consisted of demographic and disease-related questions, health literacy, and numeracy. The DCE had an attribute-based D-efficient experimental design. Each respondent answered 15 hypothetical choice questions characterised by varying attribute levels. Participants were asked to choose their preferred treatment from two alternatives. Each attribute was described by three levels based on current clinical knowledge of existing biologics and JAK inhibitors (table 1).																			
	<p><b>Table 1.</b> Attributes and levels</p> <table border="1"> <thead> <tr> <th>Attribute</th> <th>Level 1</th> <th>Level 2</th> <th>Level 3</th> </tr> </thead> <tbody> <tr> <td>Route of administration</td> <td>Tablet</td> <td>Injection</td> <td>Drip</td> </tr> <tr> <td>Frequency of use</td> <td>Daily</td> <td>Weekly</td> <td>Monthly</td> </tr> <tr> <td>Probability of mild short-term side effects (nausea, vomiting or headache)</td> <td>Common, 1 in 10</td> <td>Uncommon 1 in 100</td> <td>Rare 1 in 1000</td> </tr> <tr> <td>Probability of side effects changing appearance (hair loss, weight changes or skin rash)</td> <td>Common, 1 in 10</td> <td>Uncommon 1 in 100</td> <td>Rare 1 in 1000</td> </tr> </tbody> </table>	Attribute	Level 1	Level 2	Level 3	Route of administration	Tablet	Injection	Drip	Frequency of use	Daily	Weekly	Monthly	Probability of mild short-term side effects (nausea, vomiting or headache)	Common, 1 in 10	Uncommon 1 in 100	Rare 1 in 1000	Probability of side effects changing appearance (hair loss, weight changes or skin rash)	Common, 1 in 10	Uncommon 1 in 100
Attribute	Level 1	Level 2	Level 3																	
Route of administration	Tablet	Injection	Drip																	
Frequency of use	Daily	Weekly	Monthly																	
Probability of mild short-term side effects (nausea, vomiting or headache)	Common, 1 in 10	Uncommon 1 in 100	Rare 1 in 1000																	
Probability of side effects changing appearance (hair loss, weight changes or skin rash)	Common, 1 in 10	Uncommon 1 in 100	Rare 1 in 1000																	

	<p>Probability of psychological side effects (anxiety, mood changes, depression or sleep disturbance)</p> <p>Probability of severe side effects that requires hospitalisation such as severe infections or allergic reactions</p> <p>Effectiveness (the ability to decrease inflammation and swelling of the joints, also pain and other symptoms)</p>	<p>Common, 1 in 10</p> <p>Common, 1 in 10</p> <p>30 % improvement So out of 100 persons taking the treatment, 30 will get enough improvement, the rest will get a small or no improvement</p>	<p>Uncommon 1 in 100</p> <p>Uncommon 1 in 100</p> <p>50 % improvement So out of 100 persons taking the treatment, 50 will get enough improvement, the rest will get a small or no improvement</p>	<p>Rare 1 in 1000</p> <p>Rare 1 in 1000</p> <p>70 % improvement So out of 100 persons taking the treatment, 70 will get enough improvement, the rest will get a small or no improvement</p>
	Outcomes calculated were: attribute level estimates, relative importance scores, minimum acceptable benefit, preference heterogeneity.			
<b>Study Setting</b>	Recruitment of patients with rheumatoid arthritis started in November 2018 and ended in October 2019. In total, 673 respondents were included in the analysis. Patients were recruited via three sources: a research panel (n=162) (dynata.com); the Swedish Rheumatism Association (n=228); and the rheumatology clinic at Uppsala University Hospital, Sweden (n=283).			
<b>Statistical Methods</b>	<p>Assessing preferences: Latent class analysis models were used for the analysis of the DCE data. Such models account for the multilevel structure of the data (i.e., every respondent answered multiple choice questions) and allow for the investigation of preference heterogeneity.</p> <p>Minimum acceptable benefit was estimated as the difference between the preference weights (parameters) for two levels of an attribute divided by the preference weight.</p> <p>Assessing impact of educational tool: Random parameter logit models were used to estimate attribute levels and heterogeneity in preferences</p> <p>Relative importance scores were calculated based on results of a random parameter logit model separately for the plain text and the educational tool. The difference between the highest and lowest estimates of the attribute level was calculated for each attribute.</p>			
<b>Results</b>	<b>PREFER Clinical and Methodological Research Questions</b>	<b>How was it addressed in this preference study</b>	<b>What was found</b>	
	<b>Clinical Questions:</b>			



	<p>To elicit RA patient preferences (including risk-benefit trade-offs and relative importance of attributes) for biologics and JAK-inhibitors when changing treatment.</p>	<p>Attribute level estimates Relative importance of the attributes Trade-offs between benefits and risks</p>	<p>Latent class analysis revealed three preference patterns. When choosing treatment, the respondents found either effectiveness of treatment (34%), mode of administration (28%), or probability of severe side effects to be most important (38%).</p>
	<p>To what extent do patient characteristics explain preference heterogeneity amongst patients with rheumatoid arthritis in Sweden?</p>	<p>Latent class analysis &gt; class assignment model</p>	<p>Disease duration and experience in mild side effects influenced preferences.</p>
	<p>To estimate minimum acceptable benefit of patients with rheumatoid arthritis for changing to biologics and JAK-inhibitors</p>	<p>Minimum acceptable benefit</p>	<p>Patients noting effectiveness as most important were more willing than other patients to accept higher risks of side effects</p>
<p><b>Methodological Questions:</b></p>			
	<p>To what extent does the application of an educational tool (vs standard written information) influence patient preferences as well as the outcomes of a discrete choice experiment?</p>	<p>Compared educational tool to written information:</p> <ul style="list-style-type: none"> <li>• Scale &amp; choice consistency</li> <li>• Dominant decision making.</li> <li>• Time of completion</li> <li>• Drop-out</li> <li>• Perceived understanding and difficulty of the questionnaire</li> </ul> <p>Compared the outcomes of a DCE when using either an educational tool or standard written information</p> <ul style="list-style-type: none"> <li>• Attribute levels estimates</li> <li>• Relative importance of the attributes</li> </ul>	<p>The outcome of the scale test showed that the estimates, but not the error variances, were significantly different across the two data sets.</p> <p>There were no differences in reported difficulty or the time taken to answer the DCE questions between the study arms.</p> <p>Respondents informed by the educational tool placed relatively more importance on treatment side effects, and respondents receiving the plain text placed relatively more importance</p>

		<ul style="list-style-type: none"> <li>Trade-offs between benefits and risks</li> </ul>	on the treatment effectiveness and the administration methods.
	To what extent do psychological constructs (health literacy and numeracy) provide insight regarding the accuracy and consistency of patient preferences?	<p>The effect of health literacy and numeracy on preferences and understanding of the survey instrument will be assessed by comparing:</p> <ul style="list-style-type: none"> <li>Responses to warm up questions</li> <li>Scale &amp; choice consistency Survey internal validity checks</li> <li>Perceived understanding and difficulty of the questionnaire</li> </ul>	There were no identified effects of health literacy and numeracy.
<b>Conclusions</b>	<p>On average, respondents found either effectiveness, severe side effects, or mode of administration as the most important attribute. Patients noting effectiveness as most important were more willing than other patients to accept higher risks of side effects.</p> <p>Compared to the respondents receiving the plain text, the respondents receiving the educational tool placed relatively more importance on the side effects. However, uncertainty remains regarding how educational tools influence preferences and the results of a DCE. Further research is needed to provide guidance on how and when to use educational tools to inform and elicit patients' preferences.</p>		
<b>Limitations</b>	<p>A limitation of this study relates to the use of three different recruitment strategies. This was a pragmatic choice to facilitate achievement of a sufficient number of respondents in order to power the statistical analysis. However, a sensitivity analysis did not identify any systematic differences in respondents' preferences for the attribute levels depending on recruitment source.</p>		
<b>Potential Use of Results in MPLC Decision-Making</b>	<p>Information on the relative importance of treatment characteristics and preference, heterogeneity within patients' preferences and the minimum acceptable benefit may provide essential information in the regulatory scientific evaluation.</p>		

<b>Therapeutic Area</b>	Multiple Myeloma (MM)
<b>Core Case Study, Academic Study, Industry Study</b>	Academic Study
<b>Current study</b>	This case study will be finalized in April 2021
<b>Academic Lead, Industry Lead, Methods Lead, Clinical Lead</b>	Academic Leads: Prof. Isabelle Huys (KU Leuven), Rosanne Janssens (KU Leuven) Clinical Leads: Prof. Dr. Michel Delforge (KU Leuven/UZ Leuven), Dr. Elena Cabezudo (Hospital de Sant Joan Despi Moisés Broggi. Consorci Sanitari Integral/ICO-Hospitalet), Dr. Raija Silvennoinen (Helsinki University Hospital), Dr. Daniel Coriu (Clinica de Hematologie) Methods Lead: Prof. Martina Vandebroek (KU Leuven)
<b>Title of the Study</b>	Patient Preferences for Multiple Myeloma Treatments
<b>Background/ Study Rationale</b>	Treatments for Multiple Myeloma (MM) in development and use today, are associated with a range of different characteristics and related uncertainties, such as novel side-effects, long-term effects, and efficacy outcomes. This raises questions about how MM patients think about these characteristics, and which of them are most influential when they consider MM treatments. Understanding preferences among MM patients is therefore especially valuable in decision-making surrounding MM treatments, where uncertainty exists about the key factors that drive MM patients when considering MM treatments. Furthermore, decisions surrounding MM treatment can be labelled 'preference sensitive' as: i) multiple treatment options exist and there is no option that is clearly superior for all patients; ii) the evidence supporting one option over others is considerably uncertain or variable and iii) patients' views about the most important benefits and acceptable risks of a treatment vary considerably within a population and may differ from those of healthcare professionals <sup>1</sup> .
<b>MPLC Decision-making</b>	Main stakeholder (industry, regulatory, HTA): industry, regulatory, HTA MPLC decision-points of interest: This study may inform industry, regulatory and HTA decision-making regarding MM treatments on patient-relevant treatment attributes. More specifically, results from this study could: i) reveal the patient perspective on unmet treatment needs in early drug development, ii) inform the development of patient reported outcome

<sup>1</sup> To be revised the moment we submit the BB, tailored to match the definition of patient preference sensitive decisions in PREFER

	measures and the treatment outcomes included in MM clinical trials, iii) help understand the most important treatment outcomes, benefits and risks in regulatory benefit-risk assessment, HTA and payer decisions.
<b>Clinical Research Questions</b>	<p>The overall aim of this study is to gain insights into the preferences of MM patients regarding different characteristics of MM treatments. The study consists of two sequential phases (sub-studies), whereby the qualitative study informs the attributes, the levels and their descriptions, further investigated in the quantitative study (see “Study Design”). The research questions of these phases are:</p> <ol style="list-style-type: none"> <li>1. <b>Qualitative study (completed):</b> Which characteristics (or attributes) of MM treatments do MM patients find most important and hence should be included as attributes in the quantitative study?</li> <li>2. <b>Quantitative study (ongoing):</b> <ol style="list-style-type: none"> <li>a. What is the quantified importance (“weight”) of relevant MM treatment attributes?</li> <li>b. How are MM patients’ preferences influenced by their characteristics such as demographic characteristics, treatment experience, and disease stage?</li> </ol> </li> </ol>
<b>Methodological Research Questions</b>	<p>The methodological research questions of the quantitative study are:</p> <ul style="list-style-type: none"> <li>• How do Swing Weighting (SW) and Discrete Choice Experiment (DCE) compare in the elicitation of MM patient preferences?</li> <li>• Were the most important MM treatment attributes included in the survey according to MM participants?</li> <li>• What was MM participants’ experience with the survey, and the DCE and SW preference elicitation questions?</li> </ul>
<b>Study Design</b>	<p>The study consists of two sequential phases (sub-studies), whereby the qualitative study informed the attributes, levels and their descriptions included in the quantitative study. The design of in these phases is:</p> <ol style="list-style-type: none"> <li>a. <b>Qualitative study (completed):</b> This phase involved three stages: i) a scoping literature review, to identify the characteristics associated with MM treatments in development and use today, ii) discussions with a heterogeneous sample of MM patients (n=24) in Belgium, Finland, Romania, and Spain using Nominal Group Technique, to understand which of these are most important and should be included in the quantitative study; iii) a combined quantitative and qualitative thematic analysis involving multi-stakeholder discussions with patients and clinicians, to refine the language and explanation of the attributes, levels and their descriptions for inclusion in the quantitative study.</li> <li>b. <b>Quantitative study (ongoing):</b> An online survey widely spread among the MM patient population across Europe, particularly targeting MM patients in Belgium, Finland, Spain and Romania. The survey quantifies the relative value of these attributes and the trade-offs they are willing to make between hypothetical MM treatments or MM treatment effects. The survey uses both Discrete Choice Experiment (DCE) and Swing Weighting (SW) to quantify patients’ preferences towards the same attributes. The DCE questions ask respondents to make hypothetical choices between two hypothetical MM treatments, which differ in how they perform on the different attributes identified in the qualitative study. SW directly captures individual patient preferences for each attribute. Instead of choosing between full treatment</li> </ol>

scenarios, respondents of the SW experiment are asked for their preferences for moving from the worst to the best levels (swings) on each attribute. Two treatment alternatives per DCE choice task are included. In view of the complexity of the choice tasks, no opt-out option was included; it may trigger participants to select this option when the choice task is difficult and thereby give limited information on treatment preferences. For the same reason, more than two alternatives would make the choice task too difficult. Regarding the levels, two levels were included for each of the attributes in the DCE and SW questions, since the methodological aim of this study is to compare DCE with SW and the SW questions only elicit preferences for the swing from the worst to the best level (= two levels), and a linear preference for all attributes is expected.

The online survey contains the following components:

- Introduction
- Information form
- Informed consent form
- SW questions\*
- Patients' self-reported comprehension and evaluation of the SW questions\*
- DCE questions\*
- Patients' self-reported comprehension and evaluation of the DCE questions\*
- Background questions (clinical and demographic characteristics, quality of life of patients (EQ-5D questionnaire))
- Patients' evaluation of the complete survey including a question asking their preference for DCE vs SW questions

\*Half of the participants will be randomly assigned to receive the DCE (+ comprehension and evaluation questions) questions first and afterwards the SW (+ comprehension and evaluation questions), whereas the other half will first perform SW (+ comprehension and evaluation questions) and afterwards the DCE (+ comprehension and evaluation questions) questions. Sawtooth Software<sup>2</sup> was used to construct the DCE tasks and implement the online survey. Participants can complete the survey with a laptop, a smartphone or a tablet. The survey was developed in English and translated to the native languages of the participants (Dutch, English, Finnish, French, German, Romanian and Spanish) by a professional translation company. These translations were then checked for accuracy and lay-language by team members who are native speakers of the language used in the survey and/or patient organization members part of the study team or their members. The survey was reviewed and/or pre-tested by MM patients (5 in total across countries) to obtain patients' feedback on the overall survey and more specifically: i) their understanding of the questions (including the attributes and their descriptions), ii) their understanding of the purpose of the elicitation tasks and usefulness of the explanation preceding the elicitation tasks, iii) whether the survey was not too lengthy and iii) any other feedback.

---

<sup>2</sup> <https://www.sawtoothsoftware.com/>

	<p><b>Educational materials:</b> The DWE and SW questions are preceded by a section of written information that explains the attributes, the levels, how the DCE and SW questions work and what the purpose of the elicitation questions is. No educational video is included. To assess the usefulness of the educational component, LIKERT questions asking for patients' evaluation and comprehension were added following both the SW questions and the DCE questions.</p> <p><b>Psychosocial constructs:</b> Health literacy (the patient's ability to read, understand, and use healthcare information appropriately) of participants is assessed via the inclusion of the validated 3-item questionnaire, Chew's Set of Briefing Screening Questions (Chew et al., 2004)</p>
<p><b>Study Population</b></p>	<ol style="list-style-type: none"> <li><b>1. Qualitative study (completed):</b> Haematologists recruited 24 MM patients across 4 countries (Belgium, Finland, Spain, and Romania), who were diverse in terms of treatment experience, and disease stage. During recruitment, the haematologists kept in mind that goal of this study was to ensure that the attributes identified were not directed only to patients with a specific treatment exposure, disease history or country of origin; but rather towards all patients along the MM spectrum. During sampling, haematologists used the following inclusion criteria: i) patients diagnosed with symptomatic MM; ii) patients ability to understand the language to be used in the discussion and iii) patients ability to participate in the discussion. Six MM patients were included in each of the four countries: Belgium, Finland, Romania and Spain. These countries were included to account for potential differences in patient characteristics and, as mentioned above, and ensure the identified attributes are not geared only to patients with a specific treatment or disease history. A much larger number above 6 patients would delay the phased process of the NGT discussion which aims to reach consensus in a short time span (up to 2 hours). Saturation, the point when "no new information or themes are observed in the data", was used to define when data collection could stop. Following qualitative data collection in the four countries, it appeared that the same themes of treatment attributes were observed across different countries. Hence, it was decided that no additional data was needed to inform the attributes.</li> <li><b>2. Quantitative study (ongoing):</b> An online survey widely spread among the European MM patient population. The survey is spread via patient organisations and haematologists among MM patients across Europe, but specifically targeting MM patients Belgium, Finland, Spain and Romania. The survey includes screening questions that allow persons with asymptomatic multiple myeloma (or smouldering myeloma) or active (symptomatic) multiple myeloma. The survey is available in Dutch, English, Finnish, French, German, Romanian and Spanish. The survey seeks to include as many participants as possible. For SW, no specific sample size needs to be established upfront; Tervonen et al.<sup>3</sup> describe that since SW captures per- patient preferences in an exact manner, the required sample size does not depend on the magnitudes of utility. Regarding DCE, De Bekker-Grob et al.<sup>4</sup> describe a technique for calculating upfront the minimum</li> </ol>

<sup>3</sup> Tervonen T, Pignatti F, Postmus D. From Individual to Population Preferences: Comparison of Discrete Choice and Dirichlet Models for Treatment Benefit-Risk Tradeoffs. *Medical decision making : an international journal of the Society for Medical Decision Making*. 2019;39(7):879-85

<sup>4</sup> de Bekker-Grob EW, Donkers B, Jonker MF, Stolk EA. Sample Size Requirements for Discrete-Choice Experiments in Healthcare: a Practical Guide. *The Patient*. 2015;8(5):373-84

	<p>sample size for parameter estimates in DCEs. However, to use this approach, the following needs to be known: i) significance level; ii) statistical power level; iii) statistical model used in the DCE analysis; iv) initial belief about the parameter values and v) the DCE design (defined by the number of choice sets, the number of alternatives per choice set, the number of attributes, and the combination of the attribute levels in each choice set). Since these elements are not known, it is not possible to estimate the sample size using the technique described by de Bekker-Grob et al. A practical solution described by de Bekker-Grob et al. that does not require any sample size calculations is to maximize the sample size to facilitate in-depth analysis.</p>
<p><b>Exposure &amp; Outcome</b></p>	<p>Since the results from this study are extremely suitable to be discussed with patients and in view of the multitude of stakeholders - including drug developers, regulatory and HTA stakeholders - that might benefit from using the results from this study, it will be sought to valorize the interpretation, learnings and use of the results from this preference study by engaging with these stakeholders. For example, via a workshop towards the end of the study, bringing together patients and their organizations, regulators and other stakeholders. Patient organizations are involved throughout all steps of the study and will help dissemination and utilization of findings in advocacy efforts. Further, patient organization members are continuously asked to review and provide input on the methods, results and publication coming from this study. Patient organizations will also bring forward the results of this study to their members and stakeholders they engage with.</p>
<p><b>Study Setting</b></p>	<p>The invitation, information sheet, informed consent and link to the online survey is being disseminated via social media, telephone, hospital visit, face-to-face and e-mail. Persons recruiting will be the study contributors, hematologists, patient organizations, hematology associations and relevant scientific associations.</p>
<p><b>Statistical Methods</b></p>	<ul style="list-style-type: none"> <li>• <b>Qualitative study (completed):</b> ANOVA and Fischer exact tests were performed to investigate statistically significant differences in between groups of participants between countries. In addition, a combined quantitative descriptive analysis of patients' rankings and iterative qualitative thematic analysis on the discussion transcripts was used to determine the attributes. Participants' self-reported characteristics (including health literacy) were analysed descriptively using Microsoft Excel. The average grades for the characteristics were calculated per country to derive rank orders and averages at country level. To obtain a final rank of the themes capturing the characteristics, the averages for each of the treatment characteristics pertaining to a theme was calculated by combining the averages obtained in the four countries.</li> <li>• <b>Quantitative study (ongoing):</b> The type of statistical analysis will depend on what is required for meeting the research objectives, and the type of data collected:             <ul style="list-style-type: none"> <li>○ <b>To quantify the importance (“weight”) of patient-relevant treatment attributes and to investigate how the attribute weights are influenced by patient characteristics, namely clinical and demographic characteristics and quality of life</b></li> </ul> </li> </ul>

	<ul style="list-style-type: none"><li>▪ <b>SW questions:</b> Per-participant attribute weights are obtained directly by participants' answers to the SW questions. In addition, it will be explored to estimate the distribution of the attribute weights in the population, based on the per-participant SW weights, such as via the Dirichlet distribution<sup>5</sup>. Heterogeneity will be analysed using standard statistical methods external to SW such as cluster analysis.</li><li>▪ <b>DCE questions:</b> Conditional logit (CL), latent class (LCA), Hierarchical Bayes (HB), and mixed logit (MXL), will be fitted to the data to explore which one has the best fit. In addition, cluster analysis will be used to cluster the individual attribute weights obtained from the MXL model to detect homogeneous groups. The MXL and LC model allow obtaining per-participant attribute weights and analyzing preference heterogeneity.</li><li>○ <b>To compare SW and DCE:</b> Individual per-participant preferences obtained through SW vs DCE will be compared. More specifically, it will be sought to evaluate whether the attribute weights a patient assigns in the SW questions are consistent with his/her choice behavior in the DCE questions; based on the attributes as obtained via the SW questions, we will assess whether the answer to the DCE questions corresponds with the choice that is expected based on the attributes as obtained through SW. It is expected that, if both methods measure the same preferences, the choice behavior of a given patient in a given DCE question does not systematically deviate from the weights the patient assigned to the attributes in the SW questions and vice versa.</li><li>○ <b>To investigate participants' opinion on whether the right attributes were included.</b> The following question is included at the end of the survey: "We asked you to state your preference for (placeholder treatment attributes). Are there any other aspects of multiple myeloma treatment that are more important for your decision-making than these ones? If so, please describe them here" The answers to this question will be analysed: i) thematically to assess which themes (=lacking attributes) were missing from the survey and ii) quantitatively, to assess how many times a specific lacking attribute was mentioned across survey participants.</li><li>○ <b>To assess participants' self-reported comprehension and overall evaluation of the SW and DCE questions</b> LIKERT questions asking for patients' evaluation and comprehension are added following both the SW questions and the DCE questions. At the end of the survey, multiple choice and open questions ask participants: i) to evaluate the entire survey, ii) whether any attribute is more important than the ones included and iii) state their preference for the SW vs DCE questions. Descriptive analyses will be conducted to summarize responses to the LIKERT and multiple-choice question. Patients' responses to the LIKERT</li></ul>
--	--

<sup>5</sup> Tervonen T, Pignatti F, Postmus D. From Individual to Population Preferences: Comparison of Discrete Choice and Dirichlet Models for Treatment Benefit-Risk Tradeoffs. *Medical decision making : an international journal of the Society for Medical Decision Making*. 2019;39(7):879-85.



	<p>questions and multiple-choice questions will be compared using standard statistics such as the Mann-Whitney U test or Chi square test. Patients' responses to the open questions will be analyzed qualitatively using thematic analysis. The results from the statistical analysis on the LIKERT and multiple-choice questions will be compared with the results of the open question asking patients to state a preference for either of the methods</p>		
Results	PREFER Research Questions	How was it addressed in this preference study	What was found
	<p><b>Qualitative study (completed):</b> Which characteristics do MM patients find most important and hence should be included in the quantitative study?</p>	<p><b>See 'study design':</b> Three stages: i) a scoping literature review, to identify the characteristics associated with MM treatments in development and use today, ii) discussions with a heterogeneous sample of MM patients (n=24) in Belgium, Finland, Romania, and Spain using Nominal Group Technique, to understand which of these are most important and should be included in the quantitative study; iii) a combined quantitative and qualitative thematic analysis involving multi-stakeholder discussions with patients and clinicians, to refine the language and explanation of the attributes, levels and their descriptions for inclusion in the quantitative study.</p>	<p>This qualitative study revealed the concerns and key treatment aspects that are most important for MM patients, and hence were included as attributes in the subsequent preference survey. In particular, MM patients voiced significant expectations and hopes that treatments extend their lives and reduce their cancer signs and symptoms. Participants were afraid of serious side-effects that are life-threatening and could cause permanent organ damage. Bone fractures and debilitating neuropathic effects (such as chronic tingling in their feet) were highlighted as major issues reducing patients' independence and mobility. Patients also discussed the adverse impact of the following symptoms and side-effects on their daily activities: difficulties to think clearly and concentrate, increased susceptibility to infections, reduced energy, pain, emotional problems, and vision problems. MM patients were concerned with uncertainties regarding the durability of positive treatment outcomes, as well as the cause, severity, and duration of their symptoms and side-effects. MM patients feared short-term positive treatment responses complicated by permanent, severe side-effects and symptoms. Participants emphasized that their individual opinions and preferences were shaped by their previous treatment and disease experience; participants more frequently raised those symptoms and side-effects they had previously experienced. Based on the qualitative thematic and quantitative analysis, the following eleven attributes were identified: i) additional life expectancy in years; ii) risk of life-threatening side-effects; iii) expected treatment response; iv) duration and severity of nerve or bone problems affecting movement; v) duration and severity of thinking problems; vi) duration and severity of increased susceptibility to infections; vii)</p>

			<p>duration and severity of reduced energy; viii) duration and severity of pain; ix) duration and severity of emotional problems; x) duration and severity of eating- and digestive problems; xi) duration and severity of vision problems. These attributes were included in the subsequent preference survey that aims to determine the relative value of these attributes and the trade-offs MM patients are prepared to make between these attributes. The DCE survey uses a partial-profile design, showing four attributes of the eleven attributes in each DCE question.</p>
	<p><b>Quantitative study (ongoing):</b> What is the quantified importance (“weight”) of relevant MM treatment attributes?</p> <p>How do Swing Weighting (SW) and Discrete Choice Experiment (DCE) compare in the elicitation of MM patient preferences?</p>	<p><b>See ‘statistical methods’:</b> An online survey among the MM patient population that quantifies the relative value of these attributes and the trade-offs they are willing to make between hypothetical MM treatments or MM treatment outcomes. The survey uses Discrete Choice Experiment (DCE) and Swing Weighting (SW) to quantify patients’ preferences towards the attributes (see ‘study design’).</p>	<p>A preliminary analysis was performed on 133 completions of the SW questions and 146 completions of the DCE questions, resulting in 133 completions of both DCE and SW questions.</p> <p><b>SW questions</b></p> <p>A preliminary analysis of the 133 completed SW questions yielded the following rank order for the attributes (from high to low importance): i) life expectancy, ii) treatment response, iii) pain, iv) life-threatening side-effects, v) mobility problems, vi) thinking problems, vii) increased susceptibility to infections, viii) vision problems, ix) reduced energy, x) eating and digestive problems xi) emotional problems.</p> <p><b>DCE questions</b></p> <p>Based on the 146 completed DCE questions, preliminary Conditional Logit (CL), Latent Class Analysis (LCA) and Hierarchical Bayes (HB) modelling was performed. The CL revealed that the MM patients’ choices were significantly affected by the attribute compositions of the MM treatment concepts (<math>\chi^2 = 483</math>). The t-ratios for each attribute level were greater than 5.8, meaning that all attribute effects are statistically significantly different from 0 at the 95% confidence level. The LCA model with 5 groups had a <math>\chi^2</math> of 619, and yielded the following rank order and average values for the attributes (from high to low importance): i) life expectancy (13.72), ii) mobility problems (11,20), iii) thinking problems (11,13), iv) pain (10,75), v) life-threatening side-effects (10,38), vi) treatment response (8.04), vii) vision problems (7.89), viii) eating and</p>

			digestive problems (7.50), ix) emotional problems (6.94), x) reduced energy (6.88), and xi) increased susceptibility to infections (5.58).
<b>Conclusions</b>	<p><b>Qualitative study (completed):</b></p> <p>Results from this study argue in favor of MM drug development that not only focuses on extending patients' lives but also on addressing those symptoms and side-effects that significantly impact MM patients' quality of life. This study underscores a need for transparent communication towards MM patients about the uncertainties regarding the long-term efficacy and safety of MM treatments. Finally, this study may help understand which quality of life-related treatment outcomes are most important to MM patients and therefore should be considered for systematic incorporation in MM drug development and evaluation.</p> <p><b>Quantitative study (ongoing):</b></p> <p>The preliminary analysis of the survey suggests that life expectancy is most important to MM participants, as revealed by both DCE and SW questions. Mobility problems, thinking problems, pain and life-threatening side-effects, and treatment response scored among the most important attributes in both SW and DCE questions, but their order differs depending on whether the responses stem from the DCE vs SW questions. Vision problems, eating and digestive problems, emotional problems, reduced energy, and increased susceptibility to infections score among the less important attributes in both SW and DCE questions, but their order differs depending on whether the responses stem from the DCE vs SW questions.</p>		
<b>Limitations</b>	<p><b>Qualitative study (completed):</b></p> <p>The impact of the COVID-19 on this research is especially relevant since this study consisted of qualitative discussions with MM patients, who have a higher susceptibility to infections. Conducting this study during COVID-19 required flexibility from both participants and study team. For the online and telephone discussions, it is likely that participants, who were not comfortable with online discussions or telephone (e.g. elder participants), were less likely to participate. The study team tried to be as inclusive as possible during recruitment, by offering both face-to-face and online discussions, according to the preferences of the participants and the local social distancing and hospital guidelines. Further, technical support for participants was given throughout the entire study. A steps-wise guideline explaining the practicalities of the discussion beforehand, and ensuring there was an opportunity, before the session for participants to test whether they could participate in the discussion. Still, the median age of diagnosis of patients included in this study was 55, which is 11 years younger than the median age reported by by Kazandijan<sup>6</sup> in 2016 and 14 years younger than the average age of diagnosis reported by ASCO<sup>7</sup> in 2020.</p>		

<sup>6</sup> <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5283695/#R2>

<sup>7</sup> <https://www.cancer.net/cancer-types/multiple-myeloma/risk-factors#:~:text=Myeloma%20occurs%20most%20commonly%20in,occur%20in%20people%20under%2040>

	<p>Further regarding generalizability, it is important to note that the purpose of this study was not to make statements about a population larger than the included sample. Rather, it aimed to gain in-depth insight into the opinions of patients participating in the discussions including which attributes are important to them and why.</p>
<b>Potential Use of Results in MPLC Decision-Making</b>	<p>The outcomes of this research will benefit stakeholders (industry, regulators, HTA/payer bodies) to identify priorities and unmet treatment needs for (new) treatments in MM. More specifically, this study will clearly identify which outcomes and endpoints matter most to people living with MM, and therefore should be included in the clinical development path as clinical trial endpoints or patient reported outcomes measures. Systematically including the patient-relevant attributes identified in this study in clinical development, regulatory benefit-risk assessment, HTA/reimbursement decisions and post-marketing decisions, could result in a more patient centric- drug development and evaluation process. Conversely, when there is no evidence that an MM drug targets any of the prioritized attributes identified in this study, it may be recommended that such evidence needs to be collected before or after marketing authorization and/or reimbursement. The results from the quantitative preference survey can also be used to assess how MM patients' trade-off between the patient-relevant attributes as indicated by the qualitative study, information useful for regulatory benefit-risk assessments of MM drugs. Finally, if results from the quantitative preference survey provide insights into how people eligible for MM treatment with the new drug prefer it relative to other treatment options available to them, this study may be used to support the reimbursement of a new MM drug.</p>

<b>Therapeutic Area</b>	Hemophilia and gene therapy
<b>Core Case Study, Academic Case Study, Industry Case Study</b>	Academic case study
<b>Current status</b>	This case study is finalized
<b>Academic Lead, Industry Lead, Methods Lead, Clinical Lead</b>	<u><b>Academic Lead:</b></u> Isabelle Huys (KU Leuven) & Eline van Overbeeke
<b>Title of the Study</b>	Patient preferences to Assess Value IN Gene therapies (PAVING)
<b>Background/ Study Rationale</b>	Recently, gene therapy for the treatment of hemophilia has been developed. Clinical trial results are promising. However, multiple challenges remain. These challenges mainly relate to the fact that it is currently still unknown whether the therapeutic effect of gene therapy will be maintained throughout the full lifespan of patients, and what side effects may occur in the long run.
<b>MPLC Decision-making</b>	<u><b>Main stakeholder (industry, regulatory, HTA):</b></u> HTA
	<u><b>MPLC decision point of interest:</b></u> HTA/reimbursement
<b>Clinical Research Questions</b>	Identify attributes of gene therapy and standard of care that are important to patients
	Understand trade-offs that patients make when choosing between gene therapy and standard of care
<b>Methodological Research Questions</b>	Question 1.11/5b: What approach can be used to determine which patient preference method to use for a given circumstance?
	2.31/11d To what degree do preference results in a disease vary with Disease experience or severity
	2.14/4a2 Which presentation formats respondents prefer (displays, video, descriptions, etc.)
	3.12/15b How to ensure patient preference information from preference studies are applicable/tailored to HTA/payer decision making

<b>Study Design</b>	Interviews and threshold technique		
	<u>Educational materials</u> : Educational tool on gene therapy		
	<u>Psychosocial constructs</u> : Chew Health literacy screening questions		
<b>Study Population</b>	Patients diagnosed with moderate or severe hemophilia A or B, 18 years or older, and living in Belgium		
<b>Exposure &amp; Outcome</b>	Minimum acceptable benefit		
<b>Study Setting</b>	Hemophilia reference centers and patient organizations from April-May 2020		
<b>Statistical Methods</b>	Interval regression and plotting of thresholds		
<b>Results</b>	<b>PREFER Clinical and Methodological Research Questions</b>	<b>How was it addressed in this preference study</b>	<b>What was found</b>
	<b>Clinical Questions:</b>		
	Identify attributes of gene therapy and standard of care that are important to patients	Interviews including: open questions, ranking exercise and case questions	Top 5: <ul style="list-style-type: none"> <li>• Annual bleeding rate</li> <li>• Factor level</li> <li>• Uncertainty long-term risks</li> <li>• Impact on daily life</li> <li>• Probability that prophylaxis can be stopped</li> </ul>
	Understand trade-offs that patients make when choosing between gene therapy and standard of care	Threshold technique using 4 attributes in 2 labelled profiles (gene therapy vs. prophylactic factor replacement therapy)	Included attributes: <ul style="list-style-type: none"> <li>• Annual bleeding rate (ABR)</li> <li>• Chance to stop prophylaxis (STOP)</li> <li>• Time that side effects have been studied (TIME)</li> <li>• Quality of life (QOL)</li> </ul>

	<b>Methodological Questions:</b>		
Question 1.11/5b: What approach can be used to determine which patient preference method to use for a given circumstance?	Method selection was based on assessment of methods according to different selection criteria	In the context of gene therapy for a rare disease the threshold technique was found most suitable.	
2.31/11d To what degree do preference results in a disease vary with Disease experience or severity	Disease severity (severe vs. moderate hemophilia) is included in the covariate-adjusted regression model to assess the effect of disease severity on preferences	Severity of disease (mod vs sev) had no influence on preferences.	
2.14/4a2 Which presentation formats respondents prefer (displays, video, descriptions, etc.)	In interviews written text was discussed with the patient and preferences were investigated on how patients wanted the information to be formulated and presented. An educational tool was then designed for the quantitative phase and was further refined with patients.	Patients wanted information on gene therapy to be communicated using animations. The educational tool that was used in the survey effectively educated participants on gene therapy and increased acceptance of gene therapy (i.e. the more time participants spent on the tool, the less benefit they required to accept gene therapy).	
3.12/15b How to ensure patient preference information from preference studies are applicable/tailored to HTA/payer decision making	The PAVING study was supported by an advisory board that included, amongst others, HTA and payer decision makers. The advisory board provided feedback on study design to ensure their design and quality needs were met.	Advisory board reached a consensus on attribute classes to explore: benefits (including clinical endpoints and QoL), risks, and administration + to exclude: costs.	
<b>Conclusions</b>	Interviews showed that most hemophilia patients have a positive attitude toward gene therapy and that besides efficacy, safety and the related uncertainties, also impact on daily life is important to patients. Results from the threshold technique survey proved that education may facilitate patients' acceptance of novel treatments. Moreover, preference heterogeneity for novel treatments was confirmed in this study.		

<b>Strengths &amp; limitations</b>	<ul style="list-style-type: none"><li>• The sample was representative of the Belgian hemophilia population</li><li>• Available therapies for hemophilia other than gene therapy and prophylactic factor replacement therapy were not considered</li></ul>
<b>Potential Use of Results in MPLC Decision-Making</b>	The identified patient-relevant attributes may be used by regulators, health technology assessment bodies and payers in their evaluation of gene therapies for hemophilia. Moreover, they may inform clinical trial design, pay-for-performance schemes and real-world evidence studies. The quantitative results may be used to inform value propositions and budget impact analysis. In gene therapy decision-making, preference heterogeneity and the impact of patient education on acceptance should be considered.



## 8.2 Annex II: PREFER: Historical case studies

Prefer Task 3.3 involves the collection and review of historical case studies (i.e. completed) performed by PREFER consortium organizations. This is a convenience sample of high-quality preference studies that are being used to supplement the contributions of the core case studies in terms of addressing PREFER methodological research questions, providing lessons learned, and contributing insights to inform the WP4 final recommendations.

In the first phase of this task, 24 case studies were submitted, reviewed, and archived. In the collection of these case studies, an intake form was used that included question to the principal investigator such as:

- How did you choose your final study design?
- What were the key challenges in getting the study initiated?
- How was the study funded?
- How did you select the attributes and levels?
- Did the study have the impact you expected? If not, why do you think so?
- How subjects recruited and what were the associated challenges?
- Were study results shared with regulators or HTA bodies?
- What was the unique contribution(s) of your study to this knowledge area?
- Were your results used as part of a multi-criteria decision-making exercise?
- Are there any aspects of the study that you consider new or novel?
- If this is in support of a pharmaceutical product, what stage of development was that product in when the study was initiated?

In addition to the completed intake form, final study reports or publications were submitted for each case study.

Below is a listing of the historical case studies that have been reviewed and archived.

**Table 8-1 Listing of historical case studies (wave 1)**

PREFER partner	Study title
<b>Eli Lilly</b>	Patients Benefit-risk trade off trade-offs for Psoriasis treatment attributes
<b>Pfizer</b>	Social preferences for rheumatoid arthritis treatments: evidence from a discrete choice experiment
<b>Janssen</b>	Relative importance of benefits and risks associated with antithrombotic therapies for acute coronary syndrome: patient and physician perspectives
<b>Bayer</b>	Patient Benefit-risk trade-offs for radioactive iodine-refractory differentiated thyroid cancer treatments
<b>Universitätsklinikum Erlangen</b>	Identifying factors of patient-preference for the treatment of rheumatoid arthritis with biologicals
<b>UMC Utrecht and the Dutch Institute for Public Health</b>	Parental preferences with regard to rotavirus vaccination for their newborn
<b>Novartis</b>	Preferences and stated adherence for antibiotic treatment of cystic fibrosis pseudomonas infections
<b>University of Birmingham</b>	Preferences and acceptabilities in relation to drug formulations in paediatric populations
<b>Janssen</b>	Patient preferences related to benefits, risks and formulations of schizophrenia treatment
<b>Janssen</b>	Physician benefit-risk preferences from two randomized long-acting injectable anti-psychotic trials
<b>Pfizer</b>	The effect of information on preferences for treatment of metastatic renal cell carcinoma

<b>AbbVie</b>	Endometriosis patient preference study
<b>Janssen</b>	Physician and patient benefit-risk preferences from two randomized long-acting injectable antipsychotic trials
<b>Erasmus MC</b>	Future pandemics and vaccination: public opinion and attitudes across three European countries
<b>Pfizer</b>	Chronic pain patients' treatment preferences
<b>Janssen</b>	Patient benefit-risk preferences for anti-coagulation and antiplatelet therapies in atrial fibrillation and venous thromboembolism
<b>Janssen</b>	Psychiatrists' judgments about antipsychotic benefit and risk outcomes and formulation in schizophrenia treatment
<b>IEO European Institute of Oncology</b>	Health orientation, knowledge and attitudes toward genetic testing and personalised genomic services
<b>University of Birmingham</b>	Palatability and acceptability of multiparticulate (placebo) formulations: adults vs. children comparison
<b>Janssen</b>	Physician benefit-risk preferences for anticoagulation and antiplatelet therapies in atrial fibrillation and venous thromboembolism
<b>Janssen</b>	How heterogeneous are preferences for delaying onset of Alzheimer's disease? It depends on how you look
<b>KULeuven</b>	Quantifying benefit risk preferences for new medicines in rare diseases for patients and care givers
<b>IEO European Institute of Oncology</b>	Direct to Consumer genetic testing: exploring psychological profile of health related decisions and behaviours of consumers

**Roche**

Patients' and physicians' quantitative benefit risk preferences in rheumatoid arthritis

A second call for case studies will be initiated to address gaps identified in the overall PREFER case study portfolio (i.e. core case studies, industry case studies, academic-led case studies, and the historical case studies). These gaps include studies using methods other than DCE, those submitted to regulatory agencies, those in geographies not well represented currently in our portfolio, and those in rare diseases.

### 8.3 Annex III: PREFER methodological Research Questions

Results from currently available case studies on PREFER methodological research questions are summarized in the tables below. More evidence will become available in Spring 2021.

Table YY.1 presents observations related to the methodological research questions included in prospective PREFER case studies from the indicated case studies in which data have become available. For some research questions, no prospective PREFER studies were available. In these instances, case studies from the PREFER consortium were requested (referred to as historic case studies) and those contributed were reviewed, categorized into the appropriate research question, and included within this table overview where appropriate. Two PREFER case studies looked at patients' preferences according to different sources of recruitment and interpretation of preference data generalizability (Table YY.2)

**Table 8-2 PREFER methodological research questions related to "Preference Question Design" and relevant case study examples**

PREFER methodological research Q	Case studies in which research question was addressed; details
To what degree do small changes in the number, type and definitions of attributes impact results for a given method	Historic Case Study – <b>Attribute Type: Face vs other Psoriasis Plaque</b> , Fairchild et al. 2017: <u>Objective</u> - Using a DCE, the study objective was to quantitatively assess patients' tolerance for therapeutic risks associated with psoriasis treatments by trading off benefits of reduced plaque severity or reduced plaque area affecting the body, face, or hands. <u>Result</u> - Of those who completed the survey, 28% were unwilling to accept any greater risk of treatment-related infection mortality to achieve a benefit. Of the remaining 72%, respondents were willing to accept higher risks of infection-related mortality associated with treatment to completely remove plaques covering only 1% of the body, compared to reducing lesions from 10 to 1% of the affected area. This finding was more pronounced for lesions on the face.
Impact of educational material on understanding and preferences	<b>UU RA case study:</b> A strong, statistically significant impact on patients' preferences was revealed when including the educational tool as an interaction. Overall, respondents informed by the educational tool placed relatively more importance on treatment side effects, and respondents receiving the plain text placed relatively more importance on the treatment effectiveness and the administration methods. Relative differences were small, and uncertainty remains regarding how patient preferences are affected by different educational tools. <b>PAVING case study:</b> The majority of patients found that the educational tool either helped them 'very much' (49%) or 'moderately' (27%) in understanding the threshold questions. The median (Q2) time spent on the educational tool was 12.5 minutes, likely meaning that at least half of patients completed the full tool as the minimum necessary time to go through the full tool was 10 minutes. Moreover, the more time patients spent on the tool, the less ABR (annual bleeding rate) and QOL benefit they required to switch to gene therapy. As self-reported limited knowledge on gene therapy was found to be correlated with time spent on the educational tool, the tool seemed to have mainly helped patients with limited prior knowledge and to have resulted in patient activation.

<p>Can psychosocial constructs including PREFER Class I constructs and Class II constructs provide insight regarding an individual's preferences</p>	<p><b>MI study:</b> Based on the DCE responses, the latent class model identified <b>two classes of preferences</b>. Class 1 contained approximately 28% of the sample population. Patients in Class 1 did not have any significant preferences for changes in risk of cardiovascular death and heart attack. Their choices were mainly driven by changes in risk of bleeding within the skull, other severe bleeding and stroke. Meanwhile, approximately 72% of the sample population falls into class 2. Their choices were mainly driven by changes in risk of cardiovascular death. Using a MNL regression model, only the presence of bleeding risk factors (at least 1) and an inadequate level of health numeracy skill were found to significantly contribute to being in Class 1.</p> <p><b>Pfizer pain study:</b> The survey instruments (DCE and BWS) included the Multidimensional Health Locus of Control (MHLC) Scale – Form C and classified 3 subgroups: ‘internal’, ‘change’, and ‘powerful others’ locus of control. For DCE, in both the US and UK, subgroup analysis did not detect statistically significantly systematic differences in preferences based on locus of control. For BWS, in both the US and UK, subgroup analysis showed statistically significantly systematic differences in preferences between patients with high and low levels of locus of control for all three types of locus of control. This was manifested in stated preferences on relative ranking of risks, with those in the higher locus of control categories being more adverse to certain risks than those in the lower categories.</p>
--	--

**Table 8-3 PREFER methodological research questions related to “Conduct: Recruitment, and Analysis and Interpretation” with relevant PREFER case study examples**

<p><b>PREFER methodological research Q</b></p>	<p><b>Case studies in which this research question was addressed; details</b></p>
<p>To what degree do preference results in a disease vary with patients recruited through different channels (PO, physician)</p>	<p><b>UU RA additional academic case study:</b> “Respondents’ preferences for attribute levels were <b>not statistically influenced by the recruitment source</b> (i.e., respondents recruited from the clinic, the research panel, and the Rheumatism Association).”</p> <p><b>Additional academic PAVING study:</b> “Across the models for ABR, STOP and QOL, the effect of the type of hemophilia (hemo_type), severity of hemophilia (hemo_sev), <b>recruitment source</b> (source), and consistency (consist) on the willingness to switch from PFRT to gene therapy was <b>never significant</b>”</p>
<p>2.33: To what degree do preference results in a disease vary with <b>region</b>.</p>	<p><b>Novartis COPD study:</b> “Our study spanned 5 countries and preferences were analysed throughout according to country, with <b>some notable differences seen between countries</b>. Across countries patients placed almost equal importance on shortness of breath, sleep quality and urinary incontinence. Some differences in relative importance did emerge across the countries: shortness of breath was the most important attribute for patients from Australia and France, and sleep quality for patients from Japan”</p> <p><b>PAVING study:</b> “<b>Residence</b>, namely the difference between patients living in Flanders vs. Wallonia, <b>had a significant effect on all benefit thresholds</b>, with patients living in Wallonia requiring more benefit (higher</p>

	<p>reductions in ABR, higher chances to stop prophylaxis, and higher increases in QoL) than patients living in Flanders to make the switch from PFRT to gene therapy (Figures 6-8)."</p>
<p>To what degree do preference results in a disease vary with <b>disease experience or severity</b></p>	<p><b>Novartis COPD study:</b> "We performed extensive analysis of how disease severity affects preferences. <b>Overall, preferences were consistent across the three severity groups</b> (mild, moderate, severe), although some minor differences were observed"</p> <p><b>PAVING study:</b> "Across the models for ABR, STOP and QOL, the effect of the type of hemophilia (hemo_type), <b>severity of hemophilia (hemo_sev)</b>, recruitment source (source), and consistency (consist) on the willingness to switch from PFRT to gene therapy was <b>never significant</b>. Patients' ABR only had an effect on the ABR threshold, with patients with higher ABR tolerating higher numbers of additional bleeds." (ABR = annual bleeding rate.)</p>
<p>To what degree do preference results in a disease vary with <b>Characteristics of patients</b></p>	<p><b>Pfizer pain study:</b> "A secondary objective of this study was to explore systematic heterogeneity in preferences for <b>nine prespecified subgroup</b> pairs, which were defined by respondent condition (OA only, CLBP only, or concurrent OA and CLBP), opioid experience, age, time living with chronic pain, baseline pain level, number of correct responses on the omprehension questions, and "internal," "powerful others," and "chance" locus of control". <b>Preferences were only statistically significantly systematically different for the subgroup pair defined by the number of comprehension questions that respondents answered correctly.</b> Although respondents who answered fewer than three comprehension questions incorrectly had statistically significantly systematically different preferences than those who answered three or more comprehension questions incorrectly, the preference weights and conditional relative importance estimates for the two subgroups in this pair are qualitatively similar, and the order of importance of the treatment attributes was the same."</p> <p><b>MI study</b> "Of all the subgroups tested, only <b>age, presence of bleeding risk factor and patients' risk of future ischaemic event</b> have significant effect on their Preferences."</p> <p><b>MI study:</b> "there was no significant difference in the preferences between acute and chronic MI patients"</p> <p><b>PAVING study:</b> "<b>Age and correct comprehension</b> (i.e. correct response to the comprehension question) only had <b>significant effects</b> on the threshold for the chance to stop prophylaxis (STOP) attribute, with older patients and patients with incorrect comprehension requiring higher chances to stop prophylaxis to switch to gene therapy (Figure 7)."</p>

Table 8-4 PREFER methodological research questions related to methods

<p>PREFER methodological research questions</p>	<p>Case studies in which research question was addressed; details</p>
---	---

<p>How do results differ between simpler/cheaper methods versus more complex/expensive methods?</p>	<p><b>Case Study NMD – Approach/Results</b> – DCE was not found to be more difficult to complete or more difficult to understand than BWS. Results of BWS and DCE differed with regard to relative importance scores for the 6 attributes. More results will be added later.</p> <p><b>Case Study RA – Approach:</b> Qualitative and quantitative comparison of results: MAR, relative importance of the attributes, proportion of respondents avoiding risks, and patient-reported ease of understanding/ completion</p> <p><b>Results:</b> In the general population, MAR values were numerically similar across methods for the attributes across the values evaluated. Differences in the relative importance of the attributes was found between methods. The relative importance was used to tabulate MAR, as analysed by DCE and MAR was tabulated for PTT. MAR was similar across methods. {additional analyses to be completed}</p> <p>In the general population, the majority of the respondents found the DCE and PTT were easy or very easy to understand and answer. However, the DCE was found to be significantly easier to understand and answer than the PTT (<math>p &lt; 0.05</math>). Respondents who received the DCE first found all tasks (both DCE and PTT) easier to understand and compared to respondents who received the PTT first (<math>p &lt; 0.05</math>).</p> <p><b>Case Study Diabetes</b> - the DCE was reported by participants to be both easier to understand and to complete, compared to the SW. For the DCE, the proportion of attribute importance is very different for all the attributes. Contrastingly, all attributes in the SW, received between 12-17% of the designated importance. The DCE had a 14.9-fold difference between the most and least important attribute, while the SW had a 1.4-fold difference.</p> <p><b>Case Study MI</b> – Comparison between DCE and BWS-1 showed difference in how patients valued the importance of the treatment attributes/ outcomes particular for prevention of stroke. This difference is likely to be due to the different underlying decision-making process between DCE and BWS-1 in this study. In the BWS-1, patients were asked to indicate the most and least important outcome to avoid without information on the risks of developing the outcomes. While in the DCE, they were asked to indicate their preferred treatment choice based on varying level of risks for each attribute.</p>
<p>How do results differ when different methods with the same set of attributes are applied in the same population?</p>	<p><b>Case Study Chronic Pain – Approach:</b> A DCE and a Case 1 BWS exercise was included. The DCE was the primary preference elicitation method; the BWS exercise allowed elicitation of preferences on many adverse events related to pain treatments that could not be accommodated in the DCE due to the number of attributes included in a DCE.</p> <p><b>Method:</b> compare the rank ordering of 3 key risk attributes between the two exercises and assess qualitatively the consistency of the conditional relative importance estimates for these attributes between the two exercises.</p> <p><b>Results:</b> In the US sample, the rank ordering of the conditional relative importance of the 3 risks included in both the DCE and the BWS was consistent between the methods. In the UK sample, the rank ordering of the conditional relative importance of the risks included in both the DCE and the BWS was inconsistent between the methods.</p>



<p>How to determine which method to use in a given circumstance (and can simulation studies inform this choice)?</p>	<p><b>Case Study PAVING</b> – <u>Approach</u>: the method should 1) estimate weights of attributes, 2) estimate trade-offs between attributes, 3) quantify preference heterogeneity, 4) incorporate internal validity measures, 5) not have technical issues, 6) have a low minimal necessary sample size, and 7) allow for incorporation in an unsupervised survey.</p> <p><b>Results</b>: DCE excluded due to sample size needs for attribute and level requirements. Through expert opinion consult and considering the attribute plus levels design needs, threshold or swing weighting methods were selected as appropriate. Concerns were raised that swing weight method may require support through interviews or workshops (due to complex choice tasks/high cognitive burden); therefore, the threshold technique was ultimately chosen.</p> <p><b>Case Study Chronic Pain</b> (Pfizer/Lilly) – <u>Approach</u>: Criteria were indirectly addressed based on the ability of the chosen method to address the study’s stated research questions.</p> <p><b>Results</b>: Since the primary objective was estimating tradeoffs between benefits, risks, and mode and frequency of treatments that reduce pain and improve function for the same disease state, but that these (potential) treatments have substantially different adverse events of concern, the willingness of patients to tolerate these risks in order to achieve the same outcome may differ. Also, these treatments vary substantially in mode and administration frequency. Thus, it is unknown whether patients would accept additional risk(s) or decreases in expected benefit to have the mode and administration frequency they prefer. As such, the preference method required trading off multiple attributes simultaneously. A DCE is a well-known, rigorous approach to such data needs.</p> <p>The second objective was to understand where multiple other treatment risks fit into patient decision-making relative to the key risks included in the DCE. Because there were many potential risks associated with each of the treatments for inclusion in the PPS, conducting a DCE that included all risks likely would not have been feasible. Therefore, a method that could accommodate a larger number of risks and provide estimates of relative importance was needed. A Case 1 BWS exercise was considered appropriate.</p>
<p>How generalizable are preferences from one specific population in a disease to different populations in that or related diseases?</p>	<p><b>Case Study Chronic Pain</b> – <u>Approach</u>: The sample included patients with related diseases—osteoarthritis only, chronic low back pain, and comorbid osteoarthritis and chronic low back pain. The sample was stratified to include nearly equal representation across all three categories to facilitate subgroup analyses. In each country, we used the same survey to elicit patient preferences for all three conditions. We believe that this was appropriate because the goals of treatment are the same for all conditions as are the available treatment options.</p> <p><b>Results</b>: In the US, subgroup analysis of the DCE data did not detect statistically significantly systematic differences in preferences among the three conditions. Latent class analysis showed a marginally insignificant (P = 0.055) difference in preferences for patients with chronic low back pain only. Subgroup analysis of the BWS data did not detect statistically significantly systematic differences in preferences among the three conditions.</p>

	<p>In the UK, subgroup analysis of the DCE data did not detect statistically significantly systematic differences in preferences among the three conditions. Latent class analysis showed a marginally insignificant (<math>P = 0.065</math>) difference in preferences for patients with chronic low back pain only. Subgroup analysis of the BWS data showed statistically significantly systematic differences in preferences among the three conditions.</p> <p><b>Conclusion:</b> The results of the DCE suggest that treatment preferences are mostly generalizable across conditions when the goal of treatment and the treatment options under consideration are the same; however, there is weak evidence that there may be systematic differences in preferences based on condition that were not captured in the DCE. In contrast, evidence from the UK suggests that preferences elicited using the Case 1 BWS are not generalizable across conditions.</p>
<p>How to determine which method to use in a given circumstance (and can simulation studies inform this choice)?</p>	<p><b>Case Study NMD – Approach:</b> Ease of use and understandability was considered in light of this patient population and impact of disease onset on symptom severity. Study research questions included benefit-risk trade-off questions for future treatment options and comparing results across methods. Included patients had neuromuscular disorders, all affecting the central nervous system that can include varying degrees of cognitive impairment and fatigue. the prevalence and severity of cognitive impairment depends on the age at onset of the disease, with earlier onset generally more severe than adult onset.</p> <p><b>Result:</b> The sample population was divided by age of onset of disease. Those with early onset were only given the ‘simpler’ survey method - Q methodology – to compare with a slightly more complex method, BWS Type 2. The later onset subgroup was also given BWS Type 2 but was thought capable of completing a more complex survey method, DCE, based on clinical knowledge of the disease state and patient population.</p>

## 8.4 Annex IV: Catalogue of identified patient preference study methods

The table below provides an overview of the 32 unique exploration (qualitative) and elicitation (quantitative) methods identified from the systematic literature review of Soekhai and colleagues (Soekhai et al 2019).

Method	Description	References
Exploration methods		
Citizens' juries <a href="#">b</a>	Group of individuals discussing issues on the basis of evidence provided by two trained moderators	C. Mitton, et al. <b>Public participation in health care priority setting: a scoping review</b> ; Health Policy, 91 (2009), pp. 219-228:  J. Street, et al. <b>The use of citizens' juries in health policy decision-making: a systematic review</b> ; Soc. Sci. Med., 109 (2014), <a href="https://doi.org/10.1016/j.socscimed.2014.03.005">10.1016/j.socscimed.2014.03.005</a>
Complaints procedures <a href="#">b</a>	Method in which stakeholders can register complaints to be investigated by experts	T. Bourne, et al. <b>The impact of complaints procedures on the welfare, health and clinical practise of 7926 doctors in the UK: a cross-sectional survey</b> ; BMJ Open, 2015 (2015), p. 5  M. Wensing, R. Grol, What can patients do to improve health care? Health Expect., 1 (1998), pp. 37-49
Concept mapping <a href="#">b</a>	Method that utilizes small groups of participants responding to various topics or issues, while ensuring each respondent is given equal opportunity to express their opinions and address other group dynamic issues	J.G. Burke, et al. <b>An introduction to concept mapping as a participatory public health research method</b> ; Qual. Health Res., 15 (2005), pp. 1392-1410  W. Trochim, M. Kane, Concept mapping: an introduction to structured conceptualization in health care Int. J. Quality Health Care, 17 (2005), pp. 187-191
Delphi method <a href="#">a</a> , <a href="#">b</a>	Structured, iterative forecasting method involving a panel of experts who provide anonymous responses to questionnaires with the opportunity to revise their responses when the anonymous summary of response from the prior round is revealed	R. Boulkedid, et al. <b>Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review</b> PLoS One, 6 (2005), Article e20476  J. de Meyrick The Delphi method and health research, Health Education, 103 (2003), pp. 7-16
Dyadic interview <a href="#">a</a> , <a href="#">b</a>	Method that utilizes two participants in a single interview, responding to open-ended questions asked by an interviewer to identify how a product, service or opportunity is perceived	Z. Eisikovits, C. Koren <b>Approaches to and outcomes of dyadic interview analysis</b> , Qual. Health Res., 20 (2010), pp. 1642-1655

		D.L. Morgan, et al. <b>Introducing dyadic interviews as a method for collecting qualitative data</b> , Qual. Health Res., 23 (2013), pp. 1276-1284
Focus group <a href="#">a</a> , <a href="#">b</a>	Method that utilizes a group of interacting individuals that provide information about a specific issue to identify how a product, service or opportunity is perceived	C.E. Basch, Focus group interview: an underutilized research technique for improving theory and practice in health education, Health Education Quarterly, 14 (1987), pp. 411-44 B. Schulze, M.C. Angermeyer, <b>Subjective experiences of stigma. A focus group study of schizophrenic patients, their relatives and mental health professionals</b> , Soc. Sci. Med., 56 (2003), pp. 299-312
In depth – individual interview <a href="#">a</a> , <a href="#">b</a>	Interview technique that allows for an intensive discussion with one interviewee to explore their perspectives on a particular topic or theme, to gain a deeper understanding of this particular topic or theme. Often only a limited amount of questions or themes are prepared by the interviewer, and the rest of the questions are based on the response of the interviewee	M.A. Harris, et al., <b>Validation of a structured interview for the assessment of diabetes self-management</b> , Diab. Care, 23 (2000), p. 1301 J.B. Williams, A structured interview guide for the Hamilton Depression Rating Scale, Arch. Gen. Psychiatry, 45 (1988), pp. 742-747
Nominal group technique <a href="#">b</a>	Method that utilizes a group process that involves making decisions by vote and ranking responses given by members of the group	J. Allen, et al., <b>Building consensus in health care: a guide to using the nominal group technique</b> , Br. J. Commun. Nurs., 9 (2004), pp. 110-114 M. Gallagher, et al., <b>The nominal group technique: a research tool for general practice?</b> Fam. Pract., 10 (1993), pp. 76-81
Public meetings <a href="#">b</a>	Method to gain public opinions on particular issues by allowing general members of the public to attend and voice their responses	C. Ham, <b>Priority setting in health care: learning from international experience</b> , Health Policy, 42 (1997), pp. 49-66 K.A. McComas <b>Theory and practice of public meetings</b> , Commun. Theor., 11 (2001), pp. 36-55
Semi-structured individual interview <a href="#">a</a> , <a href="#">b</a>	Interview technique that allows new ideas to be brought up during the interview as a result of what the interviewee says in a semi-structured setting, whereas in the structured setting the interviewer strictly sticks to an interview guide and does not ask questions based on the response of the interviewee	L.K. Barriball, A. While, <b>Collecting data using a semi-structured interview: a discussion paper</b> , J. Adv. Nurs., 19 (1994), pp. 328-335 L.S. Whiting, <b>Semi-structured interviews: guidance for novice researchers</b> , Nurs. Standard, 22 (2008), pp. 35-40
Elicitation methods		

<p>Adaptive conjoint analysis <a href="#">a</a></p>	<p>Method similar to regular conjoint analysis, but with adaptive conjoint choice tasks based on the earlier choices made within the survey, in theory allowing the survey to focus attention on those attributes or levels of those attributes that have the most influence on the choices of that individual. Unlike discrete choice experiments this method is founded in the theory of conjoint measurement (CM), which is more focused on the behavior of number systems instead of the behavior of human preferences</p>	<p>K.M. Beusterien, et al., <b>Understanding patient preferences for HIV medications using adaptive conjoint analysis: feasibility assessment</b>, Value Health, 8 (2005), pp. 453-461 L. Fraenkel, et al. <b>Understanding patient preferences for the treatment of lupus nephritis with adaptive conjoint analysis</b>, Med. Care, 39 (2001), pp. 1203-1216 J. Louviere, et al., <b>Discrete choice experiments are not conjoint analysis</b>, J. Choice Model., 3 (2010), pp. 57-72</p>
<p>Allocation of points <a href="#">b</a></p>	<p>Method that involves asking respondents to rate their conditions on scales, while knowing the weights which they attach to different criteria, indicating the relative importance of particular areas of their lives</p>	<p>K.L. Haywood, et al., <b>Patient centered assessment of ankylosing spondylitis-specific health related quality of life: evaluation of the Patient Generated Index</b>, J. Rheumatol, 30 (2003), p. 764 D.L.B. Schwappach, T.J. Strassmann, <b>Quick and dirty numbers? The reliability of a stated-preference technique for the measurement of preferences for resource allocation</b>, J. Health Econ., 25 (2006), pp. 432-448</p>
<p>Analytic hierarchy process <a href="#">a</a>, <a href="#">b</a></p>	<p>Method in which responders assess the relative importance of pairs of attributes (treatment endpoints, properties, criteria, items, objects, etc.) toward achieving a goal, where these responses are used to compute a weight for each attribute</p>	<p>Medical Device Innovation Consortium (MDIC), <b>Patient Centered Benefit-Risk Project Report: A Framework for Incorporating Information on Patient Preferences Regarding Benefit and Risk into Regulatory Assessments of New Medical Technology. Public Report 2015</b> M.J. Liberatore, R.L. Nydick, <b>The analytic hierarchy process in medical and health care decision-making: a literature review</b>, Eur. J. Operation. Res., 189 (2006), pp. 194-207</p>
<p>Best-worst scaling (types 1, 2, 3) <a href="#">a</a>, <a href="#">b</a></p>	<p>Involves respondents answering surveys that include lists of attributes or profiles and being asked to indicate the best (or most appealing/important) and the worst (or least appealing/important) of them. This method consists of three types: in type 1 a set of attributes is showed that might not reflect the characteristics of any particular treatment, of which the respondent picks the best and worst. Type 2 involves a situation in which the attributes collectively characterize a particular profile and the respondent chooses the best and worst. In type 3</p>	<p>Medical Device Innovation Consortium (MDIC), <b>Patient Centered Benefit-Risk Project Report: A Framework for Incorporating Information on Patient Preferences Regarding Benefit and Risk into Regulatory Assessments of New Medical Technology. Public Report 2015</b> T.N. Flynn, <b>Valuing citizen and patient preferences in health: recent developments in three types of best-worst scaling</b>, Expert Rev. Pharmacoecon. Outcome. Res., 10 (2010), pp. 259-267 T.N. Flynn, et al., <b>Best-worst scaling: what it can do for health care research and how to do it</b>, J. Health Econ., 26 (2007), pp. 171-189</p>

	three or more profiles are shown and the respondent selects the best and worst profiles	
Constant sum scaling <a href="#">c</a>	Constant sum scaling consists of a comparative scale where respondents are asked to allocate a fixed amount (or constant sum) of points, dollars or anything among a set of objects according to a criterion	R. Mai, S., Hoffmann <b>Taste lovers versus nutrition fact seekers: how health consciousness and self-efficacy determine the way consumers choose food products</b> , J. Consumer Behaviour, 11 (2012), pp. 316-328 C.D. Skedgel, et al. <b>Choosing vs. allocating: discrete choice experiments and constant-sum paired comparisons for the elicitation of societal preferences</b> , Health Expect., 18 (2015), pp. 1227-1240
Contingent valuation <a href="#">a</a> , <a href="#">b</a>	Method to determine the willingness to pay (WTP), where individuals are presented with a choice between not having the commodity valued and having the commodity but forgoing a certain amount of money. The money being that they are willing to forgo to have the commodity is their WTP for that commodity. WTP can be calculated directly using a threshold or indirectly using a discrete choice experiment for example	T. Bärnighausen, et al., <b>Willingness to pay for social health insurance among informal sector workers in Wuhan, China: a contingent valuation study</b> , BMC Health Service. Res., 7 (2007), p. 114 S.J. Cunningham, N.P. Hunt <b>Relationship between utility values and willingness to pay in patients undergoing orthognathic treatment</b> , Commun. Dental Health, 17 (2000), pp. 92-96
Control preference scale <a href="#">a</a>	The control preferences scale (CPS) is a method to determine the degree of control a patient wants regarding medical treatment. The preference orders are analyzed using unfolding theory to determine the distribution of preferences in different populations and the effect of covariates on consumer preferences	L.F. Degner, et al., <b>The Control Preferences Scale</b> , Can. J. Nurs. Res., 29 (1997), pp. 21-43 C.J. McPherson, et al. <b>Effective methods of giving information in cancer: a systematic literature review of randomized controlled trials</b> , J. Public Health, 23 (2001), pp. 227-234 N.B. Henrikson, et al. <b>Measuring decisional control preferences in men newly diagnosed with prostate cancer</b> , J. Psychosoc. Oncol., 29 (2011), pp. 606-618
Discrete choice experiment <a href="#">a</a> , <a href="#">b</a>	Method that utilizes an attribute-based measure of benefit, during which individuals are offered a series of hypothetical choice situations (i.e., choice sets), from which they are asked to choose between two or more profiles. There are numerous variants of discrete choice experiments. In contrast to conjoint analysis, this method relies on a theory of the behavior of human preferences [for example random utility theory (RUM)]	J.F.P. Bridges, et al., <b>Conjoint analysis applications in health – a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force</b> , Value Health, 14 (2011), pp. 403-413 V. Soekhai, et al. <b>Discrete choice experiments in health economics: past, present and future</b> , PharmacoEconomics, 37 (2019), pp. 201-226 E. Lancsar, J. Louviere, <b>Conducting discrete choice experiments to inform healthcare decision-making</b> , PharmacoEconomics, 26 (2008), pp. 661-677

		J. Louviere, et al., <b>Discrete choice experiments are not conjoint analysis</b> , J. Choice Model., 3 (2010), pp. 57-72
Measure of value <a href="#">b</a>	Method used to identify the optimal bundle of services to be provided given resource constraints. Individuals are asked to allocate a fixed amount of resources between different services. These allocations are analyzed to identify the trade-offs individuals make	M.C. Weinstein, et al. <b>Principles of good practice for decision analytic modeling in health-care evaluation: report of the ISPOR Task Force on Good Research Practices – modeling studies</b> , Value Health, 6 (2003), pp. 9-17
Outcome prioritization tool <a href="#">a</a>	Instrument that allows participants to prioritize outcomes making use of a specific tool according to the 'trade-off' principle, implying that they are willing to compromise on the less important outcomes	S.M. Case, et al. <b>Older adults' recognition of trade-offs in healthcare decision-making</b> , J. Am. Geriatr. Soc., 63 (2015), pp. 1658-1662
Person trade-off <a href="#">a</a> , <a href="#">b</a>	An extension of the time trade-off. With person trade-off an individual evaluates the health effects of interventions using persons (instead of time) as the equilibrating mechanism	C. Green <b>On the societal value of health care: what do we know about the person trade-off technique?</b> Health Econ., 10 (2001), pp. 233-243 E. Nord <b>The person-trade-off approach to valuing health care programs</b> , Med. Decision Making, 15 (1995), pp. 201-208
(Probabilistic) threshold technique <a href="#">a</a> , <a href="#">b</a>	Method that determines the maximal change in one attribute respondents are willing to accept to achieve a given change in another attribute	Medical Device Innovation Consortium (MDIC), <b>Patient Centered Benefit-Risk Project Report: A Framework for Incorporating Information on Patient Preferences Regarding Benefit and Risk into Regulatory Assessments of New Medical Technology. Public Report 2015</b> J.A. Kopec, et al. <b>Probabilistic threshold technique showed that patients' preferences for specific trade-offs between pain relief and each side effect of treatment in osteoarthritis varied</b> , J. Clin. Epidemiol., 60 (2007), pp. 929-938
Q-methodology <sup>6</sup>	Method that uses a specially designed response grid to present respondents with a set of statements and asking them to order, usually based on the extent to which they agree with them	R.M. Cross, <b>Exploring attitudes: the case for Q methodology</b> , Health Education Res., 20 (2005), pp. 206-213 J. van Exel, et al. <b>Care for a break? An investigation of informal caregivers' attitudes toward respite care using Q-methodology</b> , Health Policy, 83 (2007), pp. 332-342
Qualitative discriminant process	Method that involves a scoring and ranking process based on decision analysis technique, involving the definition of options in terms of qualitative	P.M. Mullen, <b>Public involvement in health care priority setting: an overview of methods for eliciting values</b> , Health Expect., 2 (1999), pp. 222-234

<a href="#">b</a>	categories, then deriving a numeric point estimate and finally solving a maximization problem with given constraints	
Repertory grid method <a href="#">a</a>	Method used for eliciting personal constructs (i.e., what people think about a given topic). To identify preferences overlapping and rating techniques are used	C.J. Davis, et al. <b>Communication challenges in requirements elicitation and the use of the repertory grid technique</b> , J. Comput. Inform. Syst., 46 (2006), pp. 78-86 G. Rowe, et al. <b>Assessing patients' preferences for treatments for angina using a modified repertory grid method</b> , Soc. Sci. Med., 60 (2005), pp. 2585-2595
Self-explicated conjoint <a href="#">c</a>	Method that asks explicitly about the preference for each attribute rather than the preference of several	H. Riquelme, T. Rickards, <b>Hybrid conjoint analysis: an estimation probe in new venture decisions</b> , J. Business Venturing, 7 (1992), pp. 505-518
Standard gamble <a href="#">a</a> , <a href="#">b</a>	Method in which respondents are asked to choose between a certain outcome and a gamble that might result in either a better outcome with a probability P or a worse outcome than the original with a probability 1-P	A. Gafni, <b>The standard gamble method: what is being measured and how it is interpreted</b> , Health Services Res., 29 (1994), pp. 207-224 T. Morimoto, T. Fukui <b>Utilities measured by rating scale, time trade-off, and standard gamble: review and reference for health care professionals</b> , J. Epidemiol., 12 (2002), pp. 160-178
Starting known efficacy <a href="#">a</a>	Method similar to (probabilistic) threshold techniques, but with a specific known starting point. This method is specifically used within the context of the medical product lifecycle	M. Man-Son-Hing, et al. <b>Warfarin for atrial fibrillation. The patient's perspective</b> , Arch. Intern. Med, 156 (1996), pp. 1841-1848
Swing weighting <a href="#">b</a>	Method for setting the weights in which a decision-relevant range is specified for each attribute, and the impact of 'swinging' the attribute through that entire range of values is assigned a weight relative to the impact of swinging the attribute with the largest weight	M. Ryan, et al. <b>Eliciting public preferences for healthcare: a systematic review of techniques</b> , Health Technol. Assess., 5 (2001), pp. 1-186 Medical Device Innovation Consortium (MDIC), <b>Patient Centered Benefit-Risk Project Report: A Framework for Incorporating Information on Patient Preferences Regarding Benefit and Risk into Regulatory Assessments of New Medical Technology. Public Report 2015</b>
Test trade-off <a href="#">c</a>	Method that can be regarded as an extension of the time trade-off that is specifically used to evaluate a new biomarker by using risks (instead of time) as the equilibrating mechanism	S. G. Baker, B.S. Kramer, <b>Evaluating surrogate endpoints, prognostic markers, and predictive markers: some simple themes</b> , Clin. Trials, 12 (2014), pp. 299-308



		S.G. Baker, et al., <b>Evaluating a new marker for risk prediction using the test tradeoff: an update</b> , Int. J. Biostat., 8 (2012), pp. 1-17
Time trade-off <a href="#">a</a> , <a href="#">b</a>	Method that presents individuals with a choice between living for a period in a specified, but less than perfect, state versus having a healthier life for a period of time, where time is varied until the respondent is indifferent to the alternatives	<p>Medical Device Innovation Consortium (MDIC), <b>Patient Centered Benefit-Risk Project Report: A Framework for Incorporating Information on Patient Preferences Regarding Benefit and Risk into Regulatory Assessments of New Medical Technology. Public Report 2015</b></p> <p>T. Arnesen, M. Trommald, <b>Are QALYs based on time trade-off comparable?—A systematic review of TTO methodologies</b>, Health Econ., 14 (2005), pp. 39-53</p> <p>J. Brazier, et al., <b>A review of the use of health status measures in economic evaluation</b>, J. Health Services Res. Policy, 4 (1999), pp. 174-184</p>
Visual analog scale <a href="#">a</a> , <a href="#">b</a>	A self-reporting instrument consisting of a line of predetermined length that separates extreme boundaries of the phenomenon being measured	<p>A. Holdgate, et al., <b>Comparison of a verbal numeric rating scale with the visual analogue scale for the measurement of acute pain</b>, Emerg. Med., 15 (2003), pp. 441-446</p> <p>P. Räsänen, et al., <b>Use of quality-adjusted life years for the estimation of effectiveness of health care: a systematic literature review</b>, Int. J. Technol. Assess. Health Care, 22 (2006), pp. 235-241</p>

Source: adapted from Table 1, [Soekhai et al. 2019](#).

Footnotes:

Identified in systematic review (19 methods).

Identified through analysis of previous preference method reviews (23 methods).

Identified with expert consultations (4 methods).