PREFER RECOMMENDATIONS

Why, when and how to assess and use patient preferences in medical product decision-making

prefer.

PREFER Recommendations

Why, when and how to assess and use patient preferences in medical product decision-making

Two key uses of patient preference information are to inform the choice of patient-relevant endpoints by providing information on the relative importance of what matters to patients, and to provide information on patients' views about the acceptability of trade-offs between treatment characteristics, or other attributes of treatments or health interventions.

The PREFER Recommendations provide expert and evidence-based guidance from six years of research on when and how to design and conduct a patient preference study.

The PREFER project was a joint undertaking by 33 public and private partners with more than 130 people representing academic institutions, pharmaceutical companies, health technology agencies, and patient organisations in European countries and in the US.

Foreword

Between October 2016 and May 2022, the Innovative Medicines Initiative (IMI) PREFER (Patient Preferences in Benefit–Risk Assessments during the Medical Product Lifecycle) consortium developed expert and evidence-based recommendations to guide industry, regulatory authorities, health technology and reimbursement agencies on when and how patient preferences can be assessed and used to inform medical product decision-making.

Patient preference information includes the relative importance of what matters most to patients, as well as the trade-offs that patients are willing to make between treatment characteristics, or other attributes of treatments or health interventions. The increased focus on patient preferences stands within the global aim of strengthening patient-centric decision-making throughout the life cycle of medical products, including drugs, biologics and vaccines, and medical devices.

The PREFER project was a joint undertaking by 33 public and private partners with more than 130 people representing academic institutions, pharmaceutical companies, health technology agencies, and patient organisations in European countries and in the US. It was a six-year project funded by IMI and by in-kind contributions from all industry partners. It originated from discussions within the International Conference for Harmonization of Technical Requirements for Pharmaceuticals for Human Use on how to improve the assessment of benefits and risks, and revealed the importance of having a strong focus on the patient perspective in these assessments.

In close dialogue with patients and their organisations, with regulatory authorities, health technology assessment (HTA) bodies and reimbursement agencies, and industry – as well as with the wider academic research community – the PREFER project has identified over 100 different research questions related to the assessment and use of patient preference studies, and 32 different qualitative and quantitative methods for the elicitation of patient preferences. The project prioritised 17 of these questions and 13 methods that have been tested in 10 prospective clinical case studies, in addition to the assessment of the results of 33 historical case studies. Key insights on how to involve patients in the design and conduct of patient references were gained over time.

The data supporting the PREFER recommendations include stakeholder concerns and desires – revealed as qualitative evidence through literature reviews, interviews, and focus groups in seven countries – and quantitative evidence through surveys and clinical preference studies including cross-study evaluations. Results are publicly available through open access publications and platforms, conference proceedings, webinars, Twitter campaigns, and YouTube videos.

To further support the impact and increased implementation of patient preference studies for medical product decision-making, the PREFER project initiated a joint qualification procedure with the European Medicines Agency (EMA) and the European Network for Health

Technology Assessment (EUnetHTA). The goal was to attain a Qualification Opinion for a framework that supports the development and specification of study purpose and design, conduct, analysis, and reporting of patient preference studies, as well as for 'Points to Consider' on method selection, together with additional details of five key quantitative methods. The EMA accepted the application, and the full Qualification Opinion by the Committee for Medicinal Products for Human Use (CHMP) along with public comments is the available on the **EMA website**.

The PREFER recommendations are complementary to other initiatives and frameworks, such as the work carried out by the International Academy of Health Preference Research and the Professional Society for Health Economics and Outcomes Research.

We expect the PREFER recommendations to:

- stimulate the design and execution of more patient preference studies relevant for medical product decision-making

- stimulate the publication of patient preference study results

- increase the collaboration and interactions among study sponsors, academia, patients, and decision-makers through conceptualisation of the research question, conducting patient preference studies and using patient preference information in decision-making

- be taken as an invitation for future work on research questions that have been identified by PREFER but could not be addressed in the project.

We thank the entire PREFER consortium, the Scientific Advisory and Ethics Board members, IMI and all external stakeholders from patient organisations, regulatory authorities, HTA bodies, and reimbursement agencies for their outstanding support and input throughout the project (see the **Annex** for the full list of contributors).

Conny Berlin, Project Leader

Mats Hansson, Coordinator

Rebecca Noel, Co-Project Leader

Isabelle Huys, Co-Coordinator

What informed the PREFER recommendations and what are the other different outputs of the PREFER project?

The PREFER recommendations were informed by the following four steps (**Figure i**).

Stakeholder interviews

The PREFER project started from an in-depth understanding of the concerns and needs of stakeholders involved in medical product decision-making – patients, patient organisations, industry, regulators, HTA bodies and payers, and clinicians – towards patient preference study methodology and use, and consisted of 6 literature reviews, 143 individual interviews, and 8 focus group discussions.(1-9) These revealed that patient preference studies may inform decisions across the medical product life cycle, including the earliest decisions on unmet needs for patients in terms of their disease and treatment. Stakeholders also agreed that preference studies could help guide their decisions on which endpoints could be pursued and included in evidence generation in clinical trials, real-world evidence studies, and post-marketing evidence generation plans. Stakeholders further agreed that preference studies could help provide information on the relevance of these endpoints, what treatment effect would be considered meaningful to patients, marketing authorisation applications (MAAs), and HTA or payer decisions.

Preference methods categorisation and appraisal

This step was undertaken to explore how stakeholders may use findings from preference studies to inform decision-making across the medical product life cycle, as well as to provide insights into existing preference study methods and how to appraise them. The available literature on different study methods was reviewed, and relevant stakeholders (e.g. preference study researchers and experts) were interviewed to convey insights into how to select and appraise different methods for conducting patient preference studies.(1-9)

Identification and appraisal of psychological and educational feature methods

An assessment was also made of different options to measure psychological constructs that may be associated with preference heterogeneity. Also, the educational needs of patients participating in preference studies were identified and addressed.

Ten prospective case studies across different diseases

- Research from the previous three steps informed the design and conduct of 10 subsequent prospective case studies eliciting patient preferences in a subsequent phase so they would be as useful as possible for these stakeholders in their decision-making.(10-31) In these case studies, different qualitative methods (e.g. interviews, literature reviews, focus group discussions) and quantitative methods (e.g. discrete choice experiment [DCE], swing weighting, threshold technique, best-worst scaling [BWS] case 1 and 2) were used to explore and elicit patient preferences. The qualitative

methods were used to gain in-depth information on why patients would or would not take a treatment, what key positive and negative treatment outcomes (attributes) were most important to them, and what they would likely consider to be the best and worst treatment outcomes. The quantitative methods aimed to quantify the importance of the treatment attributes identified in the qualitative phases and assess preference heterogeneity through statistical analysis (e.g. how attribute weights are influenced by patient characteristics such as treatments used and experience with the disease). All case studies aimed to involve patients and patient organisations closely throughout the study design and conduct, and to elicit preferences in a variety of disease and treatment areas, particularly those where there was little or no experiential and evidence-based information from patients regarding what matters most to them. The disease areas for the prospective case studies were rheumatoid arthritis,(11, 17, 18) neuromuscular disorders (NMDs),(12, 15) lung cancer,(13, 14, 20, 22) haemophilia,(10, 16, 21) diabetes, multiple myeloma,(19) chronic pain,(25, 27) myocardial infarction,(26, 31) and chronic obstructive pulmonary disease (COPD).(28-30)
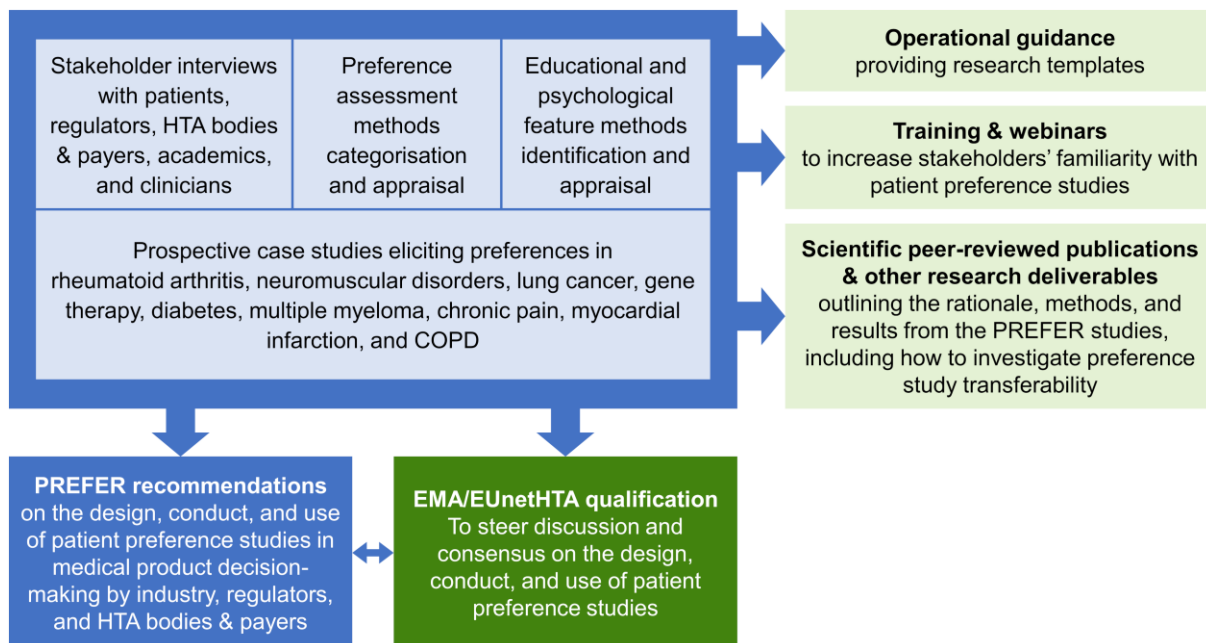

Figure i. Inputs and outputs of the PREFER project.

In the **PREFER recommendations**, preference study experience gained from the 10 case studies was supplemented with information from the published literature on experiences, good research practices, and quality criteria within relevant research disciplines and disease areas. This helped the development of experience-based recommendations applicable to qualitative approaches and quantitative survey-based approaches in preference studies across disease areas. PREFER encourages consistent and appropriate application of patient

preference studies as complimentary to other evidence in informing regulatory and HTA body and/or payer guidelines and scientific advice.

Alongside development of the recommendations, PREFER undertook a parallel **EMA/EUnetHTA Qualification Opinion procedure**, which enabled feedback from regulatory authorities and HTA bodies and payers to be incorporated into the recommendations.(32) Additionally, as part of the methods qualification procedure, the EMA undertook a public consultation to obtain input on the content of the qualification package from PREFER i.e. content on the PREFER framework for designing, conducting, and using patient preference studies as well as points to consider for method selection and details on five preference methods. Topics included in both, the qualification package and the recommendations, are summarised in sections 1 to 5 of this document. Sections 6-8 of this document, on educational materials and psychological constructs for inclusion in patient preference studies, were not included in the EMA/EUnetHTA Qualification Opinion document.(32)

PREFER also developed specific tools to help interested stakeholders design, conduct, and assess patient preference studies: **operational guidance** including research templates (e.g. informed consent templates, information sheets) and training materials, namely **webinars on patient preference studies and methods**. PREFER published its findings – including the initial qualitative stakeholder studies and the 10 prospective case studies – in more than 30 scientific peer-reviewed publications. Several research deliverables describe the rationale, methods, and results of the different research efforts funded and conducted within the context of the PREFER project.

What to find in the PREFER recommendations and how can they be used by stakeholders?

The PREFER recommendations consist of eight sections, plus a related annex document (**Figure ii**), that provide stakeholders with evidence-based insights into how patient preference studies should be designed, conducted, and used to inform decision-making throughout the medical product life cycle.

- **Section 1** outlines the objective of the recommendations and introduces the different aspects and considerations for designing and conducting patient preference studies.

- **Section 2** explains what information can be obtained from patient preference studies, and why and when these studies can be conducted and applied to medical product decision-making by industry, regulators, and HTA bodies and payers.

- **Section 3** describes the PREFER framework for patient preference studies. The PREFER framework aims to inform study research teams on key considerations when designing, conducting, and applying the results of a fit-for-purpose preference study, and guide decision-makers when assessing and using preference study results to inform medical product decision-making.

- **Section 4** focuses on the involvement of patients and other stakeholders, such as regulators and HTA bodies, in the design, conduct, and analysis of these studies so that the information they generate is meaningful for the patient population and useful for decision-makers.

- **Section 5** focuses on different qualitative and quantitative preference methods and describes how stakeholders can select an appropriate method for a given context.

- **Section 6** offers insights into when and how the psychological characteristics of participants, in addition to demographic and clinical variables, should be investigated so that preference heterogeneity among patients can be explored and understood.

- **Section 7** provides information on how to develop supporting materials so that patients can be educated about the questions and elements they are asked to evaluate and can make informed choices that will ensure validity and meaningfulness of the results.

- **Section 8** provides insights into important avenues for further research, including recommendations for which topics and research questions should be explored and incentivised to further increase the quality of patient preference studies and gain wider consensus by all stakeholders involved.

SECTION 1	**Introduction**	
SECTION 2	**Value of patient preferences**	Inform stakeholders about why and when patient preference studies may benefit their decisions
SECTION 3	**PREFER framework**	Provide clear and step-wise insights into how to design, conduct, and evaluate patient preference studies
SECTION 4	**Involvement of patients & other stakeholders**	Provide insights into how patients and other stakeholders may contribute to patient preference study design and conduct to help ensure the studies provide useful information for patients and decision-makers
SECTION 5	**Preference methods**	Help guide preference study method selection, a crucial step for patient preference studies that require considering multiple factors
SECTION 6	**Psychological constructs**	Understand how participants' psychological characteristics may be assessed to understand how they may influence patients' answers in patient preference studies
SECTION 7	**Educational materials**	Explain which tools (e.g. survey component and multimedia) may help ensure patients' understanding in patient preference studies
SECTION 8	**Areas for future research**	Describe areas for future work on research questions that have been identified by PREFER but could not be addressed in the project

Figure ii. Sections of the PREFER recommendations.

Table of Contents

Glossary

Term	Definition
Attribute	A feature or characteristic – such as efficacy, effectiveness, safety, means of administration, duration of effect, duration of use, or burden on patients or caregivers – that may affect preferences for a medical product, which in turn could inform decision-making in the medical product life cycle.
Benefits	The favourable effects of a medical product. Types of benefit include clinical benefit but may also include other 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.(33)
Best-worst scaling (BWS)	A survey method in which respondents are asked to indicate which attributes in a set (BWS case 1), attribute level in a profile (BWS case 2), or profile in a choice task (BWS case 3) is best or most preferred and which is worst or least preferred. The set of attributes or profiles of attribute levels are determined by an experimental design. The pattern of choices over a series of these questions yields data to rank and score the importance of each attribute (case 1) or attribute level (case 2 and 3).
Case study	A patient preference study that aimed to investigate one or more PREFER research questions.
Choice share	In population terms, a choice share reflects a hypothetical estimate of the proportion of the population that, in the scenario of being provided with 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.
Clinical overview	A document written by a company and submitted to the regulatory agency in support of an application for regulatory authorisation. The clinical overview should provide a critical analysis of the clinical data, describe the strengths and limitations of the development program and study results, analyse the benefits and risks of the medical product in its intended use, and describe how the clinical results support critical parts of the Summary of Product Characteristics (SmPC), labelling and package leaflet.

Term	Definition
Discrete choice experiment (DCE)	A survey method in which respondents are asked to choose among a set of profiles in a series of questions. Each profile is defined by attributes (e.g. benefits, risks, mode of administration, cost), each of which can take on varying levels (e.g. high, medium, low). Each profile and the set of profiles in each question are determined by an experimental design. The pattern of choices over a series of these questions yields data to estimate the relative preference weight for each attribute level.
Decision	A judgment, conclusion, or determination reached after consideration. A decision is a response in a situation that is composed of three parts: • there is more than one possible course of action in the choice set • the decision-maker can form expectations about the outcomes that follow from each course of action • the consequences of the outcomes can be assessed relative to current goals and values.
eCTD (electronic Common Technical Document)	The structure (as defined in the ICH M2 guidance) that pharmaceutical companies are required to use when submitting information to regulatory agencies.
Educational materials	The part of the survey or instrument that explains the context of the study, treatment alternatives, the attributes and levels, and how to complete the choice tasks.
Fit-for-purpose	The level of validation associated with a method or tool sufficient to support its context of use.
Framework	A set of principles, guidelines, and tools, or a process that frames decision-making or certain activities. The PREFER framework includes considerations to guide decisions on the design, conduct, and use of patient preference studies that aim to inform medical product decision-making.
Maximum acceptable risk (MAR)	The greatest increase in probability or magnitude of a harm for which a patient would accept a given benefit.(34)
Medical product	Any product used to diagnose or treat patients, including medical products, devices, and services.

Term	Definition
Medical product life cycle	The development, authorisation, and post-authorisation phase of a medical product can be divided into multiple different steps, the sum of which is called 'life cycle'. The medical product life cycle herein is defined as the lifecycles of drugs, biologics, and medical devices.
Minimum acceptable benefit (MAB)	The smallest increase in probability or magnitude of a benefit for which a patient would require to offset a given risk. It is synonymous with minimum required benefit.(34)
Patient preferences	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. Of note, the term 'patient preferences' is used within this document to describe patients' preferences (referring to the preferences of multiple patients rather than the preferences of an individual patient).(35)
Patient-relevant endpoint	An endpoint that is meaningful to the patient. Examples include endpoints related to how a patient feels or functions in daily life.
Patient-reported outcome (PRO)	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.(36)
Patient research partner	A patient, informal carer, or patient advocacy organisation representative with experience and/or knowledge of the disease, who serves as a member of a research team.
Patient preference sensitive decisions	Decisions where evidence from patient preference studies is particularly valuable, namely when: • it is unclear what are the most important disease or medical product characteristics to patients; these can include existing or potential future characteristics (e.g. actual/hypothetical treatment outcomes, and mode of treatment administration); *or* • there are multiple treatment options and no option is clearly superior or has a clear added value for all patients; *or* • the evidence supporting one option over another is very uncertain or variable, and patients' tolerance for this uncertainty might impact their decisions; *or* • there is potential for considerable heterogeneity in views between patients or between patients and other stakeholders.(37)
Psychological construct	A psychological characteristic that is abstract and latent rather than concrete and observable.(38)

Term	Definition
Psychological instrument	Any tool, device, or other means by which researchers assess or gather data about psychological characteristics of people.(39)
Qualitative preference (exploration) method	A method that collects descriptive data through the application of qualitative research techniques (e.g. interviews, focus groups), through observation of decisions made by patients, or through phenomenon observation, examining the subjective experiences and decisions made by patients.(5)
Quantitative preference (elicitation) method	A method collecting quantifiable data that can be reported through statistical inferences or analysis.(5)
Quality-adjusted life years (QALYs)	A measure of the state of health of a person or group in which the benefits in terms of years of life gained are adjusted to the quality of life in these life years gained.(40)
Relative effectiveness assessment (REA)	A type of assessment that evaluates the effectiveness of a new technology compared with alternative treatments.(41)
Risks	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. a 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. The term 'risk' includes both the probability of the unfavourable effect and the unfavourable effect itself.(42)
Study participant	A person who voluntarily agrees to participate in a research study. The data provided by a study participant in a research study will help answer the research question.(43)
Swing weighting	A survey method in which respondents are asked to rank and weigh different attributes. These ranks and weights describe the relative importance of the improvement of an attribute from its worst possible level to its best possible level, compared to the improvement from worst to best possible levels of the other attributes.

Term	Definition
Threshold technique	A survey method that quantifies preferences between two healthcare options from the perspective of a patient. The process is implemented by asking a patient to choose between two alternatives and systematically varying the level of a key attribute of one alternative until the patient indicates the level at which he or she is indifferent to 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.
Trade-off	A measure of the extent to which a change in the level of one attribute of a medical product is offset by a change in another attribute of that medical product.(34)
Treatment characteristics	These include both the favourable health and non-health effects of a medical product (e.g. symptom reduction, higher convenience), unfavourable effects (e.g. adverse events, abuse) of medical product, and uncertainties related to these effects. Treatment characteristics can include aspects related to existing and potential future medical products such as the actual and hypothetical outcomes of a medical product. Treatment characteristics may be assessed in a patient preference study through the inclusion of specific attributes.
Unmet patient needs	Therapeutic and other needs related to a patient's health condition that are currently not met according to the patients. Therapeutic needs are needs for (better) treatment as perceived by patients, either because there is currently no available reimbursed treatment or care or because available reimbursed treatments are insufficiently effective. Other unmet patient needs can relate to both healthcare use (e.g. financial/geographical access to care), relations with healthcare providers, and broader needs (e.g. social support, information, education, and spiritual needs).(44)
Unmet therapeutic needs	Needs as perceived by the patients that are not met by currently available reimbursed treatments or care, either because they do not exist or because they are insufficiently effective.(44)

Abbreviations

ACA	adaptive conjoint analysis
AHP	analytical hierarchy process
AOP	allocation of points
BIBD	balanced incomplete block design
BIP-Q	Brief Illness Perception Questionnaire
BRA	benefit–risk assessment
BWS	best-worst scaling
CBER	Center for Biologics Evaluation and Research
CERSI	Center of Excellence in Regulatory Science and Innovation
CDRH	Center for Devices and Radiological Health
CHMP	Committee for Medicinal Products for Human Use
COA	clinical outcome assessment
COPD	chronic obstructive pulmonary disease
CPIM	critical path innovation meeting
CPS	control preferences scale
CSS	constant sum scaling
CV	contingent valuation
DCE	discrete choice experiment
EMA	European Medicines Agency
EPAR	European Public Assessment Report
EUnetHTA	European Network for Health Technology Assessment
EUPATI	European Patients' Academy on Therapeutic Innovation
FDA	Food and Drug Administration
FDR	first-degree relative
HRQoL	health-related quality of life
HTA	health technology assessment
ICH	International Conference on Harmonization
IMI	Innovative Medicines Initiative
IQWIG	German Institute for Quality and Efficiency in Health Care

ISPOR	Professional Society for Health Economics and Outcomes Research
IRB	institutional review board
MAA	marketing authorisation application
MAR	maximum acceptable risk
MCDA	multi-criteria decision analysis
MDIC	Medical Device Innovation Consortium
MOV	measure of value
MRB	minimum required benefit
NICE	National Institute for Clinical Excellence
NSCLC	non-small cell lung cancer
NMD	neuromuscular disorder
OMEP	orthogonal main effects plans
OPT	outcome prioritisation tool
PARADIGM	Patients Active in Research and Dialogues for an Improved Generation of Medicines
PFMD	Patient Focused Medicines Development
PREFER	Patient Preferences in Benefit–Risk Assessments during the Medical Product Lifecycle
PRFT	prophylactic factor replacement therapy
PRO	patient-reported outcome
PTO	person trade-off
PTT	probabilistic threshold technique
QALY	quality-adjusted life year
QDP	qualitative discriminant process
QM	Q-methodology
REA	relative effectiveness assessment
RGP	repertory grid method
SEC	self-explicated conjoint
SG	standard gamble
SKE	starting known efficacy
SMAA	stochastic multicriteria acceptability analysis
UK	United Kingdom

US	United States
TeTO	test trade-off
TiTO	time trade-off
VAS	visual analogue scale
WP	work package
WTP	willingness-to-pay

1   Introduction

Key message

Patient preferences are regarded as important in decision-making about medical treatments, and patient preference studies can make use of robust, scientific, and structured methodologies.

The PREFER recommendations address gaps in awareness, and provide guidance, on how to conduct, assess and incorporate the outcomes of patient preference studies in decision-making throughout the medical product life cycle, specifically in product development decisions, regulatory decision-making, marketing authorisation, and HTA or reimbursement decision-making.

The PREFER recommendations provide a comprehensive framework, from ideation to communication of results, that includes a distinctive depth on how to design a preference study and apply its results relative to existing frameworks and guidance documents. The recommendations also uniquely emphasise the need to engage all stakeholders throughout the preference study process and incorporate learnings on using educational materials and psychological assessments. Much of the content in Sections 1 to 5 was included in the application for a Qualification Opinion.

The recommendations in this document are intended to educate and advise stakeholders – industry, regulators, health technology assessment (HTA) bodies and payers, academia, physicians, and patient organisations – about many aspects of patient preference studies, specifically:

- what a patient preference study is

- why such a study might be conducted

- how the results can be used to inform decision-making

- which considerations should be taken into account when designing, conducting, using, and evaluating a preference study

- what the points to consider are for selecting a method that would be appropriate for a given circumstance

- how the measurement of relevant psychological constructs addresses the challenge of preference heterogeneity

- how the use of educational materials (type and content) can support patients' understanding of preference study material.

There is broad consensus that the needs and experiences of patients should be considered in decision-making across the medical product life cycle, to ensure that medical products are developed and evaluated to meet the needs of patients. The role of patient preferences is, therefore, becoming increasingly important for all stakeholders involved in decision-making throughout the medical product life cycle, especially to guide the research and development of new drugs and devices,(3, 45-47) benefit–risk assessments,(48) and value assessment/reimbursement.(49, 50) Furthermore, the EMA, the United States (US) Food and Drug Administration (FDA), HTA bodies/payers, and industry have all expressed interest in collecting such information to inform their decision-making.(34, 37, 51-54) One of the underlying reasons for the increasing interest in patient preferences is that these provide unique information on patients' needs.(34, 55-59) Patient preferences can also highlight the extent to which existing treatment options fail to meet the needs of patients, which may, in turn, influence the development of new outcomes and/or the assessment of medical products by regulators and HTA bodies/payers.

Patient preferences represent one type of patient input (**Figure 1-1**) to help inform medical product decision-making, and have been defined by the FDA as 'patient preference information':

> *"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."* (35)

Although input on patient preferences is normally provided by the patients themselves, there may be specific circumstances when this is not possible, such as when the patient is very young or (cognitively) impaired (**Section 3.3.2.2**). In such cases, 'patient preferences' might be assessed by parents or carers as proxies. When referring to patients participating in preference studies, these could include those with the disease of interest, healthy participants (e.g. at risk of the disease of interest), or caregivers (see also **Section 4.1.3**).

22
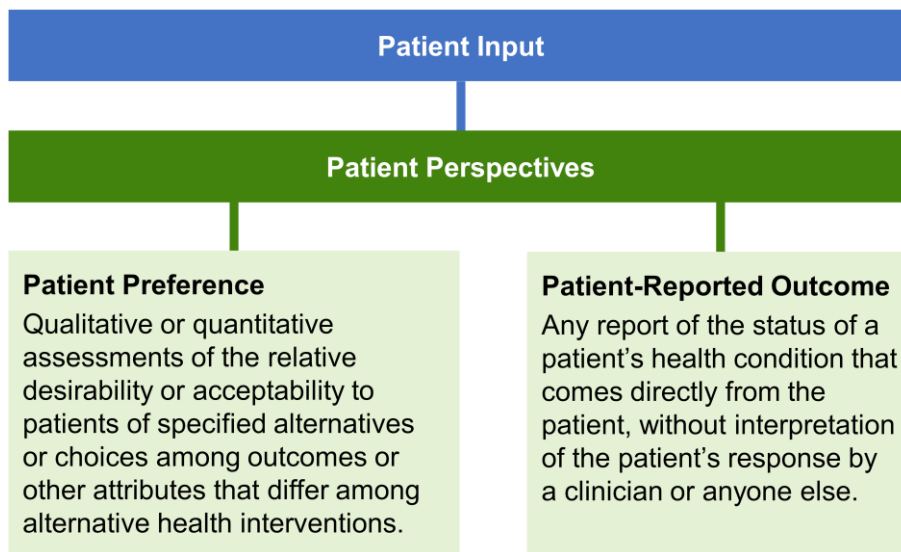
Figure 1-1. Patient preferences as a type of patient input.

Adapted from FDA Final Patient Preference Guidance, 2016(37) and FDA Patent-Focused Drug Development.(60)The term 'Patients' Perspectives' in the Center for Devices and Radiological Health Guidance is analogous to Patient Experience Data in 21st Century Cures.

For regulators and HTA bodies/payers to incorporate patient preferences in decision-making, a common understanding of patient preferences and approaches for measuring them during the development and assessment of medical products is needed. On the regulatory level, the FDA has issued guidance on using patient preference information relating to medical devices that has been generated from patient preference studies,(35) and has developed several guidance documents related to patient-focused drug and medical device development.(61, 62) The Council for International Organizations of Medical Sciences is also developing guidance in this context, and on a national level, several HTA bodies have developed processes to obtain input from patients and patient groups. Previous efforts by academic groups and HTA bodies have outlined how and when to involve patients in HTA processes, acknowledging that patient involvement is not a substitute for the use of patient preference data collected through scientifically valid patient preference studies.(63, 64)

Despite these developments, there is no comprehensive framework on how to design, conduct and analyse a scientifically robust patient preference study, or how to assess its validity. There is a corresponding lack of clarity on how and when to incorporate the results of a patient preferences in decision-making throughout the medical product life cycle – specifically in product development decisions, regulatory decision-making, and HTA or reimbursement decision-making.

Thus far, there are only a few published examples on the use of patient preferences in regulatory decisions, and these have mainly come from the US with very few examples from the European Union (EU) (**Annex Table A1-1**). According to the FDA Patient Experience

Data review, only 3% of approvals of a new molecular entity included a patient preference study based on applications received between 12 June 2017 and 12 June 2020.(65) Other examples are the regulatory benefit–risk assessment of a medical device for obese patients (55) and the evaluation of different routes of administration of medicines.(66) Additionally, there are two US regulatory decisions that were informed by unpublished preference studies: the NxStage home haemodialysis label expansion and Tula ear tubes performance criterion for the pivotal clinical study.(67) At the HTA or reimbursement level, two patient preference studies have been performed by the German Institute for Quality and Efficiency in Health Care (IQWiG),(68-71) with the aim of understanding the usefulness and feasibility of assessing and including patient preferences for informing the HTA of medical products.

The use of patient preference studies can be particularly helpful in decision-making contexts that are especially sensitive to the preferences of patients (see **Section 2.1** and the **glossary**), called patient preference-sensitive decisions. The latter was first described in the Medical Device Innovation Consortium's (MDIC) report (34) and the FDA's Center for Devices and Radiological Health guidance,(35) and PREFER has identified a need for a greater understanding of this concept. In essence, in certain situations decision-makers may feel the need to better understand which characteristics or features related to a disease or to a medical product matter to patients, how much they matter, which trade-offs between treatment characteristics they consider acceptable, and patients' tolerance level for uncertainty. Not all patients have the same preferences, so assessing and addressing heterogeneity requires careful consideration, as explained later in this document (**Section 3.5.1**).

PREFER's initial research (1, 72) revealed further detail on stakeholder expectations about the feasibility of generating and using patient preference information, and its impact on decision-making. As a general theme, the value of patient preference studies would depend on how, by whom, and when the data are generated. Specific areas of concern were the lack of:

- familiarity and experience among stakeholders regarding patient preference studies and how to critically assess the quality of a patient preference study (e.g. to identify potential bias or methodological errors)

- a clear, practical framework for defining the research question, the organisation (team, timing) and the design, conduct, and application of a patient preference study

- awareness among stakeholders (regulators, HTA bodies/payers) regarding the overall added value of patient preference studies and the type of information the studies may contribute to their decisions (e.g. about patients' unmet needs, disease and treatment experiences)

- knowledge among stakeholders about the situations in which a patient preference study is likely to add value to decision-making

- patients' understanding of preference studies and knowledge about developing questions that are easily understood by patients so that they can make informed choices in such studies

- methodological understanding and familiarity by all stakeholders regarding the selection and use of fit-for-purpose preference elicitation methods

- clarity on how patient preferences are influenced by patient characteristics (preference heterogeneity) and how to assess this in patient preference studies

- clarity on how stakeholder interaction could occur during the design and conduct of a patient preference study, as well as in the adoption and communication of the results.

A clear, dedicated framework for patient preference studies is therefore key for implementing patient preferences in medical product decision-making throughout the medical product life cycle. Such a framework should include points to consider when selecting high quality and fit-for-purpose methodologies, and guidance on the level of interaction required between stakeholders. The PREFER recommendations have, at their core, a framework (**Section 3**) that provides insights into how to design, conduct and incorporate patient preferences in decision-making. A further key section of this document is the checklist of points to consider when seeking scientific advice from regulators and HTA bodies (**Box 4-3**).

The PREFER recommendations are the final outcome of the six-year PREFER project on patient preferences in benefit–risk assessments during the medical product life cycle, supported by the Innovative Medicines Initiative (IMI) as a public–private partnership between industry, academia, patient organisations, and other non-profit institutions (**Annex Section A1.1**). PREFER, therefore, takes the first steps in creating an evidence-based approach to patient preference studies that aims to provide patient-based evidence that is useful for medical product developers in development decisions (such as clinical trial programs), regulators for endpoint selection, development program and benefit–risk decision-making, and HTA bodies/payers for value assessments and reimbursement decisions.

2   Value of patient preferences

<table>
<tr><td>

Key message

Patient preferences are useful for informing preference-sensitive decisions during multiple phases of the medical product life cycle.

The results of a single patient preference study can inform decisions of multiple stakeholders, such as industry, regulators, HTA bodies and payers.

Patient preference data is intended to inform decision-making as supplementary information to other evidence generated by clinical data.

Two especially valuable use cases are where patient preference studies inform:

- the choice of patient-relevant endpoints by showing which characteristics of a medical product or disease are most important to patients (qualitative assessment), and how much they matter (quantitative assessment)
- the acceptability to patients of trade-offs between the various medical product characteristics.

</td></tr>
</table>

Evidence from patient preference studies is particularly valuable when there is a need to better understand which disease or medical product characteristics matter most to patients, how much these characteristics matter, which trade-offs between medical product characteristics patients consider to be acceptable, and patients' views towards uncertainties about the effects of medical products. Decisions that relate to these situations are called 'preference sensitive' and occur when:(37)

- it is unclear what are the most important disease or medical product characteristics to patients; these can include existing or potential future characteristics (e.g. actual/hypothetical outcomes, and mode of treatment administration)

- there are multiple treatment options and no option is clearly superior or has a clear added value for all patients

- the evidence supporting one option over another is very uncertain or variable, and patients' tolerance for this uncertainty might impact their decisions

- there is potential for considerable heterogeneity in views between patients or between patients and other stakeholders.

In the Qualification Opinion,(32) the CHMP states that its current view is that "the value of this definition of preference-sensitive decisions is 'still to be established'" and "any eventual judgment of whether they would apply in a certain situation would remain subjective". Therefore, it is important to determine, through early consultation with regulators and/or HTA

bodies (as per Section 4.2), that the decision to be informed by a patient preference study is viewed as preference-sensitive by the decision-maker.

2.1 Situations where a patient preference study can inform decision-making

Figure 2-1 highlights the stages and potentially preference-sensitive decisions during the medical product life cycle that can be informed by patient preference studies. A single research question can inform decisions at different life cycle stages and can be of relevance to multiple decision-makers. Certain decisions can be informed by patient preference data alone while other types of decision require both preference and clinical data (**Figure 2-1**, with additional discussion in **Section 3.4**).

Preference-sensitive research questions about the acceptability of uncertainty are briefly discussed in **Section 3.4** and noted as an area of further research in **Section 8**.

Of note, the CHMP Qualification Opinion (32) identifies both "endpoint selection" and "identify and value trade-offs for benefits and risks" as potential applications of preference studies.

2.1.1 Early clinical development

Understanding patients' views on which disease and medical product characteristics matter most to them can help identify areas of unmet patient need and inform decisions about which medical products to develop.(73)

A preference study to assess the relative importance of disease or treatment characteristics during early phase:

- can help identify unmet needs with respect to treatment options, and so inform decisions on which medical product to develop (e.g. informing target product profiles)

- can be instrumental in identifying patient-relevant endpoints and support the inclusion of these endpoints in clinical trials.(1, 3, 6, 9, 37, 74)

Preference studies to understand the trade-offs patients are willing to make:

- can be conducted to provide information on the relative importance of particular treatment characteristics over others and/or the risks patients are willing to trade-off for certain benefits of a medical product. For new medical products, risks are typically not well-known during early phase. Thus, preference studies for informing early phase decisions could incorporate hypothetical side-effects traded against the anticipated benefits.

- can inform decisions about a meaningful effect size, which could be especially relevant for novel indications or endpoints (**Section 3.4**).
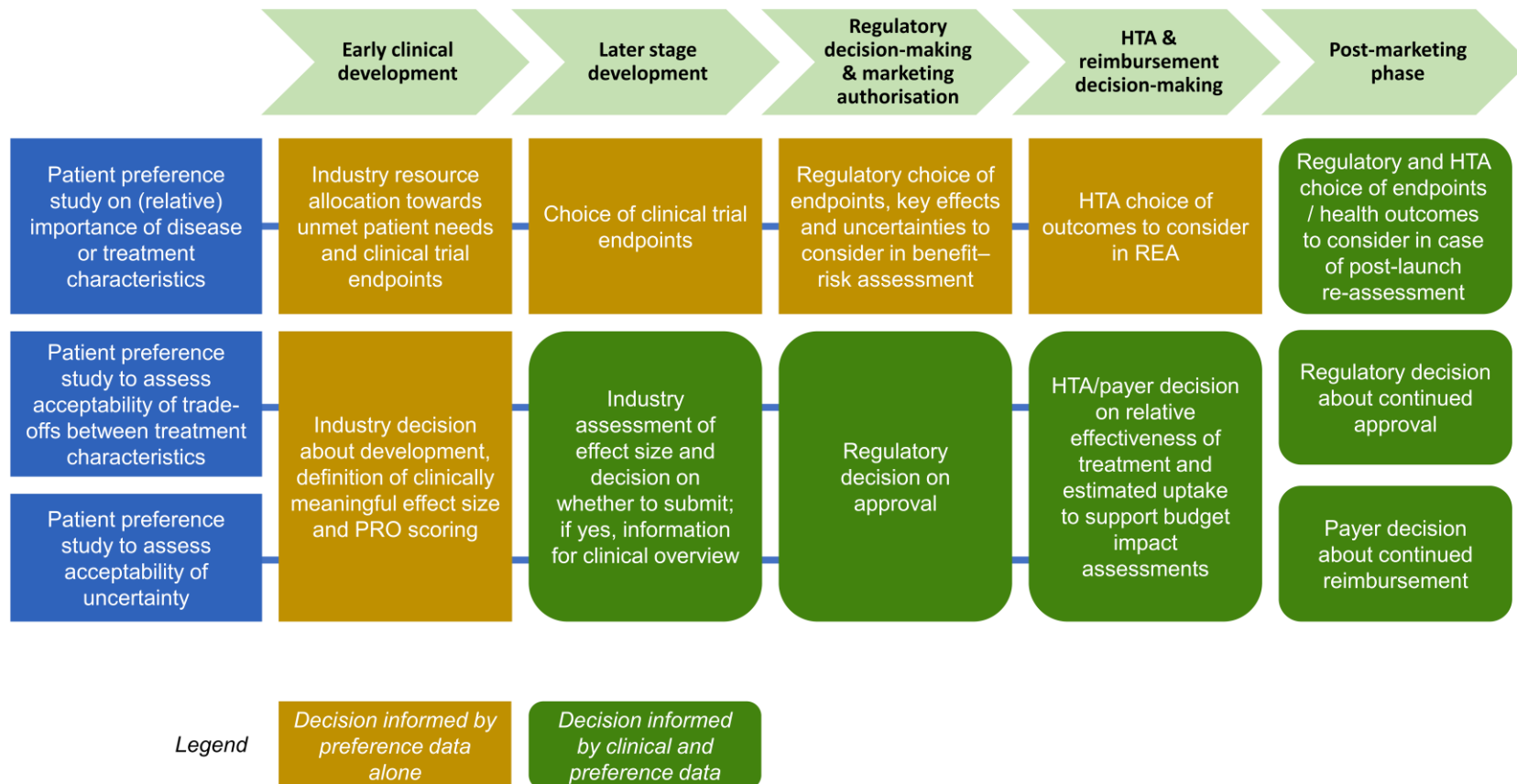
Figure 2-1. Decisions during the medical product life cycle that can be informed by patient preference studies.

2.1.2 Later stage clinical development

Types of decisions in this period that can be informed by earlier preference studies

Later stage clinical development can be informed by preference studies conducted earlier in development. For example:

- Preference studies that previously identified patient-relevant endpoints can inform the design of phase 3 clinical trials – care should be taken that these endpoints are identified in sufficient time to allow consultation with the appropriate regulatory bodies, if they are to be used to inform regulatory decision-making.

- Preference studies that previously identified patient-informed and meaningful effect sizes for treatment characteristics can inform a decision on whether to continue with development of a specific medical product (go/no-go decisions) into phase 3 studies, or to pursue marketing authorisation, when later phase clinical trial data are available.

Types of preference studies that could be run during later stage clinical development

During later stage clinical development, patient preference studies can be conducted to:

- provide information on the acceptability of trade-offs between characteristics related to medical products within a specific treatment landscape – at this stage of development, a medical product's efficacy and safety profile would be better characterised than during early phase development. Thus, a preference study at this stage could include more specific product characteristics than at an earlier stage.

- assess a patient-informed meaningful effect size, particularly on a newer beneficial endpoint.

2.1.3 Regulatory decision-making

At the regulatory decision-making stage, the most relevant preference studies are those described above: those informing the choice of patient-relevant endpoints, those informing decisions about meaningful effect sizes, and those providing information on the acceptability of trade-offs between treatment characteristics. These studies can contribute to decision-making by:

- informing the selection of primary and key secondary endpoints in the pivotal submission study, and thereby informing the selection of key benefits in the regulatory benefit–risk assessment (since a regulatory agency's choice of key benefits will often be based on the primary and key secondary endpoints in the pivotal submission study, e.g. EMA Day 80 assessment report template.(75)

- informing the decision on meaningful effect size in the pivotal clinical trial and thereby providing a benchmark for the assessment of the relevance of the observed clinical data (does the new medical product offer a meaningful difference relative to the current standard of care?) (**Section 3.4**).

- informing benefit–risk assessments that are not self-evident.

Such uses of preference studies, in combination with standard, required clinical trials, are supported by regulatory bodies as described in seminal patient preference information guidance from the FDA,(37) ICH M4E(R2),(76) and recent guidance or strategy documents.(52, 77)

Key considerations about preference studies intended to inform regulatory decisions

If a patient preference study is intended for use at the stage of regulatory review, it will only be useful if the study meets the needs of regulators (e.g. in terms of study population, survey design, and analysis), and therefore it is critical to engage in scientific advice (**Section 4.2**).

2.1.4   HTA and reimbursement decision-making

At the HTA and reimbursement decision-making stage, the most relevant preference studies are those that have identified the unmet needs of patients, those that have informed the choice of patient-relevant endpoints, and those that provide information on the acceptability of trade-offs between treatment characteristics.(33-35) These studies can contribute to decision-making by:

- informing the prioritisation of HTA topics

- informing relative effectiveness assessments (REAs) (78-80)

- informing budget impact analysis (78-80)

- providing insights for health economic evaluations (e.g. cost-utility analyses) if utility is based on data elicited directly from patients.(81, 82)

Patient preference data are complementary to clinical evidence and cost-per-quality-adjusted life year (QALY) estimates based on public preferences.(54) Additionally, HTA bodies use country-specific criteria to assess the value of health technologies, including new medical products. One criterion that is relevant in any jurisdiction is the relative effectiveness of the new technology compared with the current standard of care or other treatments used for a specific indication. For the REA of a new treatment, it is important to know patient perspectives on aspects of their health or the current treatment strategy they would like to see improved. Besides the relative importance of different health benefits of treatment, patient preference studies can also reveal the relative importance of non-health benefits (e.g. mode of administration and convenience of treatment), which can both be included in a HTA.(83, 84)

Key considerations about preference studies intended to inform HTA/reimbursement decisions

If a patient preference study is intended for use at the stage of HTA and reimbursement, it will only be useful if the study can meet the needs of HTA bodies and payers in terms of study population, survey design, and analysis, and therefore it is critical to engage in scientific advice (**Section 4.2**).

Health equity

PREFER acknowledges the potential impact on health equity of using patient preferences in decision-making, because patient preference data will be available for some but not all patient populations. The intention of the PREFER recommendations is to promote the broader use of patient preference studies in the research, development, and evaluation of novel medical products (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 regulatory and value dossiers, leading to a situation where there is greater equity and more patient-relevant information incorporated into decision-making processes.

However, even when preference data are collected for more patient populations, equity issues might arise in the methodological approaches, including patient sampling (covering patients with different socio-economic backgrounds, demographic characteristics, inclusiveness in terms of patients with different levels of health and digital literacy, etc).

2.1.5   Post-marketing phase

At the post-marketing stage, the most relevant patient preference studies are those that can inform acceptability of trade-offs between treatment characteristics. For example, the assessment of a post-approval rare but serious safety signal might gain from a patient preference study to understand how much risk patients are willing to accept for how much benefit.

For a HTA, patient preference studies are useful when the treatment landscape changes and the reimbursement status of medical products needs to be revised in view of new products or other treatments entering the market. For this, new patient preference studies might need to be conducted because the relevant endpoints and relative importance of disease and treatment characteristics might have changed due to the changed treatment landscape.

2.2 Situations where a patient preference study is less likely to inform decision-making

A decision on whether to conduct new preference studies needs to be assessed on existing evidence. For example, patient preferences are less likely to add value to a decision about the choice of patient-relevant endpoints when patient preferences are clear from previous high-quality and up-to-date research. Similarly, patient preferences are less likely to add value to a decision about the acceptability of trade-offs between treatment characteristics when this decision is straightforward (e.g. the new product offers more benefit and/or less risk than existing treatment options). Patient preference studies also have less value when others determine the choice of medical product (e.g. a decision by surgeons to use one particular surgical tool over another). A flowchart for understanding if a preference study is likely to be useful is shown in **Figure 2-2**.
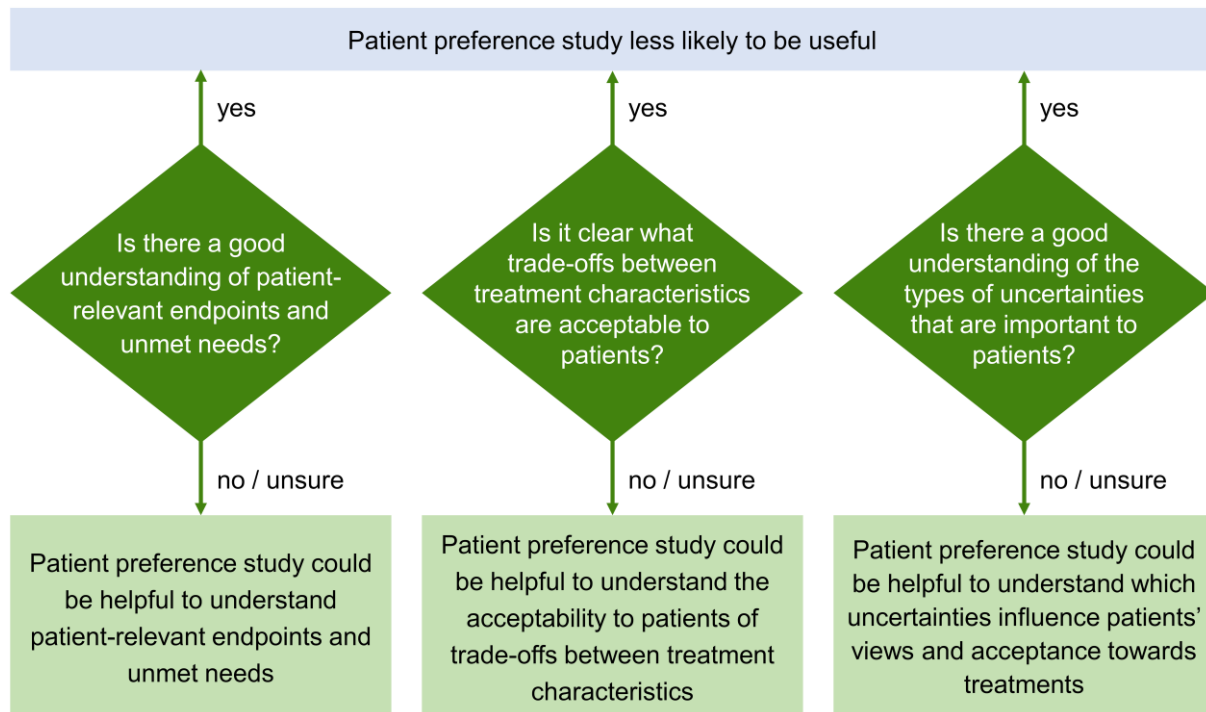
Figure 2-2. Flowchart for understanding whether a patient preference study is likely or less likely to be useful.

3 PREFER framework for patient preference studies

Key messages

The proposed PREFER framework for patient preference studies is intended to be used by study sponsors to facilitate stakeholder review and discussion about objectives, design, conduct, analysis, and interpretation of patient preference studies. The CHMP confirm in the Qualification Opinion that "the framework is suitable for informing on objectives, design and conduct, and reporting of PPS."

The framework complements other published best practices and checklists relating to patient preference studies.

3.1 PREFER framework objectives and overview

The objectives of the PREFER framework for patient preference studies are to:

- inform study research teams 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 medical product decision-making

- support discussions between industry, regulators, and HTA bodies and payers about preference studies intended to inform medical product decision-making.

The CHMP confirm in the Qualification Opinion that "the framework is suitable for informing on objectives, design and conduct, and reporting of PPS". (Note that the framework described in the documents submitted for qualification was slightly different to the framework in these recommendations, which includes some very minor adaptions for clarity; the differences between the two versions are summarised in **Annex Table A3-14**.)

The framework builds upon previous work (3, 34, 37, 85, 86) and also synthesises work of the PREFER project that included systematic literature reviews, comprehensive stakeholder interviews, and case studies. This section is a high-level description of the PREFER framework. **Annex Section A3** includes more detailed descriptions of each framework component (**Figure 3-1**). The PREFER framework is intended to cover all types of patient preference study, irrespective of the method used, and is thereby applicable to both qualitative and quantitative studies.

The PREFER framework has three broad components that can be mapped to the stages of a preference study (**Figure 3-1**). **Sections 3.1** to **3.4** provide a high-level description of the components (with more detail in **Annex Sections A3.1** to **A3.4**). **Section 3.5** describes how the framework can address important issues such as preference heterogeneity and how it supports the scientific integrity of preference studies.
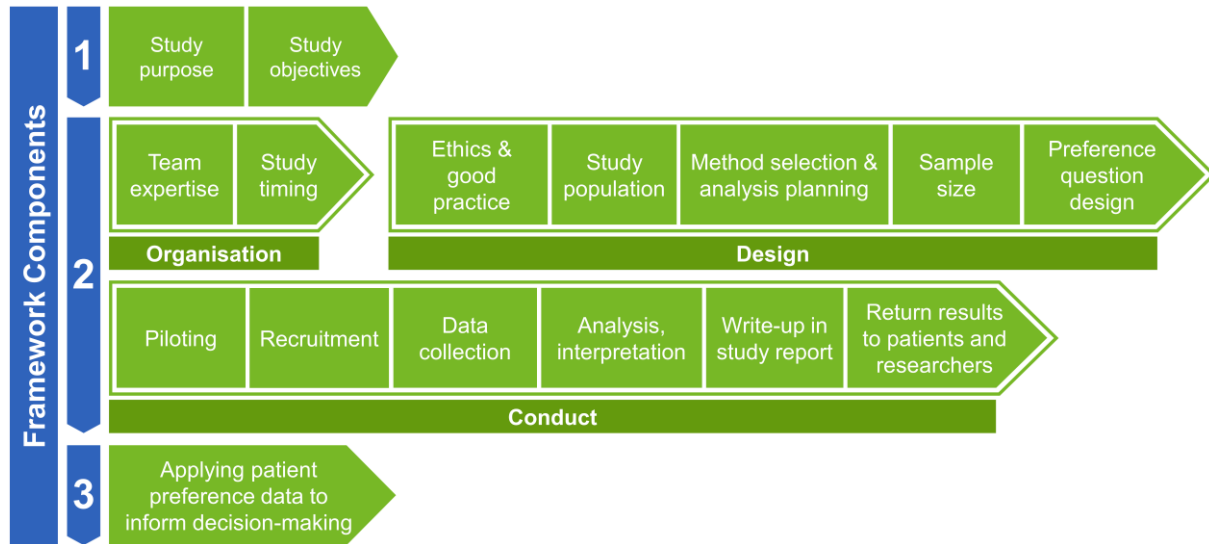
Figure 3-1. PREFER framework structure, aligned with stages and steps of patient preference studies. *Adapted from van Overbeeke et al, 2019.*(3)

3.2   PREFER framework component 1: study purpose and objectives
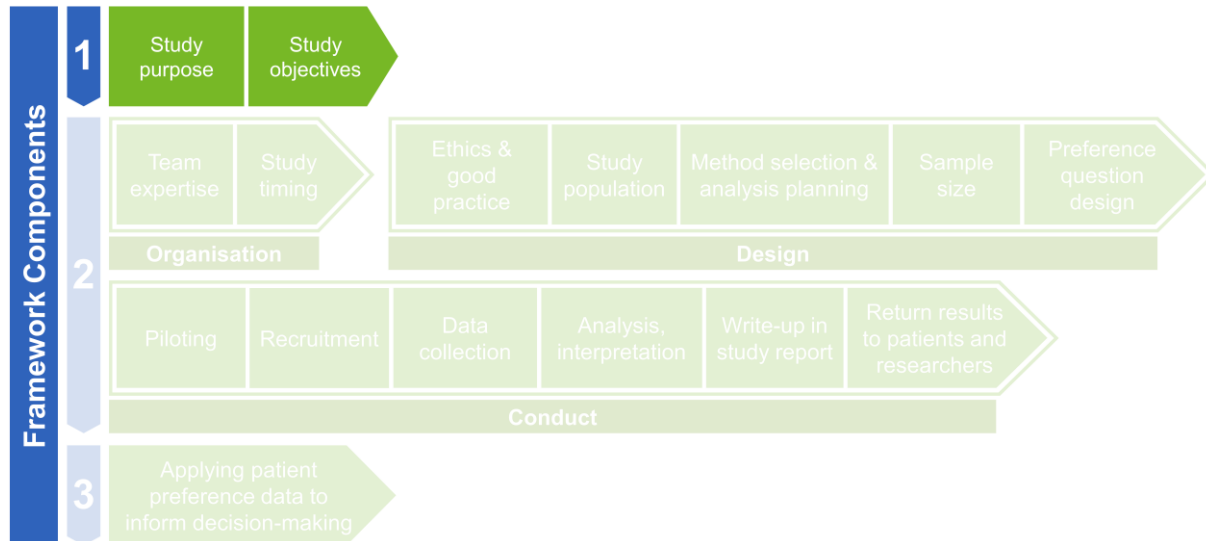
Figure 3-2. Stages of PREFER framework component 1.

3.2.1   Framework component 1: the preference study purpose – what decision will be informed by the preference study

It is critical to identify and contextualise the purpose of the preference study because it should only be performed if it can inform decision-making and the purpose cannot be addressed by existing information. In-line with the advice from Bridges and colleagues,(85) the study purpose should include detailed information about the intended decision and whose preferences are of interest. The purpose should include information about the:

- **decision + decision-makers:** what decisions, and by whom (industry and/or regulatory and/or HTA body), will be informed by the results from the patient preference study (see **Annex Section A3.2.1** for more details).

- **decision context:** for decisions relating to patient-relevant endpoints for a specific indication, the decision context will typically describe what is currently known about the topic. For decisions relating to patients' views on the acceptability of trade-offs between treatment characteristics, other relevant treatment options should be considered, along with the key benefits and risks relevant to the decision, and the known/expected magnitude of the clinical effect of the new medical product relative to the other relevant treatment options (see **Annex Section A3.2.1** for more details).

- preference sensitivity of the decision (see **Section 2.1** and **Annex Section A3.2.1**)

- **whose preferences are of interest:** this would typically include details of the disease and associated population. The study purpose should clearly describe whose preferences are of interest even if this group's preference cannot be collected directly, such as young children (see **Annex Section A3.2.1**).

3.2.2 Framework component 1: preference study objectives – how the preference study will inform this decision

After alignment on the study purpose, the team should identify the objectives and associated preference study endpoints. It is helpful to develop the primary objectives of a preference study with the end use of the preference data in mind, namely, to inform the specific decisions as set out in the study purpose. This connection can be made by considering the link from study objective to preference study endpoint, and from preference study endpoint to application of preference data to inform the medical product decision, as exemplified in **Figure 3-3** (see also **Annex Section A3.2.2**).

Aligned with the agreed preference study purpose:		
Preference study objective	**Preference study endpoint**	**Application of preference study data to inform a decision** [a]
For example: assess the risk tolerance of patients in a specific disease	**For example:** maximum acceptable risk (at specified level of benefit)	**For example:** preference study data used in parallel with expected clinical data to inform industry decision about development of candidate drug in this disease area

Figure 3-3. Example of a link from study objective to study endpoint to application.

[a] *The application of preference data to inform decisions can involve the use of preference data in isolation, preference data in parallel with clinical data, and/or preference data mathematically combined with clinical data (see **Annex Section A3.4.2** for further discussion of this topic).*

The secondary/exploratory objectives of a patient preference study are often intended to provide supportive evidence for the primary objective (e.g. exploring preference heterogeneity/consistency across subgroups).

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
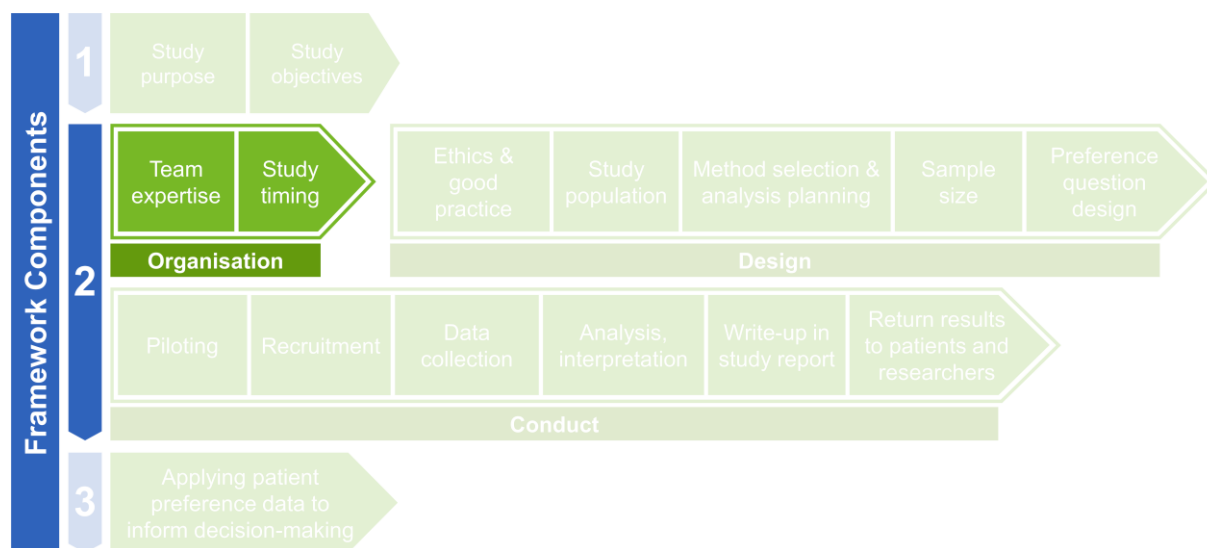
Figure 3-4. Stages of PREFER framework component 2: organisation.

3.3.1.1 Team expertise

Patient involvement in preference studies is critical (37) as described by van Overbeeke *et al*,(3) they should be the central focus of the study – as participants – but also be involved as partners in the study's design and conduct (**Section 4.1.3**). Moreover, the preference study team should include members with expertise in:

- medical aspects of the disease and its treatments
- statistics used in preference study design and analysis
- preference study conduct.

Additional team members may be required, depending on the study purpose and objectives (see **Annex Section A3.3.1.1** for further details).

3.3.1.2   Preference study timing

For a study investigating patient-relevant endpoints that will inform industry decisions about the design of a clinical study, the results should be available before designing that clinical study. For a study investigating the acceptability of trade-offs between treatment characteristics that will inform regulatory decision-making, the results should be available for inclusion in a regulatory submission. For all preference studies, details of the study design should be available so that there is sufficient time to consult with key decision-makers, such as regulators and HTA bodies, and incorporate their feedback into the study design. Examples of preference study timings are shown in **Figure 3-5**; actual timings will depend on many factors, in particular the feasibility of recruiting the sample. For further details on study timings, see **Annex Section A3.3.1.2**.
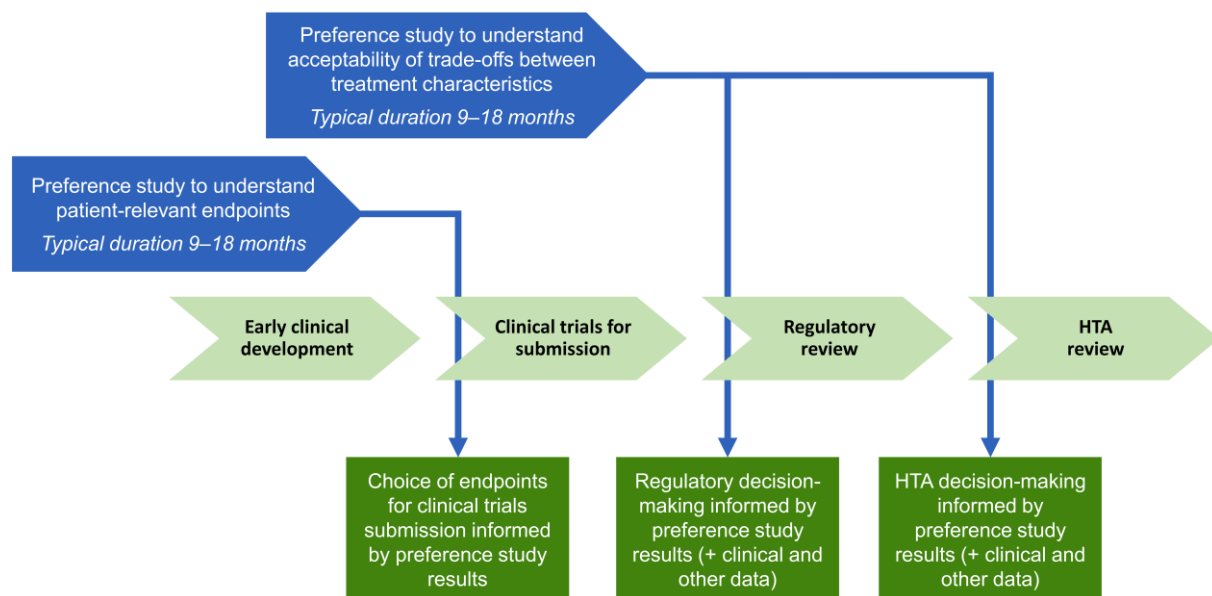
Figure 3-5. Examples of preference study timings relative to the medical product life cycle.

3.3.2 Framework component 2: design

The study purpose and objectives, as well as the planned approach to study design, conduct, and analysis, should be described in a study protocol. Within the PREFER operational guidance a **protocol template** is available; key design elements to include are shown in **Figure 3-6**.
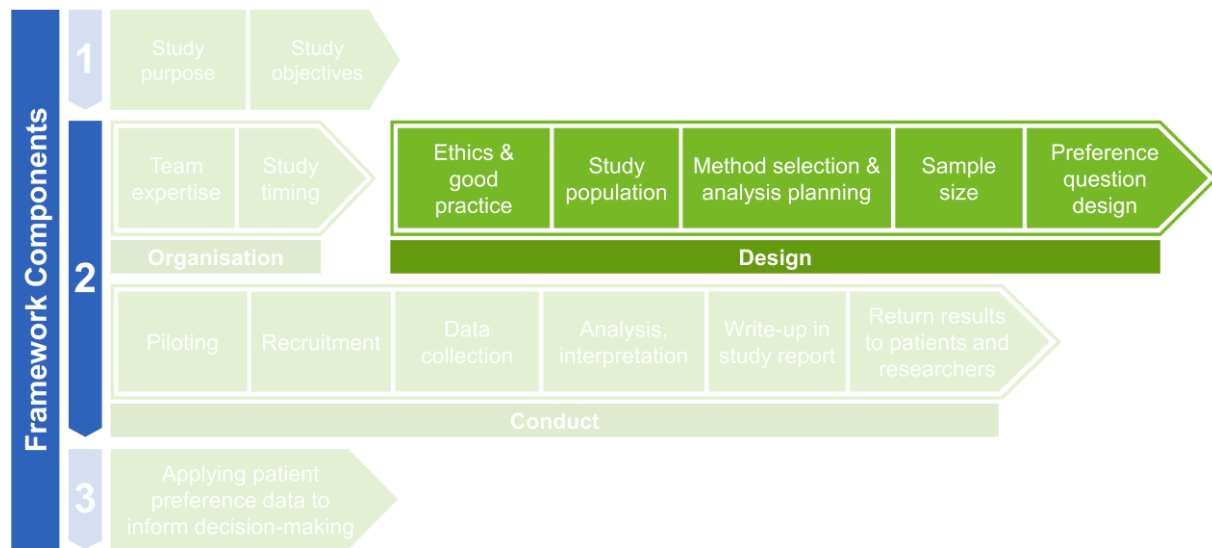
Figure 3-6. Stages of PREFER framework component 2: design.

The activities described in the 'design' section of component 2 are interrelated and therefore the sequence of activities described in the figure could be adjusted as needed; for example, some aspects of analysis planning might only be feasible after work is completed on the preference question design.

3.3.2.1 Ethics and good practice

Patient preference studies should adhere to ethical principles in the same way that clinical trials do, including ensuring that participants receive all the information they need, in a way that is easy to understand, to be able to provide informed consent. Recommendations for this are described in **Annex Section A3.3.2.1**.

3.3.2.2 Study population

When defining the study population, the inclusion and exclusion criteria issues to be considered include:

- alignment of the preference study population with the preference study purpose and objectives

- the representativeness of the preference study sample

- whether a self-reported diagnosis of disease is acceptable, or whether a physician-confirmed diagnosis is required

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

See **Annex Section A3.3.2.2** for further advice about the study population, including a discussion of when it may be appropriate to collect preferences from caregivers.

3.3.2.3 Method selection and analysis planning

Within the PREFER operational guidance, a template for a **statistical analysis plan** (SAP) is available.

Points to consider for method selection

A key point to consider when selecting a method is its alignment with the study purpose and objective, as well as considerations relating to participant and feasibility factors. See **Table 5.3** for examples on how different case studies approached method selection. See **Section 5.2.2** and **Section 5.3.2** for points to consider for method selection, and **Annex Section A5** for detailed considerations by five different quantitative methods.

Points to consider when planning analyses

At a high-level, the PREFER framework is aligned with key principles of the ICH harmonised guideline E9 (R1) Addendum on Statistical Principles for Clinical Trials (87) and the addendum on estimands and sensitivity analysis in clinical trials. Specifically, concepts include the use of detailed study objectives, the choice of method that aligns with these study objectives, selection of study metrics that align with estimand requirements appropriate to the sample, and considering how the preference data will inform the decision (see **Annex Section A3.3.2.3**). Regardless of whether a qualitative or quantitative method is selected, the following recommendations apply:

- Specify and describe in advance what analytical approach will be taken (e.g. descriptive statistics, thematic analysis, modelling if relevant), including any planned statistical testing and explain its relevance to the study objectives.

- Describe how the preference data will be used to support the agreed decision and pre-specify how data from the study will be applied (see **Section 3.5** for application approach).

- Summarise the basic approach for the planned pre-specified analyses or assessments in the protocol (as a minimum); an SAP for more detailed planning can be developed.

- Describe whether, when, and how patient partners will contribute to the analysis of data and interpretation of results of the study.

- Establish and follow a data management plan as part of analyses planning (note that a PREFER **data management template** is available).

3.3.2.4   Sample size

A preference study protocol should include a justification for the proposed sample size based on the primary objective, and, if applicable, also on secondary objectives.(37, 88) Further discussion of sample sizing for five quantitative preference methods is described in **Annex Section A3.3.2.4**.

3.3.2.5   Preference question design

The core components and general considerations to inform patient preference question design – broadly defined as developing an interview or a discussion guide (for qualitative studies) or a survey instrument (quantitative study) – are summarised hereafter.

Some aspects depend on the method and whether the results from the preference study will be combined with clinical trial data (see **Annex Section A3.3.2.5** for additional considerations for these circumstances).

Background: context description

A critical part of preference question design is the study introduction because it orients the participant to the entire study. Among other aspects, the introduction informs the participant about the role they should assume for the study and fully describes the scenario in which the preference questions will be asked.

Background: baseline characteristics

Consideration should be given to what baseline participant or disease characteristics would inform the stated preferences and, by extension, interpretation of the results.

Discussion guide and survey development: description of alternatives

Preference questions are used to understand the relative desirability or acceptability of treatments or of the attributes of those treatments; see **Annex Section A3.3.2.5** for examples. (34) These attributes must be identified and described in a patient-friendly way, avoiding framing bias as much as possible, when creating the discussion guide or survey. The selected attributes of treatments depend upon what is considered important to the patient, the study objectives, and the method selected. Two main approaches are used to identify treatment options or treatment attributes for inclusion in preference questions: 'top-down' (starting with existing knowledge and medical product development expertise) and 'bottom-up' (starting with direct conversation with the patient/caregiver to understand what matters most to them in the management of the disease). See the PREFER **case study reports** and **Annex Section A3.3.2.5** for further details.

Discussion guide and survey development: development of levels

Some quantitative methods (e.g. DCE, swing weighting, threshold technique) also require the selection of levels for each attribute. Selection of the type, number, and way of presenting associated levels for each attribute depends upon the study objectives and method selected, the clinical relevance, and the clarity to patients. Level descriptions should be free of framing bias as much as possible. See **Section 5** for method-specific considerations and **Annex Section A3.3.2.5** for general considerations across methodologies.

Discussion guide and survey development: inclusion of assessments

- **Patient education and comprehension:** the acceptance of preference data as valid scientific evidence requires that patients participating in preference studies understand the design and context of the study, the disease, and the choice tasks in which their preferences are explored or elicited. Patient education materials can facilitate this (**Section 6**).

- **Psychological constructs:** these can measure participants' psychological characteristics such as personality traits, social-cognitive factors, experiences with their disease, treatment, and decision-making styles (**Section 7**). Investigating these characteristics may offer important insights into why preference heterogeneity exists within a study population and/or the factors that influence the formation of patient preferences.

- **Internal validity:** addressing issues of internal validity (the degree to which results are trustworthy and meaningful) should be pre-specified in the protocol and SAP. The chosen internal validity assessments will depend on the study objectives, the preference method, the sample size, and the length and cognitive burden of the study. Best practices should be followed as documented in the literature, (89, 90) and examples are available in the **PREFER case study reports**.

Discussion guide and survey development: question and exercise development

There are multiple considerations when developing the questions or exercises that will be presented to a participant for collecting stated preference data. All questions should be free of framing bias as much as possible and be appropriately open-ended and unbiased. For more information on question development see **Annex Section A3.3.2.5** and the PREFER case studies – recommendations are based on best practices.(85, 86, 91)

Finalising preference question design: considerations of cognitive burden and capacity

Consideration should be given to the characteristics of the patient population (e.g. age, presence of cognitive impairment, educational level, health literacy, numeracy) using best practices in communicating the attributes, the length of the interaction/survey, and how the study will be conducted (e.g. computer-based only, face-to-face). For example, a computer-based only format can impede the ability of the intended participants to complete the study, as shown in the NMD case study, **protocol** and report.

Finalising preference question design: assessment of study materials

It is recommended that study materials are assessed by patients before study initiation (i.e. pre-tested), usually by one-on-one interviews or talk-aloud exercises with a convenient sample of patients. This enables assessment of whether the content is understood in the intended manner, whether questions and exercises are clearly understood, realistic, adequate in terms of length, and, if applicable, whether levels are sufficient to induce trade-offs between treatment characteristics.

Finalising preference question design: translation of study materials

For study materials that require translation, it is recommended that the ISPOR Principles of Good Practice for Translation of Patient-Facing Material (92) are followed. Initial translations can be made by suitably qualified (e.g. International Organization for Standardisation [ISO] certification) translation companies; however, patients should then review certain study components (e.g. interview questions) and other study materials to affirm they are understandable in their native language.

3.3.3   Framework component 2, conduct

During the 'conduct' stage of a preference study (**Figure 3-7**), teams should continue to apply the same principles of ethics and good practice as discussed in the previous section.
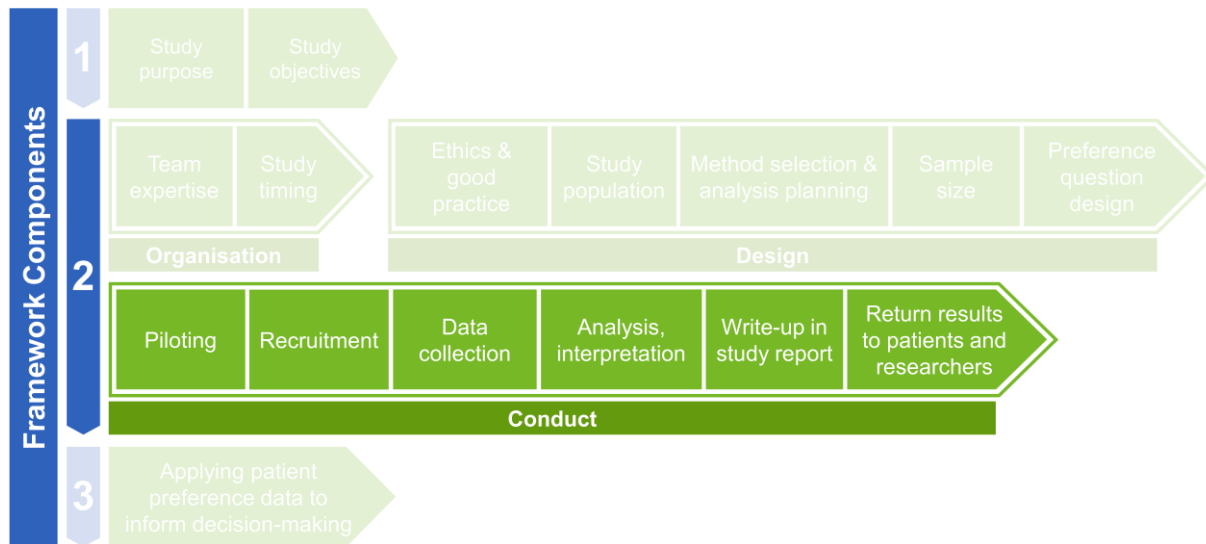
Figure 3-7. Stages of PREFER framework component 2: conduct.

3.3.3.1   Piloting, participant recruitment and data collection

In the PREFER framework, piloting is typically only completed for quantitative studies. Piloting comprises a soft launch of a survey with a small subset of the full participant sample to check that the survey and data collection work as expected, and for excessive cognitive burden (**Annex Section A3.3.3.1**). Participant recruitment and data collection are operational aspects of a preference study (**Annex Sections 3.3.3.2** and **A3.3.3.3**). A clear data collection plan is critical to the quality and success of the study; as such, PREFER has developed a general **template** for use.

3.3.3.2   Analysis, interpretation, and write-up in a study report

The analyses conducted in a preference study should be those specified within the SAP and/or protocol and summarised in a study report. Additional analyses or deviations should also be described and provided in accordance with the ICH harmonised guideline E9 (R1) (see also **Annex Sections A3.3.3.4** and **A3.3.3.5**).(88) PREFER has developed a **study report template** to assist with creating a study report.

3.3.3.3 Returning results to patients and researchers

Returning results to patients

The final step is the return of results to study participants, in accordance with best practices on plain language.(93-96) These plain language principles are summarised in the PREFER **plain language summary template**. See **Annex Section A3.3.3.6** for more details.

Making preference study results available to researchers

Preference study teams should make every effort to publish the study results. Guidelines on reporting practices have been published that are relevant to qualitative (91) and quantitative (85) studies.

3.4 PREFER framework component 3: applying preference data to inform medical product decision-making
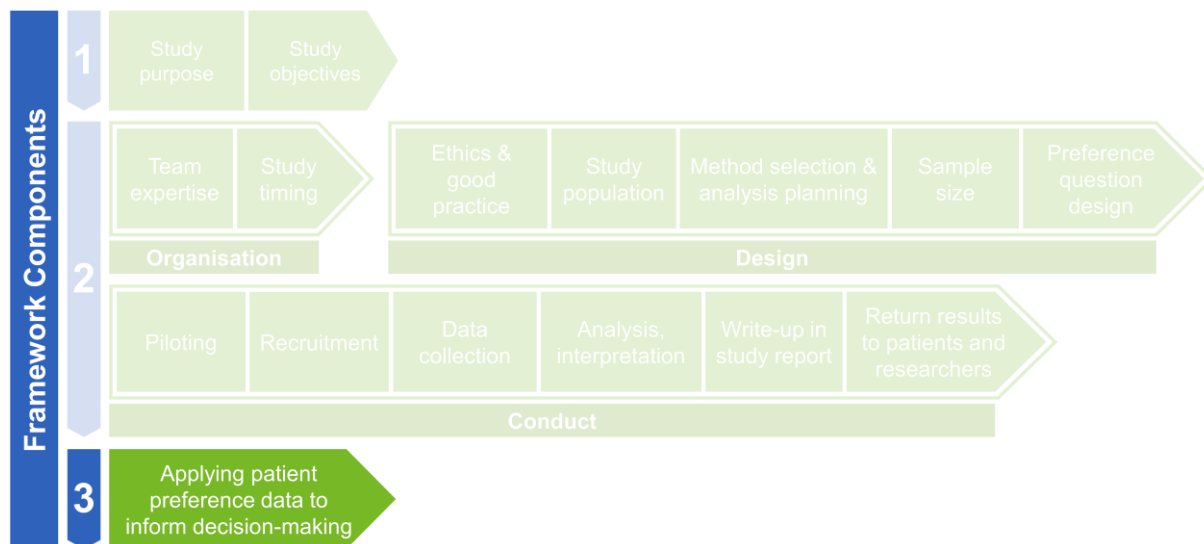
Figure 3-8. Stages of PREFER framework component 3.

This section focuses on two key applications of patient preference study results:

- choice of key endpoints
- acceptability of trade-offs between treatment characteristics.

There are many potential applications of patient preference studies in medical decision-making that are not discussed in this section; see **Annex Section A3.4.1** for more information. Technical methods for the application of preference data are discussed in **Annex Section A3.4.2**, and proposals about incorporating preference information into industry, regulatory, and HTA/payer documents are discussed in **Annex Section A3.4.3**.

3.4.1 Applications of preference data to inform medical product decision-making for choosing patient-relevant endpoints

The choice of patient-relevant endpoints can potentially be informed by a single patient preference study, as shown in **Figure 2-1**. More information on the decision types outlined below is available in **Annex Section A3.4.1.1**.

3.4.1.1 Industry decisions

In a situation where there are no established endpoints for an indication, a typical approach would be to base the decision on the preference weights (i.e. preference data in isolation – **Annex Section A3.4.2.1**). This approach can also be used to understand the importance to patients of a non-health benefit such as convenience or mode of administration.

In a situation where the preference study is used to re-assess presumed 'established' endpoints for an indication, one approach is the use of choice share information derived from preference data in combination with hypothetical clinical data. See **Annex Section A3.4.2.3** for more details on this topic.

3.4.1.2 Regulatory decisions

The regulatory decision about the choice of patient-relevant endpoints would generally be driven by the endpoints used in the clinical study. The choice of key favourable effects can be achieved by including the primary efficacy endpoints and the most clinically-relevant secondary endpoints.(75) If the choice of clinical trial endpoints has been informed by the results of a preference study, the regulator could also consider these results to inform their decision about the choice of patient-relevant endpoints.(97)

3.4.1.3 HTA/payer decisions

HTA/payer decisions about which endpoints to incorporate into an REA would generally be driven by the endpoints used in the clinical trial, along with patient preference data. For instance, patient preference data can be directly used in HTAs to define/confirm the clinical outcomes to be included in the REA, or to quantify patient preferences on non-clinical outcomes, such as improved convenience of a new treatment.(54)

3.4.2 Applications of preference data to inform medical product decision-making about the acceptability of trade-offs between treatment characteristics

As shown in **Figure 2-1**, all these decisions can potentially be informed by a single patient preference study.

The decision context for a regulatory decision could be different to the context for an HTA/payer decision; for example, the regulatory decision might rely on a different assessment of the treatment landscape to the HTA body/payer.

3.4.2.1 Industry and regulatory decisions

- Patient preference data about the acceptability of trade-offs between treatment characteristics can inform **industry** development or submission decisions, and also **regulatory** approval decisions and post-approval decisions (e.g. if a safety signal prompts a re-think of a medical product's benefit–risk profile). Industry development decisions could be informed by side-by-side approaches to preference data and hypothetical clinical data to support discussions on maximum acceptable risk (MAR) for a specific hypothetical level of benefit/minimum required benefit (MRB) for a specific hypothetical level of risk.

- Both industry submission decisions and regulatory decisions could be informed by:

- data displays combining both preference and clinical data (**Annex Section A3.4.2.2**), *and/or*

- a side-by-side approach to preference and clinical data to support discussions on MAR for a specific level of benefit/MRB for a specific level of risk (**Annex Section A3.4.2.2**) – this approach is only applicable 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 or multi-criteria decision analysis (**Annex Section A3.4.2.3**).

3.4.2.2   HTA/payer decisions

Patient preference data about the acceptability of trade-offs between treatment characteristics can inform HTA/payer decisions about the hypothetical uptake of a new treatment to inform budget impact calculations and organisational decisions – these could be informed by the choice share (**Annex Section A3.4.2.3**). The decision context relevant to the HTA decision could be different to the decision context relevant to the regulatory decision (**Annex Section A3.2.1**).

Such patient preference data can also be used to inform reimbursement revisions when the treatment landscape has changed from when the initial reimbursement request was submitted. These require information on the REA of all treatments for a specific indication and could be informed by patient preference studies covering relevant characteristics of all treatments. Data displays covering both clinical and preference data as well as information on choice share will help in the assessment, potentially leading to reimbursement revision (**Annex Section A3.4.2.2**).

3.5   Key considerations relating to the PREFER framework

Key factors when considering the inclusion of patient preference data in medical product decision-making are preference heterogeneity and ensuring scientific integrity. Further key factors relating to the PREFER framework are discussed in **Annex Section A3.5** and cover the link between the PREFER framework and preference methodology (**Annex Section A3.5.1**), and how the framework addresses operational issues (**Annex Section A3.5.2**).

3.5.1   How the PREFER framework addresses preference heterogeneity

Preference heterogeneity refers to the degree to which preferences at an individual level – which are, by nature, subjective – differ from preferences expressed at a collective level.(98) For example, some patients might be more willing to accept a higher level of risk for a specific level of benefit than others.(37) The PREFER framework covers issues of population preference heterogeneity at all stages – see **Table 3-1**. See **Annex Section A3.5.3** for further discussion on this topic.

Table 3-1. How the PREFER framework covers issues of population preference heterogeneity.

Framework section	Advice relating to heterogeneity issues
Section A3.2.2, preference study objectives	Consider the need for study objectives that investigate preference heterogeneity across the patient sample.
Section A3.3.2.3, method selection and analysis planning	Consider to what extent a preference method enables the investigation of preference heterogeneity, and the *a priori* planned approach to any analyses assessing patient heterogeneity (i.e. analyses linked to the study objectives relating to heterogeneity).
Section A3.3.2.4, sample size	If a study objective relates to a specific sub-group of patients, consider the need for the study to include sufficient patients in all subgroups of interest.
Section A3.3.2.5, section on collection of baseline data	Consider the need to collect data on baseline characteristics, disease characteristics or any other characteristics of the anticipated patient population that may influence their response choices.
Section A3.3.3.4, analysis, interpretation	Consider if/how patient heterogeneity influences the interpretation of results.

3.5.2 How the PREFER framework supports scientific integrity and credibility of patient preference studies

One concern raised by stakeholders is whether the results of patient preference studies are unbiased.(1) As with clinical trials,(99) 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. Areas of the PREFER framework that specifically address these concerns are described in **Table A3-13** in **Annex Section A3.5.4**.

4 Recommendations for working with key stakeholders – patients, regulators, and HTA bodies

Key messages

When working with patients

It is important to empower patients as research partners in preference studies to increase the relevance, appropriateness, feasibility, and acceptability of the preference study design, as well as its conduct and the interpretation of findings. Patients should be involved as study team members or advisors throughout the entire process, from study planning through to the communication of results.

Box 4-1 provides recommendations to facilitate involvement of patient research partners in patient preference studies.

When working with regulators and HTA bodies

Early consultation with the relevant regulator and/or HTA body is needed to ensure a patient preference study adds value and delivers results that can be integrated in the decision-making process.

Awareness of the needs and expectations of regulators and HTA bodies – including any differences between these groups – is necessary so that the sponsor of the patient preference study can design a single study that is fit for purpose.

Box 4-3 provides a checklist to help study sponsors guide scientific advice discussions.

This section provides additional guidance for patient preference study teams to consider when implementing framework component 2 to involve patients (as both study participants and as research partners), regulators, and HTA bodies.

4.1 Working with patients as research partners in patient preference studies

4.1.1 Why is it important to involve patients as research partners?

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

- Patients have the moral right to be involved in research that concerns them directly or indirectly.(64, 100, 101)

- Other stakeholders (such as pharmaceutical companies, regulators, payers, clinicians, and academia) do not necessarily know which treatment features matter most to patients.

- Patients can offer a unique perspective that differs from that of healthcare professionals or competent authorities because they have experiential understanding of the disease.

- Involving patients as research partners increases the relevance, appropriateness, feasibility, and acceptability of a patient preference study.

- Involving patients in study design leads to research that is more ethical and outcomes that are of greater relevance and/or value to patients.

- Patient involvement is important because it provides complementary input to scientific evidence on patient perspectives and experiences.

4.1.2 When is it important to interact with patients as research partners?

Patients should be involved as research partners in all steps of the framework (**Section 3**). They can contribute to decision-making about the study before, during, and after the research process. For example, in the PREFER multiple myeloma study, patients were involved as team members in the development of the protocol (see **Annex Table A4-1** for other examples).

When defining the study purpose and objectives (framework component 1)

Close collaboration between patient research partners and sponsors when defining the research question of a patient preference study can help ensure it is phrased in a way that is clearly understandable to patients, which includes identifying ambiguous or difficult language. Patients can also help ensure the question is relevant to the targeted patient community (**Annex Tables A4-2** and **A4-3**).

When designing and conducting the study (framework component 2)

The involvement of patient research partners is valuable in helping to:

- define the sample inclusion and exclusion criteria and advise on patient recruitment procedures

- develop the data collection instruments (e.g. formulating the questions, selecting and defining attributes and levels) and educational materials (**Section 7**) to ensure they can be easily understood by patients and that any ambiguity and misunderstanding is avoided

- ensure that questions and answers are plausible, relevant, and meaningful, based on the patients' own experience, thereby helping to improve the validity of the study results.

When using a survey, it is strongly recommended that patient research partners are involved in deciding how to present the survey, specifically in assessing its format, layout and length, the mode of presentation (digital or paper-based), consistency in wording, and the inclusion of general elements that may contribute to increased trust between patients and researchers. See **Annex Tables A4-4**, **A4-5** and **A4-6** for examples of how patients were involved in the design of PREFER case studies.

In the conduct stage of a patient preference study (framework component 2), patient research partners can add value by, for example, interviewing participants and training sponsors to understand the language that patients use.

When interpreting and communicating the study results (framework component 3)

During the interpretation of results, patient research partner involvement is highly recommended to help clarify nuances, indicate whether the results make sense, and help explain the findings. These interpretations of both qualitative and quantitative data help to contextualise the results from the point of view of the patient and should be combined with the viewpoints of other stakeholders.

Results of the study should be communicated to all stakeholders including patients themselves – both study participants and non-participants who are part of the target patient population. Dissemination of results to patients should only begin once they have been reviewed and sponsors are confident of their accuracy.

4.1.3   How to identify and recruit patient research partners?

The PREFER definition of the term 'patient' also includes proxies for patients (i.e. individuals close to the patient), such as parents and carers, if patients are unable to express themselves due to their age or limited cognitive abilities. The term 'patient' can also include representatives

from patient organisations.(102) In some cases, the patient may be a healthy member of the public or an at-risk individual (e.g. in the context of a vaccination programme or other type of preventive medicine programme). In these situations, the recipients of the proposed therapy may not view themselves as patients and so sensitivity regarding terminology is important.

There are multiple ways that patient research partners can be identified; for example, via a patient association or referral from a healthcare professional.(103) Patient preference study teams should identify and recruit patient partners with the necessary experience and/or expertise (preferably based on collective knowledge rather than individual experience) and there are several recommendations that should be adopted to ensure inclusivity.(104) Study teams should be flexible in how patients become involved so that different patient research partners can undertake different tasks.

The intensity and level of the patient partnership can vary from a targeted or embedded consultation through 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, whereas in an embedded consultation, patients are regularly consulted throughout the entire research process. If patients are involved as co-researchers, they should share in the decision-making as integral members of the research team. For example, in the PREFER rheumatoid arthritis case study, patient research partners were consulted about the desired level of involvement in the study design. This involvement was extensive and most active during the development of the focus group schedule, analysis of focus group findings, attribute selection and survey instrument development, and survey pre-testing. However, the patient partners did not wish to contribute to regular case study team meetings or co-author the manuscript.

There is no single optimal number of patients to include as research partners in any given patient preference study. No individual can represent the perspectives of all patients and so multiple patient representatives are valuable. However, the ability of the study team to support effective patient involvement of larger groups can be limited. In the rheumatoid arthritis case study, for example, eight patient research partners representing four countries were involved, which was facilitated by remote virtual meetings and regular email interactions.

What profile constitutes a good candidate for a patient research partner?

When selecting patient research partners, the selection process should take into consideration the diversity of perspectives and experiences of the target population. Key considerations can include (but are not limited to): diversity in experience with the disease (severity) and its treatment, demographics, sociocultural background, health literacy, and experience with clinical trials for a similar disease.

The expertise and experiences needed from a patient research partner should be informed by the study objectives and activities, which may vary based on the type of input needed at different stages. Such expertise and experience may include reviewing documents, liaising with other patients/patient organisations, and interviewing other patients or their representatives. In some cases, patient research partners with a combination of experiences are needed.

Example

In the PREFER rheumatoid arthritis case study, some of the supporting patient research partners had previously worked with the study team on international projects (105, 106) addressing the development of predictive and preventive approaches for rheumatoid arthritis, and this collaboration helped shape the clinical objectives of the PREFER case study.(107) Input from patient partners across different European countries was useful in this case study as survey recruitment took place in three countries. Some of these patient partners had contributed to a previous systematic review of qualitative evidence (107) and had received training in qualitative analysis and meta-synthesis. This enabled them to actively contribute to the coding and interpretation of the focus group and the interview data that was collected.

4.1.4   How best to work and interact with patients as research partners?

The involvement of patient research partners should follow the recommendations in **Box 4-1** and be guided by communication plans that are developed in collaboration with patient research partners (see **Box 4-2**). Also, it is important to provide regular feedback at each stage of the patient preference study – for example, about the project progress and the impact of patients' contributions – to patient research partners.

While the PREFER recommendations focus on patient involvement in patient preference studies, other recommendations and working practices for patient engagement activities in medicine development are also available.(108-110)

Box 4-1. Recommendations to facilitate involvement of patient research partners in patient preference studies.

1. Use easy to understand, non-technical language, and include glossaries of technical terms where required.

2. Clearly and concisely describe the roles of patient research partners.

3. Undertake outreach work to involve patient research partners in community settings.

4. Enable flexibility around meeting times, including out-of-office hours.

5. Use easily accessible meeting venues (e.g. lifts/ramps, locations)

6. Provide opportunities for patient research partners to contribute remotely (e.g. via email, teleconferences, video meetings).

7. Ensure meetings are structured to accommodate the needs of patient research partners (e.g. frequent breaks, refreshments, lay summaries of presentations/documents, care givers can attend)

8. Reimburse any expenses and payments for time spent.

9. Provide recaps at regular intervals of the study background and objectives, progress updates, and the impact of the patient research partner activities.

10. Allow sufficient time for the completion of involvement activities.

11. Ensure there is no requirement for patient research partners to sign or review lengthy and/or complex documents or legal agreements.

12. Ensure patient research partners have the requisite skills and knowledge to support meaningful involvement (e.g. to enable patients to contribute to aspects of data analysis or study conduct, assertiveness skills to support participation in management meetings). This may require specific training or provision of information or support.

13. Provide training for study sponsors so that they can effectively involve members of the public (e.g. communication skills, needs awareness, outreach training).

Box 4-2. Recommendations for communication plans.

1. Integrate suggestions from patient research partners into draft summary result templates to ensure that the language used is understandable by patients.
2. Communicate when the study will end and when the study results will be shared.
3. Before the study results are shared, provide an overview of the process that led to the final outcomes.
4. When sharing the study results, state where a plain language summary will be made available to the public.
5. Employ user testing among a group of patients to evaluate the final version of the plain language summary.
6. In the plain language summary, acknowledge the involvement of the patient research partners and study participants, and describe how the patient research partners had an impact on the study.

4.2 Interactions with regulators and HTA bodies on patient preference studies

4.2.1 Why is it important to involve regulators and HTA bodies in the design of patient preference studies (framework component 2)?

This section provides more information to support the **Section 3** recommendation to consult relevant regulatory authorities and HTA bodies about preference study design. The EMA has confirmed that the PREFER framework can support interactions between industry, regulators, and HTA bodies and payers (as well as patients), and help guide their assessments of patient preference studies and use the study results to inform their decision-making.(32)

Despite the recognised value of a more structured approach to using patient-based evidence in decision-making along the medical product lifecycle, successful integration of patient preference information has, so far, been slow, unsystematic, or very limited.(72) Interviews with stakeholder representatives conducted within PREFER revealed that, so far, few regulators and HTA bodies around the world have recognised or structured approaches to the use of patient preference data in their decision-making processes for medical products, which has contributed to their limited acceptance.(1) This is not unexpected because patient preferences remain in the early stages of adoption and there is limited experience within many regulatory and HTA systems. As a result, there is agreement on ICH level 2 among stakeholders in the regulatory environment that guidance work needs to evolve further. This view is shared by the EMA, who state in their final Qualification Opinion on the PREFER framework:

"The systematic efforts by the IMI PREFER project to address gaps in approaches to incorporate patients' views into decision-making and to develop a framework for patient preference studies are acknowledged."

and

"There is a shared interest in structuring patient involvement, including patient preference studies, in regulatory decision-making processes" (32)

The value and weight of patient preference information in decision-making depends on several factors, such as the stage of the product lifecycle, decision-making context, type of medical product, disease area, and unmet patient need. It is therefore important for sponsors to identify the relevant factors with decision-makers at the outset of planning a patient preference study. Specifically, PREFER recommends engagement with regulators and HTA bodies to facilitate a common understanding of:

- the objective that the patient preference information is intending to support

- how to ensure that patient preference information can be measured as reliably as possible

- how to ensure that patient preference studies can be conducted with appropriate scientific rigour

- how to ensure that patient preference studies provide relevant information to inform decision-making at various stages of the medical product life cycle.

As outlined in the framework (**Section 3.4.2**), in the context of **regulatory decision-making**, it is recommended that the needs and experiences of patients are considered when selecting outcomes for studies and trade-offs between treatment characteristics to inform marketing authorisation, follow-up indications, and line extension decisions.

Similarly, in the context of **HTA decision-making**, patient preference studies are advised in several particular use cases,(111) including:

- understanding what outcomes matter to patients

- predicting patient choices when treatment uptake is particularly important

- estimating the utility generated by the treatment where existing utility measures fail to fully capture their value from the patients' point of view

- estimating the WTP for treatment benefits where patients are paying for treatment themselves.

- Regulators and HTA bodies have different criteria for making decisions. Regulators make decisions for an entire patient population on the basis of a positive benefit–risk assessment

of the medical product, and based on quality, safety, and efficacy criteria. Health technology assessment perspectives often go beyond that of patients only and include those of health insurers, insured individuals, taxpayers, and society as a whole. Additionally, depending on country-specific HTA processes, there are procedural and practical barriers for integrating patient preferences within economic evaluations.(98) In many jurisdictions, the preferences of the general public are preferred to, or used in conjunction with, those of individual patients in making resource allocation decisions.(98) It is therefore important to understand and discuss which requirements prevail, and whose preferences are being measured, with the relevant bodies at the design phase of a preference study intended to be used for HTA decision-making.

4.2.2 When and how to consult with regulators and HTA bodies

To enable regulators and HTA bodies to incorporate patient preference information in their decision-making, it is important that study sponsors have discussions with these stakeholders before the start of the study to ensure that all relevant information will be captured. This is especially pertinent when the patient preference study is intended to be used in dossiers or labels to inform regulatory/HTA decisions and/or assist in the interpretation of clinical study data generated for a specific product.

The cornerstone of successful engagement with regulators (such as the FDA and EMA) and HTA bodies is a scientific advice procedure. The purpose of scientific advice is to ensure that a patient preference study is designed such that its results are useful to the decision-maker. When patient preference research is expected to impact more than one decision-making process in the EU, multistakeholder alignment can be assisted by convergence mechanisms such as EMA scientific advice, EUnetHTA joint scientific consultation, or EUnetHTA-EMA joint scientific consultation. To maximise the benefits of scientific advice for patient preference studies, it is recommended that:

- All available guidance by relevant regulatory and HTA bodies is considered. To complement this, and in the absence of detailed patient preference guidance, the PREFER recommendations and EMA Qualification Opinion on PREFER can be considered.

- Scientific advice procedures are initiated as early as possible when the study is intended to be used in decision-making by regulators or HTA bodies.(32)

- Regulatory and HTA bodies may want to include scientific experts in patient preference elicitation into scientific advice processes because preference studies use methodologies that are different from those used in clinical trials and observational studies, and which are

more comparable to those used in utility elicitation studies. Protocol development advice may require experts who can assess the design and results of a preference study.

- Sponsors, regulators, and HTA bodies involve patients as research partners in the scientific advice process more frequently (**Section 4.1.1**) because their perspectives can complement the sponsor's scientific rationale for undertaking the study, and their views on the proposed design of the study are taken into consideration.

- High-quality guidance, in the form of a 'briefing book', is needed, which considers available guidelines at the time of filing; the PREFER framework could be used as surrogate where such guidelines do not exist or do not mention particular topics relevant for patient preference studies. The briefing book could optimally include the:

- rationale for conducting the study (e.g. to inform a specific benefit–risk assessment)

- research question and study context

- study methodology in sufficient depth, including the instrument to be used, the selection process of the instrument, and the instrument testing and revision process

- ability of the study to quantify preference heterogeneity and allow estimation of trade-offs between treatment characteristics

- internal and external validity of the research

- involvement of patients as research partners and their contributions to date, as well as their future planned contributions (e.g. in data reporting and interpretation)

- limitations of the study, in context

- Because qualitative studies are exploratory, they can be undertaken before obtaining scientific advice; however, for quantitative studies, it is better that advice is obtained before the design is finalised.

The rationale for these recommendations and further advice can be found in **Annex Section A4.2**.

The checklist in **Box 4-3** provides broad topics and questions to guide scientific advice discussions, and which can be adapted as required based on the context of the preference study (e.g. the planning of a study as part of a larger clinical study would require modification of item 6). Feedback from regulators and HTA bodies might include the perspectives of patients they recruited as research partners, complementing those of patients recruited as research partners by the study sponsor.

The CHMP Qualification Opinion includes the following comments about suitable topics for scientific advice:

"Many aspects, e.g. sample definition, method selection, experiment/question design, analysis, etc., are in line with important considerations during clinical study design and subject to regulatory assessment. These aspects may qualify as topics for seeking Scientific Advice for a specific development programme."

and

"The focus of advisory interaction with regulatory bodies could be on the choice of the set of comparable alternatives (attribute vignettes), the range and presentation of attributes and their different levels."

Box 4-3. Checklist of topics and questions to guide scientific advice discussions on the context of preparing a patient preference study for marketing authorisation or reimbursement. *This checklist is based on a checklist first presented at the Patient Focused Medicines Development (PFMD) Patient Engagement Open Forum (Hauber, 2021)(112), was subsequently summarised by PFMD, (113) and was further developed by the IMI PREFER authors to apply specifically to interactions with regulators and HTA bodies.*

1. **Use of the preference study:** will it be informative to the regulator and/or HTA body? (see also PREFER component 1 on decision)	Is the research question clearly outlined and will the study generate results that will be useful for decision-makers and relevant to the context (e.g. study endpoint selection, benefit–risk assessment at pre- or post-approval, HTA, treatment selection by the physician)?
2. **Study design:** are the method question formats acceptable? (see also PREFER component 2 on design)	Has the reasoning behind the method selection been appropriately described? Does the regulator and/or HTA body agree that the proposed preference elicitation method and preference question design can adequately capture evidence to inform their decisions?
3. **Benefit attributes:** are the benefit attributes (or the approach to determining them) acceptable? (see also PREFER component 2 on design)	Does the regulator and/or HTA body agree that the: • selected benefit attributes adequately capture what is important to patients, can be tied to planned or investigated clinical study endpoints, reflect what is important to the decision, and are described in a way that is acceptable to the regulator and/or HTA body? • proposed benefit attributes are suitable and that all relevant attributes are addressed in a single framework?
4. **Risk attributes:** are the risk attributes (or the approach to determining them) acceptable? (see also PREFER component 2 on design)	Does the regulator and/or HTA body agree that the: • selected risk attributes adequately capture what is important to patients, can be tied to identified or expected adverse events, reflect what is important to the decision, and are described in a way that is acceptable to the regulator and/or HTA body? • proposed risk attributes are suitable and that all relevant attributes are addressed in a single framework?

5. Other attributes: are the attributes other than benefits/ risks (or the approach to determining them) acceptable? (see also PREFER component 2 on design)

Does the regulator and/or HTA body agree that other proposed attributes (e.g. treatment duration, in- or out-patient treatment setting) adequately capture what is important to patients, reflect what is important to the decision, and are described in a way that is acceptable to the regulator and/or HTA body?

6. Patient population: is the sample including subgroups, sample size, method of recruitment, confirmation of diagnosis, acceptable? (see also PREFER component 2 on sample definition and size)

Does the regulator and/or HTA body agree that the:

- proposed study population inclusion and exclusion criteria are appropriate regarding sample representativeness (disease-related and unrelated), diversity, context of future use (in the case of clinical trial populations)?
- approach to determining the sample size is appropriate to ensure sufficient sample for the proposed sample-level analyses and subgroup analyses?

7. Statistical methods and results: will the results derived from statistical approach(es) be informative? (see also PREFER component 2 on method and analysis)

Does the regulator and/or HTA body agree that the

- instrument design, including selected attributes, can provide reliable results?
- planned validity tests are appropriate and sufficient to demonstrate data validity?

Does the regulator and/or HTA body agree that the proposed outputs of the statistical analysis (e.g. MAR, MAB, preference share) will likely be informative to the regulator and/or HTA body?

5 Methods identification and points to consider for methods selection

Key messages

Patient preference information can be obtained through many different preference exploration (qualitative) or elicitation (quantitative) methods.

Choosing the most appropriate method is a crucial step, and multiple factors should be considered, which can be grouped into three categories: methodological, participant and feasibility. The CHMP supports these three categories of factors affecting methods selection as outlined in the Qualification Opinion.

Some of the most prominent factors that influence the choice of method include its fit to the study purpose and objectives, the preference-sensitive situation, the cognitive burden for the patient, stakeholder acceptance, and available budget and time.

The PREFER recommendations provide detailed descriptions of the five most promising methods and points to consider for method selection.

This section outlines methods used in patient preference studies, both qualitative and quantitative, as well as criteria and points to consider when choosing methods to use in patient preference studies across the medical product life cycle – with a particular focus on the selection of quantitative methods. The original set of criteria were based on research conducted by PREFER 'Methods' Work Package (WP2), published by Whichello and colleagues (8), along with additional methodological points to consider for methods selection that were identified and developed during:

- the development of the PREFER framework and points to consider for methods selection that were submitted to EMA for a Qualification Opinion

- the conduct and comparison of PREFER case studies and their results.

5.1 Types of methods used in patient preference studies

Patient preferences can be obtained by using different exploration and elicitation methods.

- **preference exploration methods:** qualitative methods that collect descriptive data and which analyse the subjective experiences of, and decisions made by, participants.

- **preference elicitation methods**: quantitative methods that collect quantifiable data, which are analysed or inferred using statistical techniques.

Although methods can be grouped according to different classification systems (stated vs. revealed, cardinal vs. ordinal, direct vs. indirect, and compositional vs. decompositional), the PREFER recommendations follow an exploration vs. elicitation methods categorisation.(5)

In 2015, the MDIC stated 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."(34)

These statements have been confirmed in PREFER.(3) Choosing the most appropriate method for eliciting patient preferences in any situation is context-specific and depends on multiple factors.(8) To determine which criteria were most important when selecting a preference exploration or elicitation method in the medical product life cycle, 35 initial criteria were developed based on previous studies, including MDIC's patient-centred benefit–risk framework(34) and a systematic review by Ryan *et al.*(114) Q-methodology was used to identify similar viewpoints across diverse stakeholder groups, and to identify a shortlist of the most important criteria from the participants' rankings.(8)

5.2 Qualitative (exploration) methods

5.2.1 Classification of qualitative (exploration) methods

Ten preference exploration methods were identified (**Figure 5-1**). These methods can be used for in-depth exploration of the patient perspective regarding their disease, treatments, and the importance of outcomes or attributes, and are therefore best suited for early phases in the medical product life cycle. Exploration methods are also often used in a mixed-method approach

to identify what attributes are most important to patients, which, in turn, helps to guide the design of subsequent preference elicitation studies. Individual methods use interviews with one participant (n=1), and this category includes (semi-)structured individual interviews, in-depth interviews, and complaints procedures. Group methods typically direct questions to more than one participant (n>1) and include the Delphi method, focus groups, dyadic interviews, public meetings, nominal group technique, and citizen juries. Individual/group methods like concept mapping can be employed in both settings (n≥1).

5.2.2   Qualitative (exploration) methods selection: points to consider

The recommended criteria for assessing which exploration method will most likely meet the methodological, participant, and feasibility requirements of the study are shown in **Table 5-1**. PREFER identified focus groups, in-depth interviews, semi-structured interviews, and dyadic interviews as potentially the most useful across all stages of the medical product life cycle. Other methods identified as appropriate for certain stages of the medical product life cycle are the nominal group technique in early development and post-marketing, and public meetings in early development and late phase 3.(8)
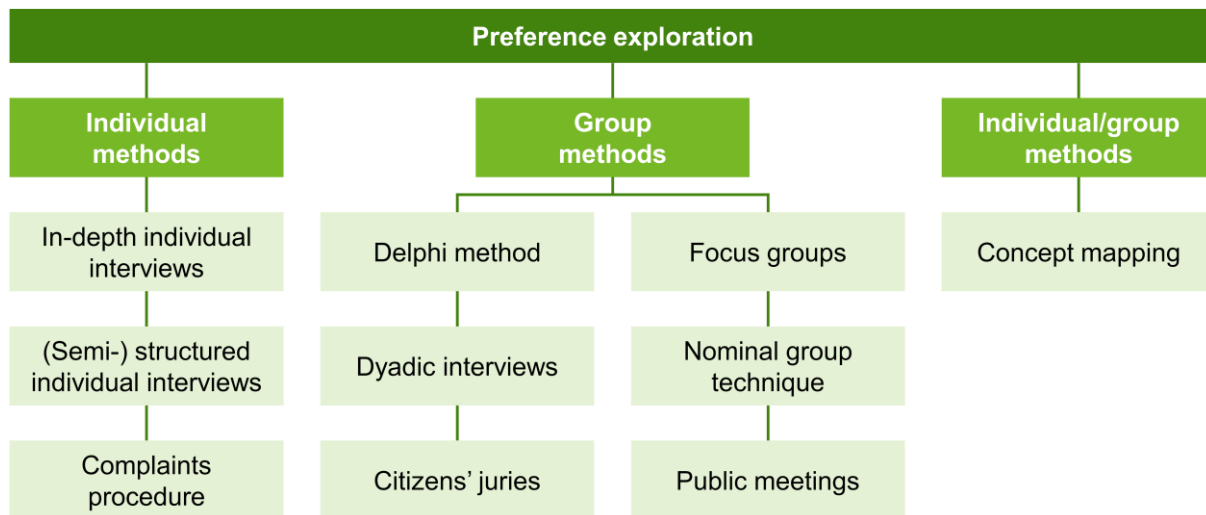
Figure 5-1. Grouping of preference exploration (qualitative) methods into three groups: individual, group, and individual/group methods. *Adapted from Soekhai et al.(5)*

Table 5-1. Performance matrix of exploration methods.(8)

SELECTION CRITERIA	EXPLORATION METHODS									
	In-depth individual interview	(Semi-) structured individual interviews	Complaints procedures	Delphi method	Dyadic interview	Citizens' juries	Focus group	Nominal group technique	Public meetings	Concept mapping
Methodological										
Sample size ≤100	✓	✓	✗	✓	✓	✓	✓	✓	✗	✓
≥8 attributes can be explored	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Easily add new attributes without invalidating previous results	✓	✓	✗	✓	✓	✓	✓	✗	✓	✓
Estimates weights (relative importance) for attributes	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗
Estimates trade-offs between attributes	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
Calculates risk attitudes due to attribute value uncertainty[a]	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
Explores reasons behind a preference in qualitative detail	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Quantifies heterogeneity in preferences	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
Can incorporate internal validation methods	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Establishes external validity	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
Public acknowledgement as acceptable method to study preferences	✓	✓	✗	✓	✗	✗	✓	✓	✗	✗
Participant										
No interaction between participants	✓	✓	✓	✗	✗	✗	✗	✗	✗	✓
Group dynamic with participants	✗	✗	✗	✓	✓	✓	✓	✓	✓	✓
Low cognitive burden on patients	✓	✓	✓	✗	✓	✗	✓	✗	✓	✗
Low complexity of participant instructions	✓	✓	✓	✗	✓	✗	✓	✗	✓	✗
Feasibility										
Low cost	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Quick sessions (≤30 mins) with participants	✓	✓	✓	✗	✗	✗	✗	✗	✗	✗
Low frequency (<2) of sessions	✓	✓	✓	✗	✓	✗	✓	✗	✓	✗
Study duration ≤6 months	✗	✗	✓	✗	✓	✗	✗	✓	✓	✓

✓ = meets criterion; ✗ = does not meet criterion. [a] Risk attitudes such as risk tolerance vs. risk aversion.

5.3   Quantitative (elicitation) methods

5.3.1   Classification of quantitative (elicitation) methods

A total of 23 elicitation methods were identified, which can be grouped into four categories (**Figure 5-2**):

- **Discrete choice-based methods:** typically examine the importance of trade-offs between attributes and their alternatives through a series of choice sets that present (hypothetical) alternatives.

- **Ranking methods:** use ranking exercises to capture the relative order of importance of alternatives or attributes within a presented set.

- **Indifference techniques:** vary the value of one attribute in one of the alternatives until the participant is indifferent to, or has no preference between, alternatives.

- **Rating methods:** use comparative rating approaches, often allowing participants to express the strength of their preferences along a labelled scale.

Depending on the chosen method, these can provide information on preference weights (estimates), relative importance of attributes, trade-offs between attributes including MAR, MRB, willingness-to-pay (WTP), and preference heterogeneity. Moving from left to right in **Figure 5-2**, methods are generally able to answer a smaller subset of research questions. For example, a DCE can provide information about multiple trade-offs and estimate WTP, along with other preference measures, whereas contingent valuation is designed to provide WTP information only. However, methods that are likely to provide more information are also likely to be more complex for participants.

Preference elicitation

Discrete choice-based methods	Ranking methods	Indifference methods	Rating methods
DCE / BWS type 3	BWS type 1	Contingent valuation	Allocation of points
Adaptive conjoint analysis	BWS type 2	Person trade-off	Analytic hierarchy process
	Control preference scale	(Probabilistic) threshold technique	Constant sum scaling
	Q-methodology	Standard gamble	Measure of value
	Qualitative discriminant process	Starting known efficacy	Outcome prioritisation tool
	Self-explicated conjoint	Test trade-off	Repertory grid method
		Time trade-off	Swing weighting
			Visual analogue scale

Decrease in ability to answer multiple questions / Increase in method complexity

Figure 5-2. Grouping of preference elicitation (quantitative) methods into four categories. *Adapted from Soekhai et al.*(115)

5.3.2   Quantitative (elicitation) method selection: points to consider

A key point to consider when selecting an appropriate method is its alignment to the research question(s). Other important considerations are the specific stage in the medical product life cycle for which the study will be conducted, population size and characteristics, and resources (budget and time).

5.3.3 Selection criteria for quantitative (elicitation) methods by stage of the medical product life cycle

PREFER identified 19 criteria for the selection of quantitative (elicitation) methods. While the criteria related to establishing validity, reliability, preference heterogeneity and those ensuring a low patient burden, are relevant for all medical product life cycle stages, the importance of these and the other criteria differs between stages. For example, in early development, sponsors may place greater priority on the study's cost, duration, and sample size, whereas demonstrating internal validity and quantifying preference heterogeneity may become more important in the later stages of medical product development.(8)

5.3.4 Evaluation of quantitative (elicitation) methods against selection criteria and points to consider for methods selection

Preference elicitation methods can be assessed against the 19 methodological, participant, and feasibility criteria (**Table 5-2**). PREFER identified 11 promising preference elicitation methods likely to meet the needs of decision-makers during all stages of the medical product life cycle:

- DCE / BWS case 3

- adaptive conjoint analysis

- (probabilistic) threshold technique

- standard gamble

- time trade-off

- swing weighting

- visual analogue scale

- analytical hierarchy process (AHP)

- BWS case 1

- BWS case 2

- Q-methodology.

Overall, allocation of points, constant sum scaling, and repertory grid method meet fewer of the criteria than other methods. **Table 5-3** shows how some of the PREFER case studies approached their method selection.

Table 5-2. Performance matrix of elicitation methods. *Adapted from Whichello et al, 2020.*(8)

SELECTION CRITERIA	DCE/BWS3	ACA	BWS1	BWS2	CPS	QM	QDP[a]	SEC	CV	PTO	PTT	SG	SKE[a]	TeTO	TiTO	AOP	AHP	CSS	MOV[a]	OPT[a]	RGM	SW	VAS
Methodological																							
Sample size ≤100	✗	✓	✓c	✓	✗	✓	✓	✗d	✗	✗e	✗	✓	✓	✓	✓	✓	✓	✗c	✓	✓	✗	✓	✓
≥8 attributes can be explored	✗	✓c,d	✓c	✓	✗	✓	✓	✗d	✗	✗	✗c	✗	✗	✗	✓	✓e	✓	✓	✓	✗	✓	✓	✓
Can easily add new attributes without invalidating previous results	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗c	✓	✓c	✓	✗	✓	✓d	✗	✗	✗	✓	✗	✓	✓d
Estimates weights (relative importance) for attributes	✓	✓	✓	✓	✗	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✗	✓	✓d
Estimates trade-offs between attributes	✓	✓	✗	✗	✗	✗	✗	✓	✓	✗	✓	✓c	✓	✓	✓	✓	✓	✓	✗	✗	✗	✓	✓d
Calculates risk attitudes due to attribute value uncertainty[b]	✓	✓	✗	✗	✗	✗	✗	✓	✗	✗	✓	✓	✓	✗	✓d	✗	✓	✗	✗	✗c	✗	✓	✗
Explores reasons behind a preference in qualitative detail	✗	✗	✗	✗	✗	✓d	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗
Quantifies heterogeneity in preferences	✓	✓	✓	✓	✓	✓c	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓e	✗	✓	✗	✓	✓
Can incorporate internal validation methods	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Establishes external validity	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗

SELECTION CRITERIA	DCE/BWS3	ACA	BWS1	BWS2	CPS	QM	QDP[a]	SEC	CV	PTO	PTT	SG	SKE[a]	TeTO	TiTO	AOP	AHP	CSS	MOV[a]	OPT[a]	RGM	SW	VAS
Public acknowledgement as acceptable method to study preferences	✓	✗[c,d]	✓[c]	✓	✓	✓[c]	✗	✗[d]	✓	✓	✓	✓	✗	✗	✓	✓	✓	✗[d]	✗	✗	✗	✓	✓
Participant																							
No interaction between participants	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓
Group dynamic with participants	✗	✗	✗[e]	✗[e]	✗[c,d]	✓[c]	✗	✗	✓	✓	✗[c]	✗[d]	✗[c]	✗[c,d]	✗[d]	✓	✓[e]	✗[e]	✓[c]	✗[c]	✓[c,d]	✓[c]	✓
Low cognitive burden on patients	✗	✗	✓[c]	✗[e]	✓[c]	✓	✗	✗[d]	✓	✓	✓[c]	✗[d]	✗	✓	✗[d]	✗	✗[e]	✓	✓	✓[c]	✓	✓[c]	✓
Low complexity of participant instructions	✗	✗[c,e]	✗[c]	✗	✗[c]	✗	✗	✗[d]	✓	✗[c]	✗[c]	✗[c,d]	✗	✗	✗[d]	✗[d]	✗[e]	✓	✓	✓[c]	✓	✗	✗[d]
Feasibility																							
Low cost	✗	✗	✓[c,e]	✓[e]	✓	✓	✓	✗[d]	✓[c]	✓	✓[c,e]	✓	✓[c]	✓	✓	✓	✓[e]	✓	✓	✓	✓	✓[c]	✓
Quick sessions (≤30 mins) with participants	✓	✓	✓	✓	✓	✗	✗	✓	✓[c]	✓	✓[c,e]	✓	✓	✓	✓	✓	✓[e]	✓	✗	✓	✓	✓	✓
Low frequency (<2) of sessions	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Study duration ≤6 months	✓	✓	✓[d]	✓[d]	✓	✓	✓	✓	✓[c]	✓	✓[e]	✓	✓[c]	✓	✓	✓	✓	✓	✓	✓	✓	✓[c]	✓

✓ = meets criterion; ✗ = does not meet criterion. [a] Informed exclusively by literature and not expert interviews. [b] Risk attitudes such as risk tolerance vs. risk aversion. [c] A lack of unanimous consensus in the decision among the experts. [d] Literature conflicted among experts. [e] No clear majority; literature broke the tie.

Table 5-3. Examples of preference elicitation method selection approaches from PREFER case studies.

Study	Approach	Criteria	Selection
PREFER additional academic case study: Patient preferences to Assess Value IN Gene therapies (PAVING) (116) **Medical product life cycle phase: HTA**	Using expert consultation, the study team determined the most appropriate method based on the research question, patient population (rare disease meaning limited population and sample size), and decision context.	The method should: • estimate weights of attributes • estimate trade-offs between attributes • quantify preference heterogeneity • incorporate internal validity measures • not have technical issues • have a low minimal necessary sample size • allow for incorporation in an unsupervised survey.	DCE was excluded due to the large sample size needed for the required number of attributes and levels. Threshold or swing weighting methods were selected as appropriate. Concerns were raised that swing weighting may require support through interviews or workshops (due to complex choice tasks/high cognitive burden) and thus require more resources; therefore, the threshold technique was chosen.
PREFER prospective case study: NMDs (12) **Medical product life cycle phase: early development**	The study team determined the appropriate choice of methods for inclusion based on research objectives and cognitive ability of the sample population.	The study aimed to address benefit–risk trade-off questions for future treatment options and heterogeneity questions, while comparing results across methods. The sample included patients with NMDs affecting the central nervous system leading to 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.	The sample population was divided into subgroups 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 case 2. The later onset subgroup was also given BWS case 2 but was also found capable of completing a more complex survey method (DCE) based on clinical knowledge of the disease state and patient population.

5.3.5 The five promising preference elicitation methods used in PREFER case studies

Based on the relative importance of method criteria and performance, one method from each of the four elicitation method categories was selected for use in the PREFER case studies to answer the research questions **(Figure 5-3**). As five methods could be explored in-depth in case studies, both BWS case 1 and 2 were included. The selected methods included DCE, BWS case 1 and 2, threshold technique, and swing weighting. Following EMA qualification of the PREFER framework and case studies, additional criteria were identified, and an overview of these is provided in **Table 5-4**, along with how these apply to the selected five elicitation methods. See **Annex Section A5** for further detailed guidance on points to consider and on the use of these five methods.
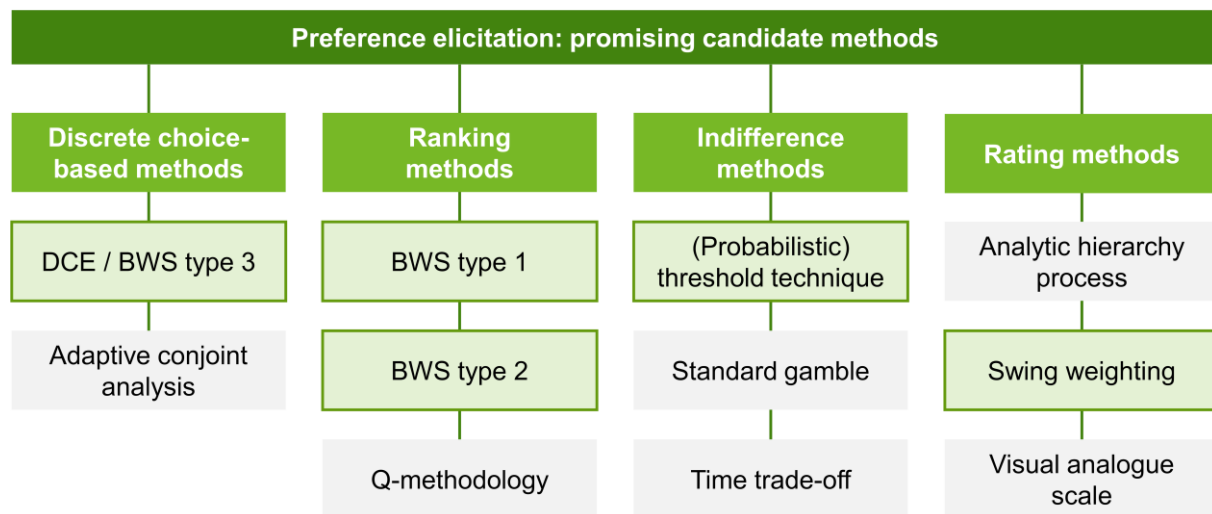

Figure 5-3. From the 11 promising candidate methods, five were selected for use in PREFER case studies.

Table 5-4. Performance matrix of elicitation methods from Whichello et al. 2020 and additional criteria identified during EMA qualification.

SELECTION CRITERIA	ELICITATION METHOD				
	BWS1	BWS2	DCE / BWS3	PTT	Swing weighting
Methodological					
Provides estimates at level of individual	✓	✓	✗	✓	✓
Provides estimates at level of sample/population	✓	✓	✓	✓	✓
Sample size ≤100	✓	✓	✗	✓	✓
Estimates preferences for individual attributes	✓	✓	✓	✓	✓
≥8 attributes can be explored	✓	✗	✗	✓	✓
Can easily add new attributes without invalidating previous results	✗	✗	✗	✓	✓
Estimates preferences over multiple levels of each attribute	✗	✓	✓	✗	✗
Estimates weights (relative importance) for attributes	✓	✓	✗[d]	✗[d]	✓
Estimates trade-offs between attributes	✗[d]	✗[d]	✓	✓	✗[d]
Simultaneous estimation of trade-offs between multiple attributes	✗	✓	✓	✗	✗
Pairwise estimation of trade-offs between attributes	✗	✗	✗	✓	✗
Can accommodate interactions between treatment characteristics	✗	✓	✓	✗	✗
Calculates risk attitudes due to attribute value uncertainty[a]	✗	✗	✓	✓	✓
Explores reasons behind a preference in qualitative detail	✗	✗	✗	✗	✗
Quantifies heterogeneity in preferences	✓	✓	✓	✓	✓
Internal validation methods can be incorporated	✓	✓	✓	✓	✓
Participant					
Establishes external validity	✗	✗	✗	✗	✗
Public acknowledgement as acceptable method to study preferences	✓	✓	✓	✓	✓
No interaction between participants	✓	✓	✓	✓	✓
Group dynamic with participants	✗[b]	✗[b]	✗	✗	✓
Feasibility					
Low cognitive burden on patients	✓	✗[c]	✗	✓	✓
Low complexity of participant instructions	✗	✗	✗	✗	✗
Low cost	✓[b]	✓[b]	✗	✓[b]	✓
Quick sessions (≤30 mins) with participants	✓	✓	✓	✓[b]	✓

✓ = meets criterion; ✗ = does not meet criterion. [a] Risk attitudes such as risk tolerance vs. risk aversion. [b] No clear majority; literature broke the tie. [c] A lack of unanimous consensus in the decision among the experts. [d] This information can be inferred (see **Annex Section A5**).

These performance matrices (**Tables 5-2** and **5-4**) should be interpreted with caution. Results from the case studies show that, while the performance of three methods (swing weighting, BWS case 1, probabilistic threshold technique [PTT]) was consistent with that in the initial assessment at the earlier stage of PREFER,(8) the performance of DCE and BWS case 2 was different. Initially, DCE and BWS case 2 were thought to be cognitively more demanding than other preference elicitation methods, but respondents in the case studies reported that this was not their experience. Similar patterns were reported in all case studies, as well as in studies conducted outside PREFER,(117-122) which called for a re-evaluation of the cognitive demands made by DCEs.

Owing to the ongoing advancements and improvements in preference elicitation, performance matrices of preference methods should be continually re-evaluated. For example, the current matrix is based on a step value function allowing methods to comply or not with a certain criterion ('yes' or 'no' answers), and a more nuanced performance matrix should be developed to allow for a less strict value function. On average, swing weighting and PTT had the highest suitability scores across decision points of the medical product life cycle. The scoring of DCE is high across decision points for the outcome-oriented methods criteria, but lower on the operational aspects. Also, BWS case 1 and BWS case 2 scored similarly across decision points, but overall scored lower than DCE, swing weighting and PTT. In the healthcare setting, DCEs are mostly applied for eliciting preferences; however, other methods should be equally considered when designing future preference studies because, they comply with the top-weighted methods criteria.

5.4   Strengths, limitations, and uncertainties of method selection recommendations

For a full discussion of the strengths, limitations, and uncertainties in these recommendations, see **Annex Section A5.6**.

6 Understanding preference heterogeneity by measuring relevant psychological constructs

Key messages
Investigating participants' psychological characteristics may offer important insights into why preference heterogeneity exists within a study population and/or the factors that influence the formation of patient preferences.
The measurement of psychological constructs in patient preference studies should be evidence-based and/or based on theoretical considerations concerning the strength of their associations with patient preferences and decision-making processes.
In particular, investigators should consider including at least one construct associated with cognitive abilities, such as health literacy and numeracy, given the importance of patient comprehension in preference research.

6.1 Why it is important to consider psychological constructs in patient preference studies

Preference heterogeneity can be defined as "differences in preferences among a sample"(34) with the existence of subgroups of patients with relevant differences in preferences (see **Section 3.5.1** for further discussion of how the PREFER framework addresses preference heterogeneity). In addition to demographic and clinical characteristics, understanding differences in patient psychological constructs (e.g. personality traits, social-cognitive factors, experiences with their disease, treatment, and decision-making styles) can provide important information about preference heterogeneity,(4) and this has been recently underscored by industry, regulators, and HTA bodies.(2)

Psychological constructs are abstract and latent rather than concrete psychological characteristics like age.(38) As a consequence, they cannot be measured directly but only using observable variables. For example, suppose researchers are interested in measuring risk propensity, i.e. people's tendency to take risks.(123) Given that it is impossible to observe and measure this directly, researchers may opt to infer risk propensity by measuring some observable everyday risk-taking behaviour, such as preferring risky travel destinations or choosing between medical treatments, which is a good indicator of this psychological construct. Including a measure of observable behaviour in a preference study would allow the quantification of risk propensity. The use of reliable and valid psychological instruments to measure psychological constructs is therefore essential for measuring and studying individual differences.

Psychological constructs can influence judgments and decision-making, and can be associated with possible differences in both stated and revealed preferences (for a review, see Russo *et al* 2019)(4). They can be categorised into six classes: cognitive factors, motivational factors, personality traits, emotion and mood, health beliefs, and wellbeing.(4, 124) The strongest empirical evidence for the association with preferences and decision-making is reported for the cognitive factors of health literacy and numeracy, and the personality trait of health locus of control. Specifically, health literacy has been demonstrated to be associated with differences in preferences irrespective of specific socioeconomic variables such as age, gender, ethnicity, and level of education.(125, 126) Health locus of control – a generalised expectation about whether one's health is controlled by one's own behaviour or forces external to oneself (127) – has been shown to play a role in shaping patient preferences and decisions;(128) specifically, people with higher internal health locus of control are more likely to prefer and use complementary and alternative medicine. Other psychological constructs vary in their associations with preferences and decisions (for a detailed overview, see Russo *et al*, 2019).(4)

Given the importance of understanding preference heterogeneity, and potential impact of patients' psychological characteristics on preferences,(4) PREFER provides guidance for including psychological characteristics in patient preference studies.

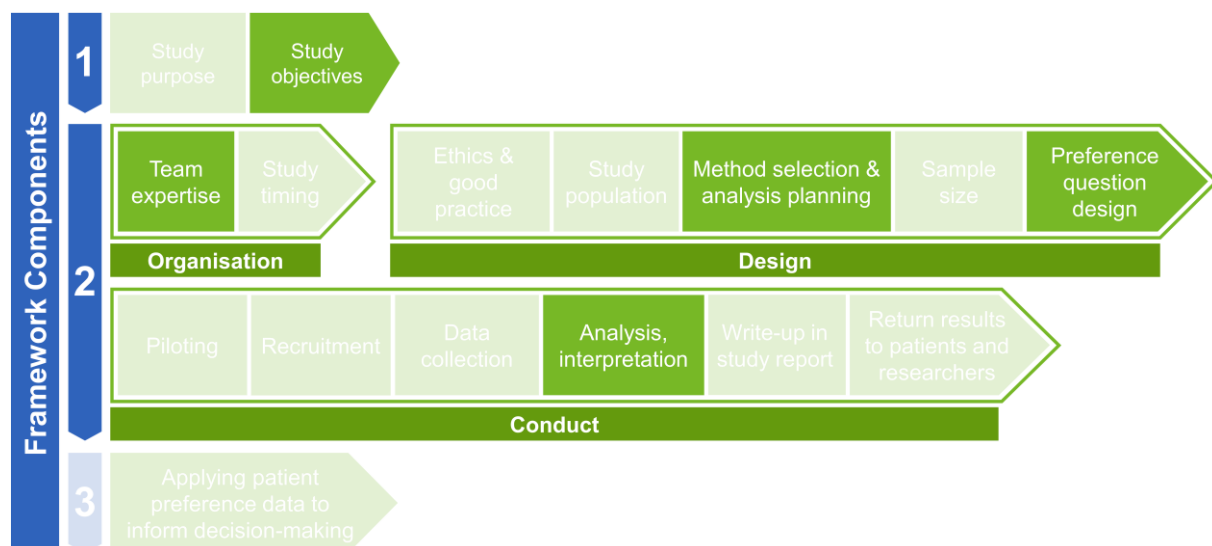## 6.2 When to consider using psychological constructs in patient preference studies

Figure 6-1. Stages of PREFER framework that should include the use of psychological constructs.

6.3 How to apply psychological constructs in patient preference studies

It is recommended that study teams use a systematic and evidence-based approach to identify the most relevant psychological constructs, measurement tools and feasibility assessments (**Figure 6-2**).
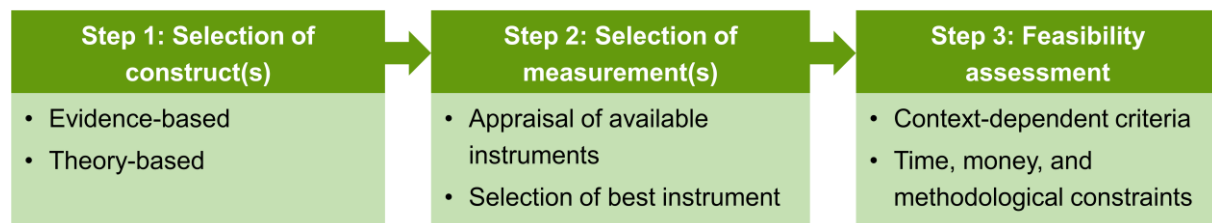

Step 1: Selection of construct(s)	Step 2: Selection of measurement(s)	Step 3: Feasibility assessment
• Evidence-based • Theory-based	• Appraisal of available instruments • Selection of best instrument	• Context-dependent criteria • Time, money, and methodological constraints

Figure 6-2. The recommended steps for including psychological constructs in patient preference studies.

6.3.1 Step 1: Which psychological construct(s) should be included in patient preference studies?

The decision to evaluate psychological constructs as part of a patient preference study should be evidence-based and based on theoretical considerations relating to their association with patient preferences, the decision-making processes and the preference study objectives. PREFER has made considerable progress in identifying which constructs should be considered for preference research, and has developed a robust classification system based on systematic reviews of existing literature and consensus-based recommendations. These classifications were developed through a two-round Delphi panel involving experts in health preference research, clinical and health psychology, health economics, public health, risk communication, and decision-making.(23) (For an overview, see Russo *et al*, 2019.)(4) Based on the results of this consensus-based process, psychological constructs have been split into three classes depending on their suitability for use in decision-making processes and patient preference studies:

- **class I:** unanimity or high levels of agreement among experts (i.e. ≥80% in agreement)

- **class II:** agreement among a majority of experts (i.e. ≥70% and <80% in agreement)

- **class III:** disagreement among experts (i.e. <70% in agreement).

Class I constructs includes health locus of control, health literacy, health numeracy, control preference, illness perception, patient activation, risk propensity, and treatment-related beliefs. Health literacy, health numeracy, and health locus of control have the greatest empirical evidence supporting their association with preferences and decision-making processes (**Annex Table A6-1**). It is recommended that at least one construct associated with cognitive abilities, such as health literacy and numeracy, is considered for inclusion

given the importance of patient comprehension in preference research. As stated below, the assessment of these constructs may be especially relevant to identify patients who may be unable to properly understand numerical and medical information, or the scenario proposed by the researcher.

Class II constructs include autonomy preference, decision-making style, and health orientation. The decision to include or not include a class II construct should be made by balancing the advantages of acquiring additional information regarding preference heterogeneity with the disadvantages of having less empirical evidence supporting its link with decision-making. The decision should also take into account operational considerations, such as potentially increasing the cognitive load for patients, costs, and the time needed to implement the research and analyse the results.

Among the 11 constructs across class I and II, health literacy and health numeracy were identified by experts as constructs that can help understand differences in patient preferences due to patients not understanding study questions. In this regard, patient preference heterogeneity between subgroups with different levels of health literacy or numeracy might not reflect true differences in benefit–risk preferences but can be simply due to biases from lack of patients' comprehension of complex numerical and/or medical information or the choice task. The remaining nine constructs were considered as those that could identify relevant psychological characteristics that account for patient preference heterogeneity.

Class III psychological constructs, including anxiety and depression, are those for which there is a lack of expert consensus regarding their potential contribution towards describing preference heterogeneity. It should be noted that this lack of consensus can be due to a lack of empirical evidence for the relationship between preferences and the psychological dimensions considered.

6.3.2   Step 2: How should psychological measures be appraised and selected?

The need to appraise and identify the best psychological instruments for evaluating the targeted psychological aspect(s) is especially relevant when more than one instrument is available to measure the same construct, and shares many aspects with those applied to the development and validation of patient-reported outcome (PRO) measures to support claims in medical product labelling.(129) For example, validity and reliability are relevant properties for appraising both PRO and psychological measures. In the same way that a PRO instrument should be shown to accurately and reliably measure a specific aspect of a patient's health, so a psychological tool should be proven to be a valid and consistent measure of the theoretical construct. Moreover, just as a PRO instrument is credible only when there is accumulating evidence of its functioning in the target population of the clinical trial, psychological measures should be used in preference studies only when there is

documented evidence of its performance in the study population. When no empirical evidence is available, researchers should consider piloting and/or running preliminary validation studies to evaluate the overall psychometric properties of the measurement instruments. For cases where preference heterogeneity is relevant to the decision-making processes (see **Section 3.5.1**), more validation will be needed to increase the strength of evidence. A list of available measures for evaluating psychological constructs can be found in Russo *et al*,(4) and in **PREFER's task 2.5 deliverable**.

A checklist for appraising and selecting psychological measures to be included in a patient preference study was advanced within the PREFER project.(23) Some of the more salient considerations for assessing the overall quality and usability of each measurement instrument are summarised here.

- **Is the instrument valid and reliable?** Psychological instruments should produce valid and reliable results. While validity relates to the extent to which a tool measures what it is intended to measure,(130) reliability quantifies the precision of psychological measures in terms of internal consistency or the consistency of observed scores across different administrations of the same instrument.(131) Psychometric validation is a complex process for proving the validity of a psychological instrument in a specific population, and/or context of use. When a psychological instrument has been insufficiently validated in the target population, additional pre-testing and/or validation studies should be considered.

- **Has the instrument been translated and culturally adapted for the target population***?* Psychological instruments are generally developed, constructed, and validated for a specific language and culture. They can be adapted for use in a different target population through a complex 'cross-cultural adaptation', that aims to reach equivalence between the original and new versions.(132) This time-consuming and costly process involves the forward-and back-translation and adaption of individual items, instructions, and response options, the consolidation of the pre-final version of the instrument, and a pretesting phase in people from the target population (see also **Section 3.3.2.5**). It is strongly recommended that this is then followed by an appraisal process where the developers, the patient population in question (or an advisory committee) review the adaptation. When a psychological instrument requires translation, the recommendation is to follow the ISO 17100 standard and conduct pre-testing and cognitive debriefing with a few patients.

- **What is the outcome measure of the instrument?** Some psychological instruments return only raw scores, which can be difficult to interpret without knowledge of how one score compares with that of a norm-referenced group. Other psychological tools adopt a reference group and use standardised scores, which provide a clearer and more valid picture of each patient for the psychological differences being measured and are therefore preferrable to instruments that only generate raw scores.(4)

- **Does the instrument provide cut-offs for classifying patients?** Some standardised psychological instruments have cut-off scores that can be used to classify patients into groups and are generally used for screening purposes to differentiate clinical populations from non-clinical ones. For example, they can help in detecting people with mild or even severe depressive symptoms. Other psychological instruments have cut-off scores to differentiate people with adequate or inadequate abilities, such as identifying people with poor or satisfactory health literacy. Moreover, this kind of tool can help to better understand the effect of a specific psychological characteristic on patient preference heterogeneity; for instance, through their inclusion in latent class analysis models or in subgroup analysis. Thus, if relevant, psychological measures providing cut-off scores are preferable to those that do not.

- **Is the instrument protected by copyright/license?** The non-commercial use of psychological instruments is often allowed without requiring written permission, a license or a fee.(133) However, some developers can require explicit approval to use their instrument even if there is no cost. Others use the copyright status and put limits on use, adaptations, and translations. Finally, other developers can charge a licensing fee, an administration fee, or a fee for obtaining scoring instructions. Teams should budget accordingly to account for these potential needs.

6.3.3   Step 3: How to assess the feasibility of measuring psychological constructs within a specific patient preference study?

When evaluating the feasibility of including psychological constructs, stakeholders balance time, cost, methodological constraints, and cognitive burden against considerations related to the preference study objectives and the need for further investigation of candidate psychological constructs with patient preference heterogeneity. This feasibility assessment can be conducted by following a checklist of interrelated criteria that are highly context-dependent for a given study, such as:

- **What is the total cost associated with the use and implementation of the instrument in this specific patient preference study?** Investigators should consider sources of additional cost for measuring psychological differences in their study, including any fees payable for using the tool (see Step 2, final bullet point, the cost associated with administration and data collection (see points below), and the cost associated with scoring, analysing, and interpreting results from psychological profiling.

- **What is the time needed to implement, administer, and analyse the results in this specific patient preference study?** The design, administration, and data analysis of individual responses arising from psychological measures can be very time-consuming, and investigators should estimate and plan for any additional time needed to complete their study. Specifically, it is important to minimise the respondent burden of study participants, which encompasses cognitive burden, the time commitment needed, and participants' perceived psychological, physical, and economic discomfort caused by

participation in research.(134-136) Psychological instruments can vary in both their completion time and required cognitive effort. Patient research partners can help identify the most suitable instruments by pretesting and piloting studies to evaluate and compare the respondent burden across alternative psychological measures.

- **Does the production, use, or analysis of the instrument require specific software?** Does the instrument require a computer/device? Some psychological instruments require specific software or a compute/device to enable the administration, data collection, and analysis of patient responses and outcomes. The feasibility of using administration software or a computer/device (along with potential costs) should be evaluated by considering its possible harmonisation with the specific research design and target population of the patient preference study.

- **Does the administration of the instrument require the presence of a researcher/clinician?** Psychological instruments can vary in how they are administered. For example, compared with interviewer-administered instruments, self-administered measures are designed to be completed by a respondent without the intervention of a researcher/clinician. Self-administered instruments can be distributed through web surveys, email, or in-person to large groups. The main advantage of self-administered measures is that they can be delivered to a large sample of people quickly with less effort and cost. Interviewer-administered instruments can only be used in face-to-face surveys or interviews. The main advantage of interviewer-led instruments is that they enable probing for additional information thereby obtaining superior quality information.

- **What is the time commitment required from your patient population?** The lung cancer case study was originally planned to measure illness perception to test its influence on patient preference heterogeneity. However, this was dropped after pre-testing of the survey because patients agreed the survey was too long and illness perception was judged to be the least relevant construct.

6.4   How psychological constructs can explain preference heterogeneity and related results

This section includes three examples of how PREFER case studies measured different class I constructs – health literacy, health numeracy, and illness perception – and related the results to patient preference heterogeneity. Results from PREFER case studies are generally consistent in supporting the association of health literacy with observed differences in patient preferences. Specifically, people with different degrees of ability to understand and use health information display meaningful differences in their stated preferences. However, because health literacy and health numeracy can also be linked with people's difficulties in understanding information and choice tasks, future studies might investigate whether these two cognitive factors could account for true patient preference heterogeneity, or if differences

in benefit–risk preferences are only due to patients not understanding or not interpreting the patient preference study questions and information as meant by the researcher.

Health literacy

In the **lung cancer case study**, patient preferences were elicited through DCE and swing weighting. Patients were categorised as having 'adequate' or 'limited' health literacy based on responses to a brief health literacy screener (Chew *et al*, 2004).(137) This dichotomous variable was then used to explain membership to two different classes identified through latent class analysis of DCE data. Specifically, patients with adequate health literacy were more likely to belong to class 1, in which people preferred oral treatment over infusions. In the **PAVING case study**, scores from the brief health literacy screener (137) were used to identify patients with adequate vs. inadequate health literacy, but the research team adopted a different analytical approach. Instead of entering this dichotomous variable in latent class models or subgroup analysis to explain patient preference heterogeneity, researchers focused on the correlation of health literacy with performance on a DCE comprehension task. The results demonstrated that adequate health literacy was positively correlated with correct comprehension, and that patients with inadequate health literacy may not have completely understood the choice task.

Health numeracy

In the **MSD case study**, the effect of patients' subjective health literacy on their preferences for anti-thrombotic treatments was explored using a latent class model. Following identification of different classes/groups of preferences, the probability of belonging to the different classes was analysed as a function of subjective health literacy – as well as patients' other observable characteristics and health numeracy – in a multinomial logit model. The latent class model identified two distinct classes of preferences. Patients with an inadequate level of health numeracy were more likely to fall into class 1 compared to patients with an adequate level. The effect of patients' subjective numeracy skills – as well as health literacy – on preferences for anti-thrombotic treatments were explored using a latent class model. While health literacy was related to observed patient preference heterogeneity, health numeracy skills were not related to class membership.

The measurement of health literacy and health numeracy is particularly useful because it can additionally be used for screening purposes. Identifying patients with inadequate health literacy and/or health numeracy allows researchers to support and empower them through the implementation of enhanced educational material so they can better understand and interpret information and tasks within preferences studies (**Section 7**).

Illness perception

In the **rheumatoid arthritis case study**, illness perception was measured so that its effect on preferences for very early treatment of rheumatoid arthritis could be evaluated using a latent class model. The case study assessed treatment preferences of both first-degree relatives (FDRs) and the general public for preventive therapies for rheumatoid arthritis. Illness perception was measured through an adapted version of the Brief Illness Perception Questionnaire (BIP-Q) for lay persons (i.e. individuals without the disease) that has been used in previous research in other disease areas. The median split of the eight scores from the BIP-Q was used to classify participants as having low or high levels relating to the following factors: consequences, timeline, personal control, treatment control, identity, illness concern, coherence, and emotional representation. Following identification of different classes/groups of preferences with the latent class model, the probability of belonging to the different classes was analysed as a function of illness perception – as well as health literacy, numeracy, perceived risk of developing rheumatoid arthritis, and beliefs in medicine – in two different models: one for the general population and one for FDRs. In the former, the latent class model identified five distinct classes of preferences and several of the eight illness perceptions subscales contributed to preference heterogeneity. In FDRs, the model identified two different classes of preferences, but none of the illness perception subscales significantly contributed to the class membership.

7 Addressing educational needs among preference study participants

<div>

Key messages

Patient educational materials should be customised to the target population, choice task, and study context by using a systematic and evidence-based approach.

Educational materials should be planned, developed, and tested in collaboration with patient research partners and participants early in the study to ensure maximum suitability and effectiveness.

The format and content of educational materials should be selected and adapted based on the specific educational needs of the participants.

</div>

7.1 Why educational materials in preference studies should be systematically developed

Key to stakeholder acceptance of preference data as valid scientific evidence is ensuring that patient preferences are well-informed.(37, 138-140) This requires that patients participating in preference studies understand the design and context of the study, the disease, and the choice tasks in which their preferences are explored or elicited.(1, 37, 85, 141) The importance of patients' understanding is highlighted in current regulatory and HTA guidance, and best practice guidelines.(37, 139, 140) For example, the FDA has identified a minimum of 11 quality criteria that should be fulfilled for preference data to be accepted as valid scientific evidence, including the need for effective communication of benefits, harms, risks, uncertainties, and general understanding.(37) The German HTA body, IQWiG, states that:

> *"Studies on the assessment of patient preferences can be very easily biased if the following aspects are not considered very carefully…the questions have to be asked in an understandable and open way, and the options described and their advantages and disadvantages have to be realistic".*(139)

Additionally, ISPOR describes *"confusion or misunderstanding or unobserved heterogenous interpretations"* as a key source of measurement error.(140) Moreover, the importance of adequately informing patients while avoiding information bias has been broadly recognised by industry, regulatory, and HTA stakeholders during interview and focus group discussions.(6) Previous research has found that patients can find it challenging to understand health information, especially when it relates to risks and benefits of interventions as presented in preference elicitation tasks.(142) This is likely amplified in contexts where patients have little or no experience of the disease, treatments, or choice

tasks, especially in scenarios where there are complex trade-offs between treatment characteristics.

To address the challenges of patients' understanding in patient preference studies, guidelines recommend that all surveys start with an educational section to explain its context and aim.(85) Although education in patient preference studies has traditionally been text-based, recent transitions towards digitalisation has enabled the exploration of enhanced multimedia educational materials to achieve these objectives.(143) The rationale for incorporating enhanced educational materials in patient preference elicitation studies is based on their ability to:

- mimic real choice situations (through their design)

- assist in education and preference construction (through exploration)

- engage and motivate study participants.

Although there are no regulatory guidelines for the format, content, and development of enhanced educational materials, guidelines for preference research offer some insights, and recommend that they should be evidence-based, and scientifically and clinically validated.(37) Enhanced educational materials should be developed to maximise learner engagement, reflection, and motivation so that learners perform better than with traditional approaches. The level of interactivity to achieve the desired educational objectives can be varied according to the educational needs.

7.2 When to develop patient educational materials

This section provides more information on the development of educational materials to primarily support **framework component 2: preference questions design and piloting** (**Figure 7-1**).

The development of patient educational materials should occur at several stages in the design of a preference study, and should be initiated after the patient population, choice task, and context have been decided (after **framework component 1**) because this informs the required content. Early development and pre-testing of educational materials with patient partners is important to enable sufficient time to make revisions before they are incorporated in the main preference elicitation study. A critical last step in the process is thorough testing of both the enhanced educational materials and the survey itself in the final test environment. Testing should be performed using browsers and devices (tablet, laptop, mobile phone) that participants are expected to use, particularly as usability issues (e.g. formatting issues, broken functionality) can be specific to certain device–browser combinations and could lead to low participant satisfaction and/or dropout. It is recommended that final testing is performed by those who were not involved in the development of the tool and are naïve to the survey contents.
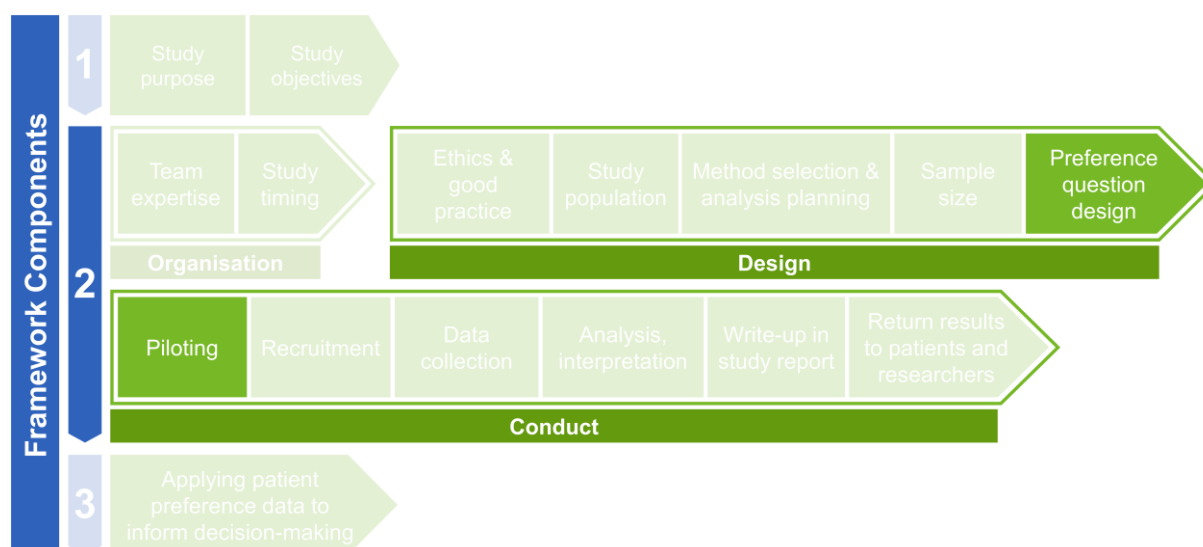
Figure 7-1. Stages of PREFER framework in which educational materials for patients should be designed and tested.

7.3   How to develop educational materials for patient preference studies

It is recommended that researchers use a systematic and evidence-based approach to identify educational needs and to design and develop educational materials in a patient preference study. This approach helps narrow down a broad list of possible educational features – including, for example, those related to format, motivation, cognition, content, and tests – to clearly identify which educational features are required, optional, or inappropriate for a given case study.(144-146) Additionally, although customisation is very important, care should be taken to avoid over-engineering development of the educational materials, particularly for preference studies that are more exploratory in nature. Traditional written text can be the best option in cases where the content is familiar and easy to understand.

One important consideration with the development of enhanced educational materials is that the timelines and costs required to develop these materials are often unknown at the time of initial planning and budgeting. The required timelines and costs, therefore, need to be balanced with the requirements for education of the study population early in the process. Budgets can be managed more closely by applying an iterative, stepwise approach, as seen in software development.(146) Additionally, the costs for developing educational materials can vary between developers, and are driven, in part, by the level of interactivity, visual design elements (e.g. animations), and volume of content.

A systematic, evidence-based approach for developing educational materials is provided in **Section 7.3.1**, and an application of a stepwise approach for a PREFER case study is

presented in **Section 7.3.2**, with an additional example available in **Annex Section A7**. A list of appraisal criteria for selecting educational features can be found in the **PREFER's task 2.5 deliverable**.

7.3.1   How to develop educational materials using a stepwise approach

The following stepwise approach is recommended when developing educational materials for preference research studies (**Figure 7-2**). It is strongly recommended that patient research partners and healthcare professionals are involved in each of these steps.
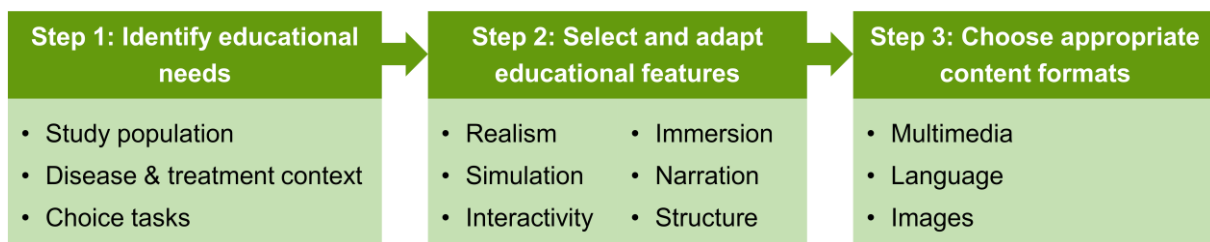
Step 1: Identify educational needs	Step 2: Select and adapt educational features	Step 3: Choose appropriate content formats
• Study population • Disease & treatment context • Choice tasks	• Realism • Immersion • Simulation • Narration • Interactivity • Structure	• Multimedia • Language • Images

Figure 7-2. Stepwise approach for developing educational materials for preference studies.

Step 1: Identify educational needs

The educational needs of the participants should be assessed by considering the study population, disease context, available treatments, and choice tasks determined by the selected preference elicitation method.(147)

Educational needs can differ depending on the age, level of health literacy and numeracy, and cognitive ability of the target population. For instance, for younger patients, more emphasis should be placed on readability scores, while for older patients, this should be on numeric, verbal skills, or visual skills. The patient population can also be inexperienced with advanced symptoms of the disease or possible side-effects of treatment options. For example, the development of educational materials for at-risk or early-stage disease populations may need to focus on disease awareness, whereas this is less important for those with more advanced disease. Moreover, as novel choice tasks can be unfamiliar to participants, some preference methods may need to be explained carefully so that participants understand how to complete the survey. In addition, for surveys with more detailed instructions or a greater number of questions/profiles, there can be a need to design the educational material so that participants stay engaged and stimulated while completing it.

Step 2: Select educational features

The selection and adaptation of appropriate educational features should be based on the educational needs identified in step 1 and consider the following aspects:

- **Realism:** relates to the amount of contextual information provided to participants so that the choice task becomes more relatable. A low level of realism means participants receive only basic facts and figures. A high level of realism means they receive more contextual information about the topic, enabling greater insight and understanding into how the facts and figures relate to real-world situations.

- **Simulation:** the extent to which participants can see the outcomes of different decisions and choices made in a preference-sensitive context. A low level of simulation focuses only on framing situations, circumstances, and consequences of decisions. A high level of simulation will additionally depict multiple decision-making scenarios so that participants can more easily visualise the short and long-term impacts of those decisions.

- **Interactivity:** the amount of input that participants provide to an educational tool to guide the information that is presented to them. A low level of interactivity means participants simply receive information with only basic navigation. A high level of interactivity requires participants to absorb, reflect on, and submit information back to the tool for active learning. It also adds layers of freedom and fosters motivation by enabling participants to explore content in multiple ways.

- **Immersion:** the extent to which participants are mentally engaged with the choice task and which can be tailored using different technical features. A low level of immersion helps build explicit knowledge for passive understanding. A high level of immersion helps build more cognitive, somatic tacit knowledge that can be put into real-life practice, and fosters motivation and engagement.

- **Narration:** the use of stories that connect and explain a carefully selected set of realistic events and experiences. A low level of narration focuses on facts, whereas a higher level provides a more in-depth, consistent, and coherent story about situations and contexts, including many emotive aspects. At the highest level, narration includes storylines that unfold differently depending on the decisions made by the participants.

- **Structure:** the structure of educational materials can determine what options participants have to choose information. For example, information can be layered so that all participants received a minimum level of content, but additional, more in-depth, content can be made optionally available for those with particular interest in finding out more.

In general, higher levels of these aspects should be considered when higher educational needs are identified. In preference studies where there is a greater need to keep participants engaged and stimulated, higher levels of interactivity, immersion and narration should be considered.

Step 3: Choose appropriate content formats

This step includes the selection and adaptation of formats, such as multimedia, language, and images, that are appropriate to the needs identified in step 1.

It is important to align the educational formats with the needs for the target population. For example, if the target population has problems accessing digital platforms, printed materials should be considered to minimise the degree of potential selection bias. In addition to text, other formats, such as visuals and voiceovers, may be considered. When planning a preference study hosted on a digital platform, researchers should consider the devices that participants are most likely to use to access the preference survey and educational materials (or, in clinical trials, which devices sponsors will supply to participants). For example, knowing if participants will typically use a mobile phone or a laptop is informative. This will help determine the conditions under which to develop enhanced educational materials or not and will inform which aspects are technically possible to integrate, particularly in the case of multimedia materials. For example, hover interactions (such as when a user moves their mouse cursor over specific content to display additional content like as pop-up text box) do not function on tablets, so alternative methods of user interaction should be implemented if researchers expect participants to use these devices. Another example is the use of complex figures and animations: these will be optimally viewed on a laptop or desktop computer yet could present challenges on mobile phones where browsers may not be responsive or may downsize content below a readable level (e.g. size 2 font). Participants should be informed of any technical limitations and advised on the devices/browsers that offer an optimal viewing experience; alternatively, researchers could choose to restrict which devices can access the survey to ensure that optimal viewing and usability is achieved.

Although preference researchers can be aware of basic technical considerations, there are unique aspects to the integration of enhanced educational materials that researchers should consider when designing online preference studies,(143) which can require consultation with experts outside of the research team. Because digital preference surveys and enhanced educational materials are typically developed using different software, they should be integrated with care. Where possible, enhanced educational materials should be directly integrated into survey design software, which should reduce survey complexity for participants and enable them to access all components within a single digital environment.

The language used should be adapted to the target population and the relevant clinical decision-making context, and should be a collaboration between patient research partners and healthcare professionals. Careful consideration should always be paid to the selection of language and images to ensure participants understand risks and to avoid framing effects.(148) Accurately communicating risks is challenging, and participants often find it difficult to understand and interpret the risks presented to them. Although the literature on risk communication in healthcare as a whole is extensive, the literature specifically relating to preference studies is considerably more limited.(147, 149) Nevertheless, general recommendations can be made:

- risk is better communicated using numbers than using words such as 'low', 'medium' or 'high'

- absolute risk estimates are preferred over relative risk estimates

- frequencies are preferred over percentages

- adding images to support numerical or text content is effective for increasing understanding, but they should still be clear, well-planned, and tailored to the context.

7.3.2 Example of developing educational material using a stepwise approach

An example for the stepwise approach taken by a PREFER case study team is presented below, along with important considerations during the process. A **short film** illustrating the education materials and design is also available.

Example case study: Uppsala-Rheumatoid Arthritis

This study elicited preferences among patients with rheumatoid arthritis from Sweden and used enhanced educational materials.(17) One aim of the study was to compare the influence of enhanced educational materials on patient preferences elicited in a DCE with traditional written information.

Step 1: Identification of educational needs

The educational needs of the study population were assessed with interviews with patients, input from two rheumatologists, and patient research partners.(17) Most patients with rheumatoid arthritis had a relatively high level of education (medium level 22%, high 47%), as well as several years of experience with the disease and treatment options (18% had 5–10 years, 51% had >10 years). The educational needs were identified as moderate, and related to the disease, the different treatment alternatives, and the description of the survey choice tasks. Because DCE was selected as the most appropriate method to answer the research questions, consideration was given to the best way of introducing the choice tasks in the survey. Particular focus was given to the attributes and their levels because responding to a DCE can be challenging for patients unfamiliar with scenario-based questions composed of attributes with varying levels.

Step 2: Selection and adaptation of educational features

Educational features were selected based on the educational needs identified in step 1. A moderate-to-high level of realism was chosen (**Figure 7-3** and **Figure 7-4B**). Decisions were made to not include actual photographs of swollen and destructed joints given the potential negative psychological reactions among participants, and artwork was used instead. Input from patient research partners was influential in this regard. Moderate interactivity levels were chosen so patients could click through the attributes and levels at the interface (**Figure 7-4A**). Patients were required to review all descriptions before advancing to the DCE. No simulations or narratives were included given the experience and education level of the participants. Patients could not speed-up or skip content.

Step 3: Choose appropriate content formats

As in the previous step, the following choices were made based on the educational needs of the study population and case study. The majority of patients had experience with the disease and treatment options although this level of experience did vary. The selections of educational features and formats made in step 2 and 3 were made for two reasons. Firstly, to aid a common understanding of the attributes associated with the treatments in the study, and, secondly, to ensure that patients' responses did not rely solely on previous experience, particularly if they were unfamiliar with the type/level of effectiveness or potential risks associated with the treatments under evaluation. The selected formats were integrated into a simple and intuitive user interface, overlayed with a Swedish voiceover (because the study was conducted among participants in Sweden), which included line-drawn visuals, stock images, and supportive icons (**Figure 7-4A**). The on-screen text was concise and focused only on key messages. The attribute levels were presented as illustrations, written text, and a voiceover, which was the same as the on-screen text. Patients could click an attribute to find out more about it, including its levels (**Figure 7-5**).

The first draft of the written descriptions and selected key terms were based on the scientific and regulatory literature, and it was then refined in several meetings with rheumatologists and patient research partners. The function of the imagery, together with the text and voiceover, was to inform about the choice tasks, the disease, and the benefits and risks of the treatment.

Overall, the educational materials in this case study functioned well. However, the survey would have benefitted from comprehension tests that would have enabled a more direct evaluation of patients' understanding of the materials, and patient partners could have been involved at the planning stage of the study.

Results of various assessments of the same educational materials can be found in **Annex Table A7-1**.

7-3A

7-3B

Figure 7-3. Example of multimedia graphics used to educate patients with rheumatoid arthritis from Sweden for a DCE in a preference elicitation study. *Figure 7-3A* shows the joints of a slightly affected hand. *Figure 7-3B* shows severe joint destruction that is a likely consequence of untreated rheumatoid arthritis.
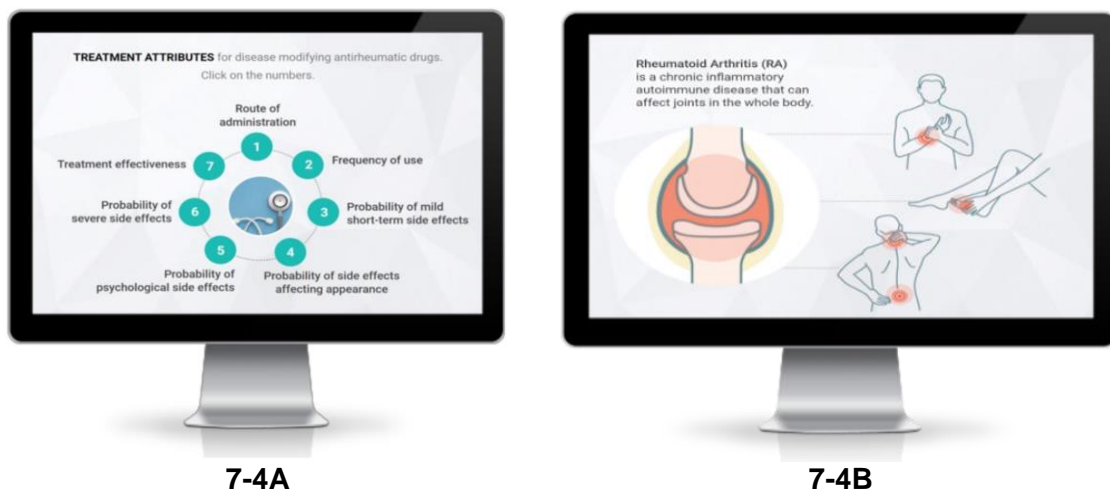

7-4A

7-4B

Figure 7-4. Example of multimedia graphics used to educate patients with rheumatoid arthritis from Sweden for a DCE in a preference elicitation study. *Figure 7-4A* shows an interactive user interface where the patient clicks on each of the treatment attributes for a more detailed description. *Figure 7-4B* shows further detail about the disease.
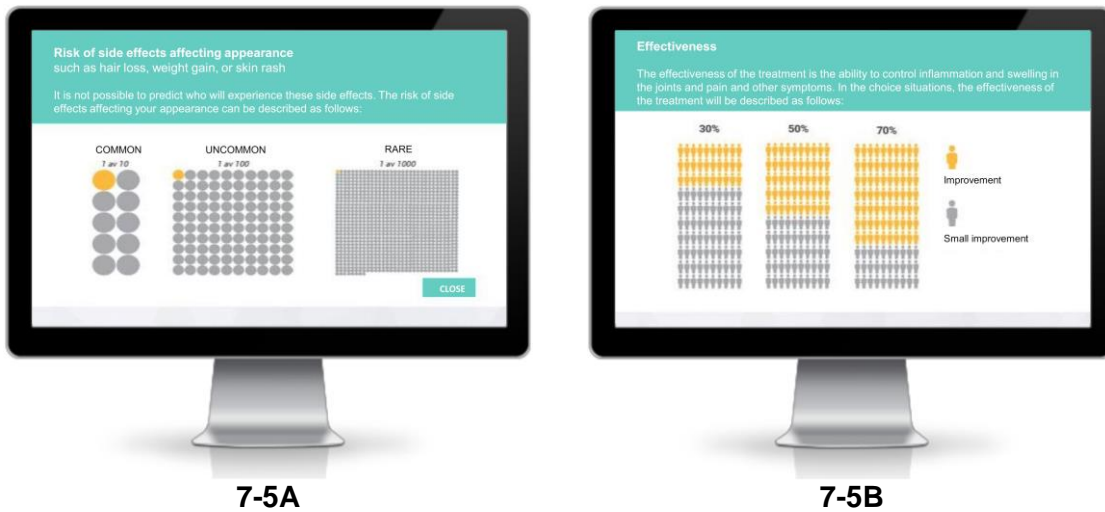
7-5A

7-5B

Figure 7-5. Example of attributes in the DCE presented by using pictograms (*7-5A*) and icon arrays (*7-5B*).

8 Future areas of research

Key messages

PREFER was able to provide a foundation on when and how to run patient preference studies but there are still questions that require additional experience or additional examination beyond this six-year project.

Future research should involve collaboration between all stakeholders – including guidance from scientific societies, regulators and HTA bodies – to build upon the results from PREFER and increase the use and understanding of patient preference studies.

8.1 Background

Patients and their advocacy groups, regulatory and HTA bodies, and industry are increasingly advocating for patient engagement in health and healthcare decisions across the medical product life cycle. Despite the growing use of patient preference studies, there are unanswered questions about their methodological requirements, including best practices on how and when they should be used to inform decisions. PREFER was launched in 2016 to answer these questions with the specific goal of providing recommendations on how and when to assess, engage and include patient preferences during the medical product life cycle.(72) PREFER brought together experts from academic research institutions, pharmaceutical companies, patients, HTA bodies, and small- and medium-sized enterprises. In addition, stakeholder advisory groups were included to ensure that recommendations were evidence-based, relevant, and useful. The resulting PREFER recommendations are based on methodological preference research conducted alongside clinical PREFER case studies.

8.2 PREFER research agenda

Early stages of PREFER focused on assessing the current state of patient preference studies to inform PREFER's research agenda. This was developed in a multi-step approach to identify, refine, and rank questions that could be suitable for examination in proposed case studies.(24) Consortium members ranked the methodological and clinical questions by importance, which were then clustered into the major themes,(24) the highest priority themes being:

- the reliability and validity of different preference methods
- generalisability and transferability
- the impact of educational materials on preferences.

Based on the final rankings of priority research questions, each PREFER case study was encouraged to include, where possible, the following:

- two preference methods for comparison of results

- psychosocial instruments, including those that assess participants' numeracy, literacy, and attitudes toward risk or health-decision making

- educational tools to assess participant understanding and preference consistency

- assessment of preference heterogeneity.

Lastly, in addition to clinical questions, PREFER case study teams were encouraged to include any additional priority methodological question(s) for assessment within the constraints of the case study context.

Ten prospective case studies addressing methodological questions were performed within PREFER. Because the case studies could not address each prioritised research question, choices were made based on what was feasible. If a priority research question was not addressed in a PREFER case study,(24) it had often been studied elsewhere, including in PREFER methods research or simulations. Briefly, changes in the number, type, and definitions of attributes have been studied in PREFER simulations (150) and otherwise.(151, 152) Attribute framing,(153-156) attribute presentation formats (143, 157, 158) and preference evaluation over time (159, 160) have also been examined elsewhere.

Generally, these priority research questions demonstrated that it was important for preference studies to use robust methods, to be able to assess preference heterogeneity, and to consider ways to minimise patient burden and maximise patients' understanding of concepts presented in the preference study. Similar research agendas have previously been identified by the MDIC.(161, 162)

8.3   Areas of future research

PREFER was able to provide a foundation on when and how to run a patient preference study based on planned research activities and completion of the EMA qualification procedure. This included providing a framework and methodological considerations to aid in the design and execution of patient preference studies. In addition, PREFER case studies investigated the methodological questions that were considered higher priority based on literature reviews, stakeholder discussions, and ranking among stakeholders. There are, however, outstanding questions that require additional experience or examination beyond this six-year project, which are presented hereafter.

8.3.1 Preference framework (Section 3)

It is recommended that stakeholders use the proposed PREFER framework when conducting patient preference studies so they are designed, conducted, and communicated in a consistent and uniform way for use in decision-making. Future research should explore how stakeholders use the framework across the medical product life cycle and if any adjustments are required.

More research is needed on how best to incorporate results from patient preference studies into regulatory documents (e.g. clinical overviews, European public assessment reports) and MAA dossiers, based on a more complete understanding about what information would be most useful for decision makers. These findings could then be used to refine the framework.

8.3.2 Involving patients and other stakeholders (Section 4)

An important theme highlighted during consultation with PREFER stakeholders (patients, sponsors, regulatory agencies, HTA bodies) was the importance of early involvement of stakeholders in study design to increase the relevance and quality of studies for decision-making. Appropriate training materials for patient partners (e.g. glossaries, introduction to research/analysis techniques, assertiveness skills) and researchers (communication skills, needs awareness, community outreach) to facilitate effective involvement are recommended where necessary. However, formal development and evaluation of such materials for different stakeholders was not undertaken and could be the focus of further investigation.

During the PREFER cases studies, patient research partners made valuable contributions to the development of educational resources and materials for disseminating study findings to participants and the general public. Further work could build on this by applying patient-centred approaches in the development of innovative digital platforms to better communicate this information.

Currently, there are checklists for reporting qualitative and quantitative preference studies to promote quality and transparency.(85, 163) Future research could investigate the potential value of extending reporting standards of patient preference studies to describe patient and other stakeholder involvement more thoroughly; for example, by adapting aspects of the GRIPP2 checklist for reporting patient and public involvement in research (164) and the IMI PARADIGM Guidance for Reporting and Dissemination of Patient Engagement Activities.(165) This could include the extent and impact of stakeholder involvement in:

- study objectives

- identification of relevant populations

- relevant study outcomes

- recruitment procedures

- information materials for study participants

- qualitative discussion guides and data interpretation

- attribute selection and presentation

- choice of demographic, clinical, and psychological measures

- survey design, content, pre-testing, and piloting

- dissemination of study findings.

8.3.3 Methodological considerations (Section 5)

8.3.3.1 Comparison of methods

Seven case studies compared two preference elicitation methods (all studies used a DCE and compared this to either PTT, BWS case 1, BWS case 2, or swing weighting), and the method selected had some impact on the results. Methods were generally well aligned for the one or two attributes that were ranked as most important; however, generally the relative importance weights or MAR, significantly differed across methods. Importantly, the preference methods themselves elicit preferences differently (e.g. ranking, sorting, weighting, one-at-a-time vs. all attributes together) or require different presentation or framing formats that can result in quantitative and/or relative rank differences.

Most case study participants reported that preference methods were relatively easy to understand and complete. Although DCEs have historically been used more frequently than other methods, results from PREFER's stakeholder survey demonstrated that PTT, BWS and swing weighting approaches tend to comply with methods criteria considered important by stakeholders across the medical product life cycle (**Section 5**). Therefore, additional research into the use of PTT, BWS and swing weighting, as well as other preference elicitation methods, should be considered to gain experience with these methods, to test their feasibility, and to develop evidence-based guidance documents.

Further, additional research should compare different preference elicitation methods to build upon the comparisons already conducted within PREFER. There is no gold standard method for eliciting patient preferences, and it is important to understand the differences between methods in terms of how they are conducted, what they measure, the burden they impose on respondents and how study outcomes compare to one another. Such studies will help researchers and other medical product life cycle stakeholders make an informed choice about which method selection is most appropriate for a given preference study. Future studies could also determine if, and to what extent, internal validity and data quality can/should be incorporated in preference studies, as well as explore measurements of external validity for different preference methods.

8.3.3.2 Qualitative and quantitative transferability

The transferability of patient preference study results to other contexts (e.g. other countries, disease areas or treatment/indications) has not been widely studied. PREFER research has resulted in a transferability checklist (see **Box A8-1** in **Annex Section A8**) to help researchers evaluate the transferability of study results. Future studies should investigate situations where transferability may be appropriate, decision contexts where the transfer of previous results is useful, and elements of the preference study affecting transferability.

In addition to considering transferability, the utility and necessity of conducting a new study vs. applying a quantitative meta-analysis approach with previously published preference studies could be further assessed. A meta-analysis of preference studies for psoriasis treatment confirmed that previously published preference information could be used to characterise patient risk tolerance,(166) and there may be other disease areas and decision contexts where a meta-analytic approach could replace the need for new research. However, understanding how existing preference evidence obtained in a certain context can be leveraged to answer different questions is challenging. Although the PREFER transferability checklist can be used to help determine if existing evidence can be used, knowing how to numerically extrapolate this evidence to a different context needs further development. Further research should assess if and what statistical methods are available to conduct transparent benefit transfers, building on evidence from environmental economics.

8.3.4 Measuring relevant psychological constructs and individual differences (Section 6)

All PREFER case studies considered the inclusion of psychological constructs (defined as a "measurable definition of a psychological aspect") and several did include them. **Section 6** makes recommendations on which measures of psychological constructs should be considered for inclusion in patient preference studies based on the available evidence literature (4) and expert consensus.(23) Consensus was reached for the inclusion of eight class I constructs (health literacy, health numeracy, illness perception, treatment-related beliefs, risk propensity, health locus of control, control preference, and patient activation). Six of these have been included in PREFER case studies (health literacy and health numeracy in eight case studies each, patient activation and health locus of control in two each, illness perceptions and treatment related beliefs in one each) to assess the relationships between these constructs and treatment preferences. None of the case studies assessed relationships between treatment preferences and measures of risk propensity and control preferences (167) (see **Section 6.3.1** class I constructs) – an area that warrants further studies.

Three constructs (autonomy preference, decision-making style, and health orientation) reached the majority agreement (class II) but were not assessed in PREFER case studies. Other constructs were not included in the consensus-based recommendations due to a lack of scientific evidence to support their inclusion at the time. Further research is therefore

needed to assess relationships between patient treatment preferences and measures of autonomy preference, decision-making style, health orientation and other constructs for which there is a theoretical rationale for a role in decision-making and/or health behaviours.

Findings from the PREFER case studies reinforce the importance of considering the role of health literacy and numeracy in patient preference studies, and in understanding preference heterogeneity. These constructs were well studied in PREFER and explained preferences across a range of diseases, treatment settings, and populations. Relationships between other constructs (e.g. illness perceptions) and preferences were also observed, but further investigation is needed to assess the extent to which these findings apply across samples and disease settings. This progress highlights a need for expansion of this nascent evidence base and subsequent development and adoption of a data-driven consensus framework by the wider research community to define which psychological constructs are relevant in which contexts.

The PREFER recommendations also highlight a range of pragmatic considerations (e.g. length of survey and burden for participants) that might limit the applicability of including psychological constructs. It is also possible that the act of responding to some measures of psychological constructs could impact on patient preferences. To a limited extent this can be managed by including psychological measures after the choice task items in the survey instrument. Pragmatic considerations such as these were managed in the PREFER case studies by the inclusion of stakeholders (particularly patient research partners) in survey development and pre-testing. However, the most appropriate placement of psychological tools within a preference survey and the minimum/maximum measures to include are yet to be established. Systematic investigation of the effect of length and complexity of measures of psychological constructs incorporated in preference studies on survey acceptability and/or the preferences elicited was not undertaken in PREFER, and is another useful avenue of further investigation.

8.3.5   Educational materials (Section 7)

Traditionally, education about the disease, treatment context, and attributes and levels in preference studies has taken the form of text-based materials. Although there has been some exploration of the use of 'enhanced' materials to achieve educational objectives (143, 157), this is a nascent area of development.

The PREFER case studies explored a range of theory-based and evidence-based educational approaches (see **Section 7** and the use of the EDUGrid), including short videos about the preference methods and the attributes of the disease and its treatment (**Section 7**). Results from these case studies indicated that participants were generally satisfied with enhanced educational materials, such as animations with voiceovers, to prepare them for study participation, which replicates previous findings from the literature.(168-170) However, no systematic effect on preference outcomes or choice consistency was found, either in

PREFER or in the literature, in split sample studies comparing groups of respondents who received enhanced educational material with those who received text-only information. Reflecting on these uncertain findings about the added value of educational materials in preference studies, future research is needed to increase insight into the impact of different types of educational materials on preferences and other study outcomes.

Future research should examine ways to tailor information so that participants are presented with educational material that is specifically relevant to them; for example, by considering their health literacy, existing knowledge of medical condition(s) and treatments, and psycho-emotional state. Psychometric constructs such as health literacy and numeracy should be considered when developing educational materials because low levels of these constructs among respondents are associated with greater difficulties in understanding health information displayed in words or numbers.(171) Future research should look into additional characteristics of patients (as well as comparing respondent groups, e.g. high vs. low literacy) and assess global understanding by using comprehension questions. These questions can help to identify parts of educational materials where a respondent may benefit from additional, tailored information. In PREFER, case studies involving more novel diseases and treatments, or less experienced respondents, found more favourable responses towards educational material than those with respondents who have more disease experience. Future studies should investigate and assess respondents' awareness/experience and psycho-emotional sensitivity with the choice area and task presented, as both these aspects could further help to inform the tailoring of the educational materials.

In relation to the design of educational materials, the PREFER case studies sought to include multimedia features (combining textual information, images, and videos) that would make choice tasks stimulating and effective, and these were tested to ensure that they were easy to understand. PREFER identified that long and/or complex educational materials should be avoided as they can increase the risk of attrition or loss of proper engagement, especially where preference elicitation tasks are already demanding for participants. Thus, future studies could focus on identifying the right balance between educational tools and features (such as realism, simulation, interactivity, immersion, and narration), and educational burden. An option would be to compare different presentation formats (e.g. graphical displays, video, descriptions) and identify the most suitable for specific patient characteristics.

8.3.6   Future research not in the initial scope of PREFER

Areas of future research that were out of scope of PREFER and which could be investigated include:

- individual preferences and shared decision-making within the healthcare setting

- creating attributes libraries within disease areas

- identifying methods for the mapping of clinical outcome assessments (COAs, e.g. PROs) to patient preference study attributes and levels

- examining revealed preferences in the post-marketing setting

- how best to characterise the uncertainty of treatment benefits and risks in patient preference studies.

8.3.6.1   Individual preferences and shared decision-making

Shared decision-making is the process by which a healthcare provider works with a patient to reach a decision about their care, based on the patient's clinical condition, patient-specific probabilities of endpoints, and wishes of the patient.(172, 173) It is similar to a patient preference study in that it requires the patient to understand the context and choices available to them. Some preference studies have been applied in shared decision-making or value clarification, but there is limited public use.(174, 175) Individual preferences can be used in clinical practice (176) but future research could explore the use of patient preference studies in the development of decision aids and individual patient preferences for shared decision-making.

8.3.6.2   Attribute reference library

The selection of outcomes that will compose the choice tasks is a crucial phase in the design of a patient preference study. Attributes and levels are mostly identified through a combination of literature reviews, qualitative research, expert opinion, and an analysis of clinical data.(85, 163) While every study may be unique in terms of its context and target sample, a level of standardisation (including standardised attributes) can offer advantages such as comparability across studies and a chance for generalisation. Attributes found in many patient preference studies, such as overall survival, disabling stroke, myocardial infarction, and fatigue, are currently defined anew in most studies, as are attributes relating to common adverse events that could appear in a wide range of treatment profiles. The use of common attributes (i.e. wording, definition, and/or representation) could save time, improve quality, and provide consistency. However, care should always be taken that the attributes are relevant for the specific context/country and consistent with how they might be explained in the clinical setting. Future research could be undertaken to evaluate whether a common library of attributes would be useful, clarifying in which context this would be of value, and what type of attributes should take priority. The design of this library should be based on a consensus approach including patient preference experts, clinical experts, patients, and other stakeholders (industry and regulators). This would ensure that relevant attributes to all parties are considered but also will result in evidence-based decisions that guarantee the selection of appropriate attributes and levels. Empirical investigation of the effect of variations in attribute framing, presentation, and definition may be considered in the consensus-seeking process.

In clinical trials, there are some disease areas that have an agreed set of clinical outcome measures collected in each study known as a core outcome set.(177-179) Similarly, the International Consortium for Health Outcomes Measurement (180) has developed standard sets of health outcome measures that matter most to patients. Patient preference researchers could learn from the selection processes of these initiatives to identify, classify, and disseminate core outcome set attributes and levels. Such harmonisation of attributes and levels could help to reduce across-studies heterogeneity and enhance evidence-based synthesis of patient preferences in different diseases. Alternatively, the focus could be on the framing of specific symptoms regardless of the disease. For example, fatigue, which is the most frequent symptom reported by patients with chronic illnesses and is commonly included in patient preference studies, has over 20 different PROs and the framing or wording of the same symptom can vary considerably.(181) Fatigue was also an attribute common across PREFER case studies but described in different ways. Identifying and prioritising attributes of relevance for such harmonisation could be a task for further development.

In this context, it is recommended that all prospective patient preference studies are registered publicly; for example, in the Health Preference Study and Technology Registry, which is maintained by the International Academy for Health Preferences Research.(182) If all past and present studies are easily retrievable, companies can assess the transferability of existing studies before deciding whether to initiate a new study.

8.3.6.3   Relating preferences to PROs and other COAs

Patient-reported and clinician-reported COAs are measures that reflect how a patient feels or functions. Patient-reported outcomes and COAs are distinct from patient preference assessments, but they are both patient-centred measures, which can inform one another across the medical product life cycle. For example, patient preferences may inform which PROs or other endpoints to include in a clinical study, whereas patient-reported or clinician-reported COA items (constructs) may be relevant to include as attributes or measures of clinical effectiveness in preference studies.

Clinical outcome assessments can be used to inform items in patient preference studies, and patient preference studies can be used to inform the scoring algorithm in COAs. Both are relatively new topics that need additional research, but the work to date suggests these are critically important and productive areas for research. Below are some several examples that provide direction for the field.

Clinical outcome assessments typically have three to ten items, each measured with Likert or other scales with multiple response levels. Clinical outcome assessments that are not completed by patients (e.g. clinician-reported outcomes) may have items that are not assessable by patients. Additionally, the items in COAs are often interdependent, while attributes in a patient preference study must be independent. The number of items, their potential complexity, and their possible interdependence makes using COA items directly as

attributes within a patient preference study challenging. However, there are a few studies that have mapped COAs to attributes; for example, improved mood descriptions in treatment-resistant depression from a depression COA,(183) and changes in positive and negative symptoms in schizophrenia from a clinician-reported outcome.(184, 185) These studies used factor analyses to identify which COA items had the most influence on the COA score, as well as correlation between the items, to determine which items to use when defining attributes. Other important examples are Hauber *et al* (186) and Arden *et al* (187) for the Western Ontario and McMaster Universities Osteoarthritis Index.(187) This topic is also of increasing importance to the FDA and has been raised at FDA Advisory Committee meetings.(188, 189)

The scoring algorithm in a COA typically assigns scores to the response levels for each item and then gives a cumulative score by summing or weighing the scores for each item. Preferences can be used to assess the scores for each level and for the weights used in the cumulative score. The earliest examples are by Osoba *et al* (190) and Johnson *et al* (191) for the European Organisation for Research and Treatment of Cancer's Quality of Life questionnaire. Mohamed *et al* summarised a number of previous studies that used similar approaches to other PRO instruments.(192)

8.3.6.4 Revealed preferences

Most patient preference studies in the medical product life cycle use stated preferences where participants are responding to hypothetical situations. Future research could examine revealed preferences – actual treatment choices being made following product approval. Such studies might help explain why certain products are or are not used, how they might be improved, and how patients manage long-term side-effects. Furthermore, combining research on stated and revealed preferences further allows for the investigation of external validity.(193-197)

8.3.6.5 Uncertainty of treatment benefits and harms

The most effective ways of including the degree of uncertainty around benefit and risk estimates in patient preference studies has not been determined and may not even be appropriate in all studies. Attributes and levels can be challenging to explain clearly to participants, and introducing uncertainty as a concept adds additional complexity to the study design.(198) Some examples of efforts to incorporate uncertainty in rheumatoid arthritis include using level of evidence (GRADE) language from treatment guidelines(199, 200) a graphical range and qualitative descriptions of confidence in the evidence,(201) and the length of time that a treatment has been in use.(202, 203) In a whole genome screening patient preference study, conveying uncertainty was considered too complex: the multiple layers of probabilities and associated uncertainties required considerable simplification to ensure experimental control over multiple factors intrinsic to the design.(198) More research

is needed to understand when to incorporate uncertainty and the most effective way for patients to assess the relative importance of uncertainty around benefits and risks.

8.3.6.6 Missing data

The topic of missing data in research is important. Data can be missing (completely) at random or missing for reasons relating to a clinical condition, patient-reported measures, or patient preferences. To our knowledge, and in contrast to clinical and epidemiological research studies, missing data and standardised methods of handling and assessing missing data in the context of patient preference studies have not been defined. However, this should be further investigated as the use of patient preference studies for decision-making across the medical product life cycle matures.

8.4 The future

Patients, regulators, HTA bodies, academic researchers, and industry stakeholders have supported the application of patient preference studies to decision-making in the medical product life cycle. However, as with any emerging area of research, this enthusiasm is tempered by questions that require additional evaluation. While there are practical limitations regarding which questions can be addressed within a fixed budget and timeframe, PREFER assessed high priority research questions across more than 10 case studies and laid the groundwork for the use of preferences in the medical product life cycle via the framework and points to consider when choosing a method. Beyond PREFER, the body of methodological research and evidence about the application of patient preference studies grows with each additional study.

Patient preference studies have enormous promise despite the existence of open questions. PREFER encourages the patient preference community to continue identifying requirements to ensure studies are sufficiently robust to inform decisions about medical products. In addition, PREFER invites more guidance from scientific societies, regulators, and HTA bodies on the use of patient preferences.

Applying PREFER's recommendations in a systematic and sustainable way, addressing open research questions in future preference studies, and gaining more experience requires continued effort after the end of the PREFER project. The PFMD (204) has been instrumental in the sustainability activities of other IMI projects (e.g. EUPATI, PARADIGM) and, with its Patient Engagement and Patient Experience Data projects, has a shared purpose with PREFER of enhancing the field and establishing patient preference studies as an accepted tool to generate evidence that supports medical-product decision making. A new patient preference workstream under PFMD will enable a continued collaboration of PREFER consortium members and new partners, as well as interactions with patients, regulators, HTA bodies, and scientific groups.

9 References

1. Janssens R, Russo S, van Overbeeke E, Whichello C, Harding S, Kübler J, *et al.* Patient Preferences in the Medical Product Life Cycle: What do Stakeholders Think? Semi-Structured Qualitative Interviews in Europe and the USA. Patient. 2019;12(5):513–26.

2. Whichello C, van Overbeeke E, Janssens R, Schölin Bywall K, Russo S, Veldwijk J, *et al.* Factors and Situations Affecting the Value of Patient Preference Studies: Semi-Structured Interviews in Europe and the US. Frontiers in Pharmacology. 2019;10(1009).

3. van Overbeeke E, Whichello C, Janssens R, Veldwijk J, Cleemput I, Simoens S, *et al.* Factors and situations influencing the value of patient preference studies along the medical product lifecycle: a literature review. Drug Discov Today. 2019;24(1):57–68.

4. Russo S, Jongerius C, Faccio F, Pizzoli SFM, Pinto CA, Veldwijk J, *et al.* Understanding Patients' Preferences: A Systematic Review of Psychological Instruments Used in Patients' Preference and Decision Studies. Value Health. 2019;22(4):491–501.

5. Soekhai V, Whichello C, Levitan B, Veldwijk J, Pinto CA, Donkers B, *et al.* Methods for exploring and eliciting patient preferences in the medical product lifecycle: a literature review. Drug Discovery Today. 2019;24(7):1324–31.

6. van Overbeeke E, Janssens R, Whichello C, Schölin Bywall K, Sharpe J, Nikolenko N, *et al.* Design, Conduct, and Use of Patient Preference Studies in the Medical Product Life Cycle: A Multi-Method Study. Frontiers in Pharmacology. 2019;10(1395).

7. van Overbeeke E, Vanbinst I, Jimenez-Moreno AC, Huys I. Patient Centricity in Patient Preference Studies: The Patient Perspective. Frontiers in Medicine. 2020;7.

8. Whichello C, Levitan B, Juhaeri J, Patadia V, DiSantostefano R, Pinto CA, *et al.* Appraising patient preference methods for decision-making in the medical product lifecycle: an empirical comparison. BMC Medical Informatics and Decision Making. 2020;20(1):114.

9. van Overbeeke E, Forrester V, Simoens S, Huys I. Use of Patient Preferences in Health Technology Assessment: Perspectives of Canadian, Belgian and German HTA Representatives. Patient. 2021;14(1):119–28.

10. van Overbeeke E, Michelsen S, Hauber B, Peerlinck K, Hermans C, Lambert C, *et al.* Patient perspectives regarding gene therapy in haemophilia: Interviews from the PAVING study. Haemophilia. 2021;27(1):129–36.

11. Bywall KS, Kihlbom U, Hansson M, Falahee M, Raza K, Baecklund E, *et al.* Patient preferences on rheumatoid arthritis second-line treatment: a discrete choice experiment of Swedish patients. Arthritis Research & Therapy. 2020;22(1):288.

12. Jimenez-Moreno A, Pinto C, Levitan B, Whichello C, Dyer C, Van Overbeeke E, *et al.* A study protocol for quantifying patient preferences in neuromuscular disorders: a case study of the IMI PREFER Project. Wellcome Open Research. 2020;5(253).

13. Durosini I, Janssens R, Arnou R, Veldwijk J, Smith MY, Monzani D*, et al.* Patient Preferences for Lung Cancer Treatment: A Qualitative Study Protocol Among Advanced Lung Cancer Patients. Frontiers in Public Health. 2021;9.

14. Petrocchi S, Janssens R, Oliveri S, Arnou R, Durosini I, Guiddi P*, et al.* What Matters Most to Lung Cancer Patients? A Qualitative Study in Italy and Belgium to Investigate Patient Preferences. Frontiers in Pharmacology. 2021;12.

15. Jimenez-Moreno AC, van Overbeeke E, Pinto CA, Smith I, Sharpe J, Ormrod J*, et al.* Patient Preferences in Rare Diseases: A Qualitative Study in Neuromuscular Disorders to Inform a Quantitative Preference Study. The Patient - Patient-Centered Outcomes Research. 2021;14(5):601–12.

16. van Overbeeke E, Hauber B, Michelsen S, Goldman M, Simoens S, Huys I. Patient Preferences to Assess Value IN Gene Therapies: Protocol Development for the PAVING Study in Hemophilia. Frontiers in Medicine. 2021;8.

17. Bywall KS, Veldwijk J, Hansson MG, Baecklund E, Raza K, Falahee M*, et al.* Does being exposed to an educational tool influence patient preferences? The influence of an educational tool on patient preferences assessed by a discrete choice experiment. Patient Educ Couns. 2021;104(10):2577–85.

18. Falahee M, Simons G, DiSantostefano RL, Valor Méndez L, Radawski C, Englbrecht M*, et al.* Treatment preferences for preventive interventions for rheumatoid arthritis: protocol of a mixed methods case study for the Innovative Medicines Initiative PREFER project. BMJ Open. 2021;11(4):e045851.

19. Janssens R, Lang T, Vallejo A, Galinsky J, Plate A, Morgan K*, et al.* Patient Preferences for Multiple Myeloma Treatments: A Multinational Qualitative Study. Frontiers in Medicine. 2021;8.

20. Monzani D, Petrocchi S, Oliveri S, Veldwijk J, Janssens R, Bailo L*, et al.* Patient Preferences for Lung Cancer Treatments: A Study Protocol for a Preference Survey Using Discrete Choice Experiment and Swing Weighting. Frontiers in Medicine. 2021;8.

21. van Overbeeke E, Hauber B, Michelsen S, Peerlinck K, Lambert C, Hermans C*, et al.* Patient preferences for gene therapy in haemophilia: Results from the PAVING threshold technique survey. Haemophilia. 2021;27(6):957–66.

22. Janssens R, Arnou R, Schoefs E, Petrocchi S, Cincidda C, Ongaro G*, et al.* Key Determinants of Health-Related Quality of Life Among Advanced Lung Cancer Patients: A Qualitative Study in Belgium and Italy. Frontiers in Pharmacology. 2021;12.

23. Russo S, Monzani D, Pinto CA, Vergani L, Marton G, Falahee M*, et al.* Taking into Account Patient Preferences: A Consensus Study on the Assessment of Psychological Dimensions Within Patient Preference Studies. Patient preference and adherence. 2021;15:1331–45.

24. Smith IP, DiSantostefano RL, de Bekker-Grob EW, Levitan B, Berlin C, Veldwijk J*, et al.* Methodological Priorities for Patient Preferences Research: Stakeholder Input to the PREFER Public–Private Project. The Patient - Patient-Centered Outcomes Research. 2021;14(5):449–53.

25. Turk D, Boeri M, Abraham L, Atkinson J, Bushmakin AG, Cappelleri JC, *et al.* Patient preferences for osteoarthritis pain and chronic low back pain treatments in the United States: a discrete-choice experiment. Osteoarthritis Cartilage. 2020;28(9):1202–13.

26. Pinto CA, Chua GN, Bridges JFP, Brookes E, Hyacinthe J, Tervonen T. Comparing Patient Preferences for Antithrombotic Treatment During the Acute and Chronic Phases of Myocardial Infarction: A Discrete-Choice Experiment. The Patient - Patient-Centered Outcomes Research. 2022;15(2):255–66.

27. Walsh DA, Boeri M, Abraham L, Atkinson J, Bushmakin Andrew G, Cappelleri JC, *et al.* Exploring patient preference heterogeneity for pharmacological treatments for chronic pain: A latent class analysis. European Journal of Pain. 2022;26(3):648–67.

28. Cook NS, Kostikas K, Gruenberger J-B, Shah B, Pathak P, Kaur VP, *et al.* Patients' perspectives on COPD: findings from a social media listening study. ERJ Open Research. 2019;5(1):00128–2018.

29. Patalano F, Gutzwiller FS, Shah B, Kumari C, Cook NS. Gathering Structured Patient Insight to Drive the PRO Strategy in COPD: Patient-Centric Drug Development from Theory to Practice. Advances in Therapy. 2020;37(1):17–26.

30. Cook N, Gey J, Oezel B, Mackay AJ, Kumari C, Kaur VP, *et al.* Impact of cough and mucus on COPD patients: primary insights from an exploratory study with an Online Patient Community. Int J Chron Obstruct Pulmon Dis. 2019;14:1365-76.

31. Tervonen T, Prawitz T, Chua GN, Hyacinthe J, Pinto CA. Net clinical benefit of antiplatelet therapy was affected by patient preferences: A personalized benefit-risk assessment. J Clin Epidemiol. 2021;144:84–92.

32. European Medicines Agency. Qualification Opinion of IMI PREFER. 3 May 2022. **qualification-opinion-imi-prefer_en.pdf (europa.eu)**

33. U.S. Food & Drug Administration. Patient-Focused Drug Development Glossary. 6 August 2018. https://www.fda.gov/drugs/development-approval-process-drugs/patient-focused-drug-development-glossary. Accessed 22 April 2022.

34. 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. 2015. https://mdic.org/wp-content/uploads/2015/05/MDIC_PCBR_Framework_Proof5_Web.pdf.

35. U.S. Food & Drug Administration. Patient Preference-Sensitive Areas: Using Patients Preference Information in Medical Device Evaluation. 29 September 2020. https://www.fda.gov/about-fda/cdrh-patient-science-and-engagement-program/patient-preference-sensitive-areas-using-patient-preference-information-medical-device-evaluation.

36. U.S. Food & Drug Administration. Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. December 2009. https://www.fda.gov/media/77832/download.

37. U.S. Food & Drug Administration. Department of Health and Human Services. Center for Devices and Radiological Health & Center for Biologics Evaluation and Research. Patient Preference Information - Voluntary Submission, Review in Premarket Approval

Applications, Humanitarian Device Exemption Applications, and De Novo Requests, and Inclusion in Decision Summaries and Device Labeling. Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders. 24 August 2016. https://www.fda.gov/media/92593/download. Accessed 22 April 2022.

38. Nunnally JC, Bernstein IH. Psychometric theory. 1994. 3rd edition, McGraw-Hill (New York).

39. American Psychological Association. APA Dictionary of Psychology. https://dictionary.apa.org/instrument. Accessed 22 April 2022.

40. National Institute for Health and Care Excellence. Glossary. https://www.nice.org.uk/glossary?letter=q#:~:text=and%20cognitive%20impairment'.-,Quality-adjusted%20life%20year,of%20life%20in%20perfect%20health. Accessed 22 April 2022.

41. Kleijnen S, George E, Goulden S, d'Andon A, Vitré P, Osińska B, *et al.* Relative Effectiveness Assessment of Pharmaceuticals: Similarities and Differences in 29 Jurisdictions. Value in Health. 2012;15(6):954–60.

42. U.S. Food & Drug Administration. M4E(R2): The CTD — Efficacy Guidance for Industry. July 2017. https://www.fda.gov/media/93569/download.

43. University of Rochester. Participating in a Research Study. https://www.rochester.edu/ohsp/subject/participatingInResearch.html. Accessed 22 April 2022.

44. de Noordhout CM, Detollenaere J, Primus-de Jong C, Kohn L, Devleesschauwer B, Charafeddine R, *et al.* Identifying Patient needs: methodological approach and application. KCE report 348. 2021. https://kce.fgov.be/sites/default/files/atoms/files/KCE_348_Patient_Needs_Report.pdf.

45. Parsons S, Starling B, Mullan-Jensen C, Tham SG, Warner K, Wever K. 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.

46. 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.

47. Crocker JC, Boylan AM, Bostock J, Locock L. 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 Expect. 2017;20(3):519–28.

48. Christiaens W, Kohn L, Léonard C, Denis A, Daue F, Cleemput I. Models for citizen and patient involvement in health care policy - Part I: exploration of their feasibility and acceptability 2014.

49. Tai BB, Bae YH, Le QA. A Systematic Review of Health Economic Evaluation Studies Using the Patient's Perspective. Value Health. 2016;19(6):903–8.

50. Abelson J, Wagner F, DeJean D, Boesveld S, Gauvin FP, Bean S, *et al.* Public and patient involvement in health technology assessment: a framework for action. Int J Technol Assess Health Care. 2016;32(4):256–64.

51. European Medicines Agency. The patient's voice in the evaluation of medicines. How patients can contribute to assessment of benefit and risk. 18 October 2013.

https://www.ema.europa.eu/en/documents/report/report-workshop-patients-voice-evaluation-medicines_en.pdf.

52. European Medicines Agency. EMA Regulatory Science to 2025: Strategic Reflection. 2020. https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/ema-regulatory-science-2025-strategic-reflection_en.pdf.

53. U.S. Food & Drug Administration. PDUFA reauthorization performance goals and procedures fiscal years 2023 through 2027. https://www.fda.gov/media/151712/download.

54. Bouvy JC, Cowie L, Lovett R, Morrison D, Livingstone H, Crabb N. Use of Patient Preference Studies in HTA Decision Making: A NICE Perspective. Patient. 2020;13(2):145–9.

55. Ho MP, Gonzalez JM, Lerner HP, Neuland CY, Whang JM, McMurry-Heath M, *et al.* Incorporating patient-preference evidence into regulatory decision making. Surg Endosc. 2015;29(10):2984–93.

56. Marsh K, Caro JJ, Zaiser E, Heywood J, Hamed A. Patient-centered decision making: lessons from multi-criteria decision analysis for quantifying patient preferences. Int J Technol Assess Health Care. 2018;34(1):105–10.

57. Mühlbacher AC, Kaczynski A. Making Good Decisions in Healthcare with Multi-Criteria Decision Analysis: The Use, Current Research and Future Development of MCDA. Appl Health Econ Health Policy. 2016;14(1):29–40.

58. Marsh K, Caro JJ, Hamed A, Zaiser E. Amplifying Each Patient's Voice: A Systematic Review of Multi-criteria Decision Analyses Involving Patients. Appl Health Econ Health Policy. 2017;15(2):155–62.

59. Mol PG, Arnardottir AH, Straus SM, de Graeff PA, Haaijer-Ruskamp FM, Quik EH, *et al.* Understanding drug preferences, different perspectives. Br J Clin Pharmacol. 2015;79(6):978–87.

60. U.S. Food Drug Administration. Patient-Focused Drug Development: Collecting Comprehensive and Representative Input Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders. June 2020. https://www.fda.gov/media/139088/download.

61. U.S. Food & Drug Administration. Patient Engagement in the Design and Conduct of Medical Device Clinical Studies. Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders. 24 September 2019. https://www.fda.gov/media/130917/download.

62. U.S. Food & Drug Administration. Patient-Focused Drug Development: Methods to Identify What Is Important to Patients Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders. February 2022. https://www.fda.gov/media/131230/download.

63. Facey K, Hansen HP, Single ANV. Patient involvement in health technology assessment. 2017. Adis Singapore. .

64. Cleemput I, Dauvrin M, Kohn L, Mistiaen P, Christiaens W, Léonard C. Position of KCE on patient involvement in health care policy research. Method. Brussels: Belgian Health Care Knowledge Centre (KCE). 2019.

https://kce.fgov.be/sites/default/files/atoms/files/KCE_320C_Patient_involvement_health_care_policy_research_Synthesis_0_0.pdf.

65. Eastern Research Group Inc. Assessment of the Use of Patient Experience Data in Regulatory Decision-Making. Final Report. 18 June 2021. https://www.fda.gov/media/150405/download.

66. Rummel M, Kim TM, Plenteda C, Capochiani E, Mendoza M, Smith R *, 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.

67. U.S. Food & Drug Administration. Patient Preference Information (PPI) in Medical Device Decision-Making. 27 September 2020. https://www.fda.gov/about-fda/cdrh-patient-science-and-engagement-program/patient-preference-information-ppi-medical-device-decision-making. Accessed 22 April 2022.

68. Hummel MJ, Volz F, van Manen JG, Danner M, Dintsios CM, Ijzerman MJ *, 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.

69. Institute for Quality and Efficiency in Health Care (IQWiG). Choice-based Conjoint Analysis – pilot project to identify, weight, and prioritize multiple attributes in the indication "hepatitis C". 23 July 2014. Executive Summary of Working Paper No. GA10-03. https://www.iqwig.de/download/ga10-03_executive-summary-of-working-paper-11_conjoint-analysis.pdf.

70. Postmus D, Mavris M, Hillege HL, Salmonson T, Ryll B, Plate A *, 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.

71. Postmus D, Richard S, Bere N, van Valkenhoef G, Galinsky J, Low E *, et al.* Individual Trade-Offs Between Possible Benefits and Risks of Cancer Treatments: Results from a Stated Preference Study with Patients with Multiple Myeloma. Oncologist. 2018;23(1):44–51.

72. de Bekker-Grob EW, Berlin C, Levitan B, Raza K, Christoforidi K, Cleemput I *, et al.* Giving Patients' Preferences a Voice in Medical Treatment Life Cycle: The PREFER Public-Private Project. Patient. 2017;10(3):263–6.

73. Biotechnology Innovation Organization and Parent Project Muscular Dystrophy. Key Considerations in Developing & Integrating Patient Perspectives in Drug Development: Examination of the Duchenne Case Study. June 2016. https://archive.bio.org/sites/default/files/BIO_PPMD_whitepaper_web.pdf.

74. Janssens R, Huys I, van Overbeeke E, Whichello C, Harding S, Kübler J *, 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;19(1):189.

75. European Medicines Agency. Day 80 assessment report overview and D120 LOQ template_en Rev 10.19. 2019. https://www.ema.europa.eu/en/human-regulatory/marketing-authorisation/assessment-templates-guidance.

76. International Conference on harmonisation of technical requirements for registration of pharmaceuticals for human use. ICH harmonised guideline. Revision of M4E guideline on enhancing the format and structure of benefit–risk information in ICH efficacy – M4E(R2). 15 June 2016. https://database.ich.org/sites/default/files/M4E_R2__Guideline.pdf.

77. U.S. Food & Drug Administration. Benefit-Risk Assessment for New Drug and Biological Products Guidance for Industry DRAFT GUIDANCE. September 2021. https://www.fda.gov/media/152544/download.

78. Mott DJ, Najafzadeh M. 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.

79. Mühlbacher AC, Juhnke C, Beyer AR, Garner S. Patient-Focused Benefit-Risk Analysis to Inform Regulatory Decisions: The European Union Perspective. Value Health. 2016;19(6):734–40.

80. Danner M, Hummel JM, Volz F, van Manen JG, Wiegard B, Dintsios CM*, 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;27(4):369–75.

81. 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):266–72.

82. Craig BM, Lancsar E, Mühlbacher AC, Brown DS, Ostermann J. Health Preference Research: An Overview. Patient. 2017;10(4):507–10.

83. Mühlbacher AC. Patient-centric HTA: different strokes for different folks. Expert Review of Pharmacoeconomics & Outcomes Research. 2015;15(4):591–7.

84. Gutknecht M, Schaarschmidt ML, Herrlein O, Augustin M. A systematic review on methods used to evaluate patient preferences in psoriasis treatments. J Eur Acad Dermatol Venereol. 2016;30(9):1454–64.

85. Bridges JF, Hauber AB, Marshall D, Lloyd A, Prosser LA, Regier DA*, 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;14(4):403–13.

86. Mühlbacher A, Johnson FR. Choice Experiments to Quantify Preferences for Health and Healthcare: State of the Practice. Appl Health Econ Health Policy. 2016;14(3):253–66.

87. International Council for Harmonisation of technical requirements for pharmaceuticals for human use. ICH harmonised guideline. Addendum on estimands and sensitivity analysis in clinical trials. To the guideline on statistical principles for clinical trials. E9 (R1). 2019. https://www.gmp-compliance.org/files/guidemgr/E9-R1_Step4_Guideline_2019_1203.pdf.

88. European Medicines Agency. ICH E9 Statistical principles for clinical trials. Note for guidance on statistical principles for clinical trials. September 1998. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5_en.pdf.

89. Janssen EM, Marshall DA, Hauber AB, Bridges JFP. Improving the quality of discrete-choice experiments in health: how can we assess validity and reliability? Expert Rev Pharmacoecon Outcomes Res. 2017;17(6):531–42.

90. Johnson FR, Yang JC, Reed SD. The Internal Validity of Discrete Choice Experiment Data: A Testing Tool for Quantitative Assessments. Value Health. 2019;22(2):157–60.

91. Lacey A, Luff D. Qualitative Research Analysis. The NIHR RDS for the East Midlands / Yorkshire & the Humber, 2007 https://www.rds-yh.nihr.ac.uk/wp-content/uploads/2013/05/9_Qualitative_Data_Analysis_Revision_2009.pdf.

92. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A*, et al.* Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. Value Health. 2005;8(2):94–104.

93. U.S. Food & Drug Administration. Plain Language Principles. 11 January 2022. https://www.fda.gov/about-fda/plain-writing-its-law/plain-language-principles. Accessed 22 April 2022.

94. Jacobson KL, Parker RM. Health Literacy Principles: Guidance for Making Information Understandable, Useful, and Navigable. 22 December 2014. https://nam.edu/wp-content/uploads/2015/06/HealthLiteracyGuidance.pdf.

95. Sroka-Saidi K, Boggetti B, Schindler TM. Transferring regulation into practice: The challenges of the new layperson summary of clinical trial results. Medical Writing. 2015;24(1):24–7.

96. National Institute for Health Research. Plain English summaries. 5 April 2021. https://www.nihr.ac.uk/documents/plain-english-summaries/27363. Accessed 22 April 2022.

97. European Medicines Agency. ICH guideline M4E(R2) - Common technical document for the registration of pharmaceuticals for human use – Efficacy. 15 July 2016. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-m4e-r2-common-technical-document-registration-pharmaceuticals-human-use-efficacy-step-5_en.pdf.

98. Huls SPI, Whichello CL, van Exel J, Uyl-de Groot CA, de Bekker-Grob EW. What Is Next for Patient Preferences in Health Technology Assessment? A Systematic Review of the Challenges. Value Health. 2019;22(11):1318–28.

99. European Medicines Agency Committee for Human Medicinal Products. Guideline for good clinical practice E6(R2) Step 5. 1 December 2016. EMA/CHMP/ICH/135/1995. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-6-r2-guideline-good-clinical-practice-step-5_en.pdf.

100. Toledo-Chávarri A, Perestelo-Pérez L, Álvarez-Pérez Y, Abt-Sacks A, Santoro Domingo P, Villalón D*, et al.* Participación de los pacientes en la Evaluación de Tecnologías Sanitarias: manual metodológico. Ministerio de Sanidad, Servicios Sociales e Igualdad. Servicio de Evaluación del Servicio Canario de la Salud. Informes de Evaluación de Tecnologías Sanitarias. 2016. http://funcanis.es/wp-content/uploads/2015/09/SESCS_2016_ImplicacionPacientesETS.pdf.

101. Veatch RM. The Patient-Physician Relation. The Patient as Partner, Part 2. January 1991. Indiana University Press.

102. Warner K, See W, Haerry D, Klingmann I, Hunter A, May M. EUPATI Guidance for Patient Involvement in Medicines Research and Development; Guidance for Pharmaceutical Industry-Led Medicines R&D. Frontiers in Medicine. 2018;5(270).

103. IMI PARADIGM. Planning Patient Engagement. Recommendations on the required capabilities for patient engagement. 2020. https://imi-paradigm.eu/PEtoolbox/pe-capabilities.pdf.

104. IMI PARADIGM. Planning Patient Engement. Patient engagement in medicines development: Recommendations on how to find the right match for the right patient engagement activity. https://imi-paradigm.eu/PEtoolbox/identification-of-patient-representatives.pdf. 2020.

105. CORDIS EU research results. Towards Early diagnosis and biomarker validation in Arthritis Management. https://cordis.europa.eu/project/id/305549/reporting. Accessed 22 April 2022.

106. IMI. RTCure. Rheuma Tolerance for Cure. https://www.imi.europa.eu/projects-results/project-factsheets/rtcure. Accessed 22 April 2022.

107. Bayliss K, Raza K, Simons G, Falahee M, Hansson M, Starling B, *et al.* Perceptions of predictive testing for those at risk of developing a chronic inflammatory disease: a meta-synthesis of qualitative studies. Journal of Risk Research. 2018;21(2):167–89.

108. IMI PARADIGM. Code of Conduct for all stakeholders involved in patient engagement activities within medicines development. 2020. https://imi-paradigm.eu/PEtoolbox/code-of-conduct.pdf.

109. IMI PARADIGM. Enhancement EUPATI industry guidance: Suggested working practices. 2020. http://imi-paradigm.eu/PEtoolbox/EUPATI/PARADIGM-Suggested-Working-Practices.pdf.

110. Patient Focused Medicines Development. The PFMD Book of Good Practice. May 2018. https://patientfocusedmedicine.org/bogp/book-of-good-practices.pdf.

111. Marsh K, de Bekker-Grob E, Cook N, Collacott H, Danyliv A. How to integrate evidence from patient preference studies into health technology assessment: a critical review and recommendations. International Journal of Technology Assessment in Health Care. 2021;37(1):e75.

112. Hauber B. Patient preference studies for regulatory decision making. Presentation at the Patient Focused Medicines Development Patient Engagement Open Forum. 23 June 2021. https://patientengagement.synapseconnect.org/resources/peof-2021-understandings-learnings-guidances-what-do-we-know-so-far-on-the-use-of-patient-experience-data-ped-for-healthcare-decision-making-slides-and-recording Accessed 12 April 2022.

113. Patient Focused Medicines Development. Highlighting Recent Trends in the Fast-Evolving Patient Engagement & Patient Experience Data Landscape. Autumn 2021. https://patientfocusedmedicine.org/docs/Trends-in-the-Fast-Evolving-PE-PED-Landscape.pdf.

114. Ryan M, Scott DA, Reeves C, Bate A, van Teijlingen ER, Russell EM, *et al.* Eliciting public preferences for healthcare: a systematic review of techniques. Health Technol Assess. 2001;5(5):1–186.

115. Soekhai V, de Bekker-Grob EW, Ellis AR, Vass CM. Discrete Choice Experiments in Health Economics: Past, Present and Future. Pharmacoeconomics. 2019;37(2):201–26.

116. van Overbeeke E, Hauber B, Michelsen S, Goldman M, Simoens S, Huys I. Study protocol article. Patient Preferences to Assess Value IN Gene Therapies: Protocol Development for the PAVING Study in Hemophilia. Frontiers in Medicine. 2021;8:595797.

117. Veldwijk J, Johansson JV, Donkers B, de Bekker-Grob EW. Mimicking Real-Life Decision Making in Health: Allowing Respondents Time to Think in a Discrete Choice Experiment. Value Health. 2020;23(7):945–52.

118. Jonker M, de Bekker-Grob E, Veldwijk J, Goossens L, Bour S, Rutten-Van Mölken M. COVID-19 Contact Tracing Apps: Predicted Uptake in the Netherlands Based on a Discrete Choice Experiment. JMIR Mhealth Uhealth. 2020;8(10):e20741.

119. Visser LA, Huls SPI, Uyl-de Groot CA, de Bekker-Grob EW, Redekop WK. An implantable device to treat multiple sclerosis: A discrete choice experiment on patient preferences in three European countries. J Neurol Sci. 2021;428:117587.

120. de Bekker-Grob EW, Donkers B, Veldwijk J, Jonker MF, Buis S, Huisman J, *et al.* What Factors Influence Non-Participation Most in Colorectal Cancer Screening? A Discrete Choice Experiment. The Patient - Patient-Centered Outcomes Research. 2021;14(2):269–81.

121. Jonker MF, Donkers B, de Bekker-Grob E, Stolk EA. Attribute level overlap (and color coding) can reduce task complexity, improve choice consistency, and decrease the dropout rate in discrete choice experiments. Health Economics. 2019;28(3):350–63.

122. Jonker MF, Donkers B, Goossens LMA, Hoefman RJ, Jabbarian LJ, de Bekker-Grob EW, *et al.* Summarizing Patient Preferences for the Competitive Landscape of Multiple Sclerosis Treatment Options. Medical Decision Making. 2020;40(2):198–211.

123. Meertens RM, Lion R. Measuring an Individual's Tendency to Take Risks: The Risk Propensity Scale1. Journal of Applied Social Psychology. 2008;38(6):1506–20.

124. Appelt K, Milch K, Handgraaf M, Weber E. The Decision Making Individual Differences Inventory and Guidelines for the Study of Individual Differences in Judgment and Decision-Making Research. Judgment and Decision Making. 2011;6(3):252–62.

125. Veldwijk J, van der Heide I, Rademakers J, Schuit AJ, de Wit GA, Uiters E, *et al.* Preferences for Vaccination: Does Health Literacy Make a Difference? Medical Decision Making. 2015;35(8):948–58.

126. Noblin AM, Wan TTH, Fottler M. The impact of health literacy on a patient's decision to adopt a personal health record. Perspect Health Inf Manag. 2012;9(Fall):1–13.

127. Wallston KA, Strudler Wallston B, DeVellis R. Development of the Multidimensional Health Locus of Control (MHLC) Scales. Health Education Monographs. 1978;6(1):160–70.

128. Ono R, Higashi T, Suzukamo Y, Konno S, Takahashi O, Tokuda Y, *et al.* Higher internality of health locus of control is associated with the use of complementary and alternative medicine providers among patients seeking care for acute low-back pain. The Clinical journal of pain. 2008;24(8):725–30.

129. U.S. Food & Drug Administration. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. 2009. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-reported-outcome-measures-use-medical-product-development-support-labeling-claims. FDA-2006-D-0362.

130. Carmines EG, Woods J. Validity. In: Lewis-Beck M, A. Bryman, Liao TF, editors. The Sage encyclopedia of social science research methods: SAGE Publishing; 2003.

131. Gushta MM, Rupp AA. Reliability. In: Salkind NJ, editor. Encyclopedia of Research Design2010.

132. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976). 2000;25(24):3186–91.

133. Hays RD, Weech-Maldonado R, Teresi JA, Wallace SP, Stewart AL. Commentary: Copyright Restrictions Versus Open Access to Survey Instruments. Med Care. 2018;56(2):107–10.

134. Lingler JH, Schmidt KL, Gentry AL, Hu L, Terhorst LA. A New Measure of Research Participant Burden: Brief Report. J Empir Res Hum Res Ethics. 2014;9(4):46–9.

135. Sharp LM, Frankel J. Respondent burden: A test of some common assumptions. Public Opinion Quarterly. 1983;47(1):36–53.

136. Ulrich CM, Wallen GR, Feister A, Grady C. Respondent burden in clinical research: when are we asking too much of subjects? Irb. 2005;27(4):17–20.

137. Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. Fam Med. 2004;36(8):588–94.

138. Eichler HG, Abadie E, Baker M, Rasi G. Fifty years after thalidomide; what role for drug regulators? Br J Clin Pharmacol. 2012;74(5):731–3.

139. Institute for Quality and Efficiency in Health Care (IQWiG). Background: What are patient preferences? https://www.iqwig.de/en/about-us/methods/background-what-are-patient-preferences.html. Accessed 22 April 2022.

140. Johnson FR, Lancsar E, Marshall D, Kilambi V, Mühlbacher A, Regier DA, *et al.* Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. Value Health. 2013;16(1):3–13.

141. Louviere JJ. What You Don't Know Might Hurt You: Some Unresolved Issues in the Design and Analysis of Discrete Choice Experiments. Environmental and Resource Economics. 2006;34(1):173–88.

142. Fischhoff B, Kadvany J. Risk: a very short introduction. Oxford: Oxford University Press; 2011.

143. Vass CM, Davison NJ, Vander Stichele G, Payne K. A Picture is Worth a Thousand Words: The Role of Survey Training Materials in Stated-Preference Studies. Patient. 2020;13(2):163–73.

144. Buffel C, van Aalst J, Bangels AM, Toelen J, Allegaert K, Verschueren S, *et al.* A Web-Based Serious Game for Health to Reduce Perioperative Anxiety and Pain in Children

(CliniPup): Pilot Randomized Controlled Trial. JMIR Serious Games. 2019;7(2):e12431.

145. Matthyssens LE, Vanhulle A, Seldenslach L, Vander Stichele G, Coppens M, Van Hoecke E. A pilot study of the effectiveness of a serious game CliniPup® on perioperative anxiety and pain in children. J Pediatr Surg. 2020;55(2):304–11.

146. Verschueren S, Buffel C, Vander Stichele G. Developing Theory-Driven, Evidence-Based Serious Games for Health: Framework Based on Research Community Insights. JMIR Serious Games. 2019;7(2):e11565.

147. IMI PREFER. D2.5 Identification of educational and psychological feature methods. 27 February 2017. https://doi.org/10.5281/zenodo.5750154.

148. Fischhoff B, Brewer NT, Downs JS. Communicating Risks and Benefits: An Evidence-Based User's Guide. 2018. https://www.fda.gov/media/81597/download.

149. Harrison M, Rigby D, Vass C, Flynn T, Louviere J, Payne K. Risk as an attribute in discrete choice experiments: a systematic review of the literature. Patient. 2014;7(2):151–70.

150. Soekhai V, Donkers B, Levitan B, de Bekker-Grob EW. Case 2 best-worst scaling: For good or for bad but not for both. Journal of Choice Modelling. 2021;41:100325.

151. Vanniyasingam T, Cunningham CE, Foster G, Thabane L. Simulation study to determine the impact of different design features on design efficiency in discrete choice experiments. BMJ Open. 2016;6(7):e011985.

152. Goossens LMA, Jonker MF, Rutten-van Mölken M, Boland MRS, Slok AHM, Salomé PL, et al. The Fold-in, Fold-out Design for DCE Choice Tasks: Application to Burden of Disease. Medical decision making : an international journal of the Society for Medical Decision Making. 2019;39(4):450–60.

153. Lempert KM, Phelps EA. The Malleability of Intertemporal Choice. Trends in cognitive sciences. 2016;20(1):64–74.

154. Veldwijk J, Essers BA, Lambooij MS, Dirksen CD, Smit HA, de Wit GA. Survival or Mortality: Does Risk Attribute Framing Influence Decision-Making Behavior in a Discrete Choice Experiment? Value Health. 2016;19(2):202–9.

155. Abhyankar P, Summers BA, Velikova G, Bekker HL. Framing Options as Choice or Opportunity: Does the Frame Influence Decisions? Medical decision making : an international journal of the Society for Medical Decision Making. 2014;34(5):567–82.

156. Howard K, Salkeld G. Does attribute framing in discrete choice experiments influence willingness to pay? Results from a discrete choice experiment in screening for colorectal cancer. Value Health. 2009;12(2):354–63.

157. Lim SL, Yang JC, Ehrisman J, Havrilesky LJ, Reed SD. Are Videos or Text Better for Describing Attributes in Stated-Preference Surveys? Patient. 2020;13(4):401–8.

158. Veldwijk J, Lambooij MS, van Til JA, Groothuis-Oudshoorn CG, Smit HA, de Wit GA. Words or graphics to present a Discrete Choice Experiment: Does it matter? Patient Educ Couns. 2015;98(11):1376–84.

159. Veldwijk J, Groothuis-Oudshoorn CGM, Kihlbom U, Langenskiöld S, Dekker E, Kallenberg FGJ, et al. How psychological distance of a study sample in discrete choice

experiments affects preference measurement: a colorectal cancer screening case study. Patient preference and adherence. 2019;13:273–82.

160. Auriemma CL, Nguyen CA, Bronheim R, Kent S, Nadiger S, Pardo D*, et al.* Stability of end-of-life preferences: a systematic review of the evidence. JAMA internal medicine. 2014;174(7):1085–92.

161. Levitan B, Hauber AB, Damiano MG, Jaffe R, Christopher S. 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):531–6.

162. Meara A, Crossnohere NL, Bridges JFP. Methods for measuring patient preferences: an update and future directions. Current opinion in rheumatology. 2019;31(2):125–31.

163. Hollin IL, Craig BM, Coast J, Beusterien K, Vass C, DiSantostefano R*, et al.* Reporting Formative Qualitative Research to Support the Development of Quantitative Preference Study Protocols and Corresponding Survey Instruments: Guidelines for Authors and Reviewers. Patient. 2020;13(1):121–36.

164. Staniszewska S, Brett J, Simera I, Seers K, Mockford C, Goodlad S*, et al.* GRIPP2 reporting checklists: tools to improve reporting of patient and public involvement in research. BMJ. 2017;358:j3453.

165. IMI PARADIGM. Guidance for Reporting and Dissemination of Patient Engagement Activities. 2020. https://imi-paradigm.eu/PEtoolbox/reporting-and-dissemination.pdf.

166. Gonzalez JM, Johnson FR, McAteer H, Posner J, Mughal F. Comparing preferences for outcomes of psoriasis treatments among patients and dermatologists in the U.K: results from a discrete-choice experiment. Br J Dermatol. 2017;176(3):777–85.

167. Degner LF, Sloan JA, Venkatesh P. The Control Preferences Scale. Can J Nurs Res. 1997;29(3):21–43.

168. Silk KJ, Horodynski MA, Rienzo M, Mercer L, Olson B, Aldrich R. Strategies to increase health literacy in the infant feeding series (TIFS): a six-lesson curriculum for low-income mothers. Health promotion practice. 2010;11(2):226–34.

169. Keselman A, Logan R, Smith CA, Leroy G, Zeng-Treitler Q. Developing informatics tools and strategies for consumer-centered health communication. Journal of the American Medical Informatics Association : JAMIA. 2008;15(4):473–83.

170. Jacobs RJ, Lou JQ, Ownby RL, Caballero J. A systematic review of eHealth interventions to improve health literacy. Health informatics journal. 2016;22(2):81–98.

171. Durand MA, Yen RW, O'Malley J, Elwyn G, Mancini J. Graph literacy matters: Examining the association between graph literacy, health literacy, and numeracy in a Medicaid eligible population. PloS one. 2020;15(11):e0241844.

172. National Institute for Health and Care Excellence. Shared Decision-Making. 14 December 2021. https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-guidelines/shared-decision-making. Accessed 22 April 2022.

173. Agency for Healthcare Research and Quality (AHRQ). The SHARE Approach. https://www.ahrq.gov/health-literacy/professional-training/shared-decision/index.html. Accessed 22 April 2022.

174. de Angst IB, Weernink MG, Kil PJ, van Til JA, Cornel EB, Takkenberg JJ. Development and usability testing of a multi-criteria value clarification methods for patients with localized prostate cancer. Health informatics journal. 2020;26(1):486–98.

175. Weernink MGM, van Til JA, Witteman HO, Fraenkel L, MJ IJ. Individual Value Clarification Methods Based on Conjoint Analysis: A Systematic Review of Common Practice in Task Design, Statistical Analysis, and Presentation of Results. Medical decision making : an international journal of the Society for Medical Decision Making. 2018;38(6):746–55.

176. Whitty JA, Fraenkel L, Saigal CS, Groothuis-Oudshoorn CGM, Regier DA, Marshall DA. Assessment of Individual Patient Preferences to Inform Clinical Practice. Patient. 2017;10(4):519–21.

177. Taylor J, Walsh T, Worthington H, Brocklehurst P, Pemberton MN, Glenny AM. Cochrane and the COMET initiative: developing the evidence base in oral medicine. British dental journal. 2017;223(9):729–32.

178. Tugwell P, Boers M, Brooks P, Simon L, Strand V, Idzerda L. OMERACT: an international initiative to improve outcome measurement in rheumatology. Trials. 2007;8:38.

179. Jimenez-Moreno AC, Nikolenko N, Kierkegaard M, Blain AP, Newman J, Massey C*, et al.* Analysis of the functional capacity outcome measures for myotonic dystrophy. Annals of clinical and translational neurology. 2019;6(8):1487–97.

180. International Consortium For Health Outcomes Measurement. https://www.ichom.org/.

181. Nordin Å, Taft C, Lundgren-Nilsson Å, Dencker A. Minimal important differences for fatigue patient reported outcome measures-a systematic review. BMC medical research methodology. 2016;16:62.

182. International Academy of Health Preference Research. Health Preference Study and Technology Registry. https://hpstr.org/landing.

183. Fairchild AO, Katz EG, Reed SD, Johnson FR, Dibernardo AB, Hough DW*, 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. 2020;37:67–78.

184. Levitan B, Markowitz M, Mohamed AF, Johnson FR, Alphs L, Citrome L*, et al.* Patients' Preferences Related to Benefits, Risks, and Formulations of Schizophrenia Treatment. Psychiatr Serv. 2015;66(7):719–26.

185. Markowitz MA, Levitan BS, Mohamed AF, Johnson FR, Bridges JF, Alphs L*, et al.* Psychiatrists' judgments about antipsychotic benefit and risk outcomes and formulation in schizophrenia treatment. Psychiatr Serv. 2014;65(9):1133–9.

186. Hauber AB, Arden NK, Mohamed AF, Johnson FR, Peloso PM, Watson DJ*, et al.* A discrete-choice experiment of United Kingdom patients' willingness to risk adverse events for improved function and pain control in osteoarthritis. Osteoarthritis Cartilage. 2013;21(2):289–97.

187. Arden NK, Hauber AB, Mohamed AF, Johnson FR, Peloso PM, Watson DJ*, et al.* How Do Physicians Weigh Benefits and Risks Associated with Treatments in Patients with Osteoarthritis in the United Kingdom? The Journal of Rheumatology. 2012;9(5).

188. Food & Drug Administration Briefing Document. Joint Meeting of Arthritis Advisory Committee and Drug Safety and Risk Management Advisory Committee. BLA 761130. Tanezumab. March 24–25, 2021.

189. Ho M, Gonzalez JM, Mansfield C, Wilson L. Three preference studies sponsored by FDA CBER to inform benefit-risk assessment for biologics. Presented at the Virtual ISPOR 2021 Conference; 20 May 2021.

190. Osoba D, Hsu MA, Copley-Merriman C, Coombs J, Johnson FR, Hauber B*, et al.* Stated preferences of patients with cancer for health-related quality-of-life (HRQOL) domains during treatment. Qual Life Res. 2006;15(2):273–83.

191. Johnson FR, Hauber AB, Osoba D, Hsu MA, Coombs J, Copley-Merriman C. Are chemotherapy patients' HRQoL importance weights consistent with linear scoring rules? A stated-choice approach. Qual Life Res. 2006;15(2):285–98.

192. Mohamed AF, Hauber AB, Johnson FR, Coon CD. Patient preferences and linear scoring rules for patient-reported outcomes. Patient. 2010;3(4):217–27.

193. Lambooij MS, Harmsen IA, Veldwijk J, de Melker H, Mollema L, van Weert YWM*, et al.* Consistency between stated and revealed preferences: a discrete choice experiment and a behavioural experiment on vaccination behaviour compared. BMC medical research methodology. 2015;15(1):19.

194. Salampessy BH, Veldwijk J, Jantine Schuit A, van den Brekel-Dijkstra K, Neslo REJ, Ardine de Wit G*, et al.* The Predictive Value of Discrete Choice Experiments in Public Health: An Exploratory Application. The Patient - Patient-Centered Outcomes Research. 2015;8(6):521–9.

195. Quaife M, Terris-Prestholt F, Di Tanna GL, Vickerman P. How well do discrete choice experiments predict health choices? A systematic review and meta-analysis of external validity. The European journal of health economics : HEPAC : health economics in prevention and care. 2018;19(8):1053–66.

196. de Bekker-Grob EW, Swait JD, Kassahun HT, Bliemer MCJ, Jonker MF, Veldwijk J*, et al.* Are Healthcare Choices Predictable? The Impact of Discrete Choice Experiment Designs and Models. Value Health. 2019;22(9):1050–62.

197. de Bekker-Grob EW, Donkers B, Bliemer MCJ, Veldwijk J, Swait JD. Can healthcare choice be predicted using stated preference data? Social science & medicine (1982). 2020;246:112736.

198. Marshall DA, Gonzalez JM, MacDonald KV, Johnson FR. Estimating Preferences for Complex Health Technologies: Lessons Learned and Implications for Personalized Medicine. Value Health. 2017;20(1):32–9.

199. Harrison M, Bansback N, Aguiar M, Koehn C, Shojania K, Finckh A*, et al.* Preferences for treatments to prevent rheumatoid arthritis in Canada and the influence of shared decision-making. Clin Rheumatol. 2020;39(10):2931–41.

200. Harrison M, Spooner L, Bansback N, Milbers K, Koehn C, Shojania K*, et al.* Preventing rheumatoid arthritis: Preferences for and predicted uptake of preventive treatments among high risk individuals. PloS one. 2019;14(4):e0216075.

201. Bansback N, Harrison M, Marra C. Does Introducing Imprecision around Probabilities for Benefit and Harm Influence the Way People Value Treatments? Medical Decision Making. 2016;36(4):490–502.

202. Fraenkel L, Nowell WB, Michel G, Wiedmeyer C. Preference phenotypes to facilitate shared decision-making in rheumatoid arthritis. Ann Rheum Dis. 2018;77(5):678–83.

203. Ho KA, Acar M, Puig A, Hutas G, Fifer S. What do Australian patients with inflammatory arthritis value in treatment? A discrete choice experiment. Clin Rheumatol. 2020;39(4):1077–89.

204. Patient Focused Medicines Development. Medicine and device lifecycles, digital health, data and health systems for better outcomes. https://patientfocusedmedicine.org/.

ANNEX TO THE PREFER RECOMMENDATIONS

Why, when and how to assess and use patient preferences in medical product decision-making

The PREFER project took place from October 2016 to May 2022.

Annex Table of Contents

FULL LIST OF CONTRIBUTORS TO THE PREFER RECOMMENDATIONS

PREFER consortium members

Kate Adcock (Muscular Dystrophy UK), Greg Anglin (Eli Lilly), Kathi Apostolidis (European Cancer Patient Coalition), Annlouise Assaf (Pfizer), Quazi Ataher (CSL Behring), Luca Bailo (European Institute of Oncology), Conny Berlin (Novartis), Andrea Beyer (Actelion), Mondira Bhattacharya (Abbvie), Susan Bhatti (Merck KGaA), Alasdair Blain (Newcastle University), Christelle Boeynaems (Sanofi), Monika Borchert (Novartis), Monika Brand (Actelion), Gudrun Briat (Belgian Health Care Knowledge Centre), Gunnar Brobert (Bayer), Thomas Brookland (Roche), Andreas Brueckner (Novartis), Connor Buffel (MindBytes), Kristin Bullok (Eli Lilly), Lyndsey Butterworth (Newcastle University), Enrica Capeletto (Istituto Tumori Bari), Antonella Cardone (European Cancer Patient Coalition), Klevisa Ceka (European Cancer Patient Coalition), Nicola Charpentier (Bayer), Dasha Cherepanov (Takeda), Antonio Ciaglia (International Alliance of Patient Organisations), Irina Cleemput (Belgian Health Care Knowledge Centre), Mireille Collombat (Actelion), Nigel Cook (Novartis), Lucia D'Apote (Amgen), Andrii Danyliv (Novartis), Evan Davis (Roche), Esther de Bekker-Grob (Erasmus University Rotterdam), Francesco De Lorenzo (European Cancer Patient Coalition), Ardine de Wit (University Medical Centre Utrecht), Sheila Dickinson (Novartis), Charalabos-Markos Dintsios (Bayer), Rachael DiSantostefano (Janssen), Bas Donkers (Erasmus University Rotterdam), Emma Du Four (Abbvie), Ilaria Durosini (European Institute of Oncology), Christine Dyer (Newcastle University), Laurent Ecochard (Novartis), Matthias Englbrecht (University of Erlangen), Mats Ericson (Amgen), Aina Laura Errando (European Cancer Patient Coalition), Marie Falahee (University of Birmingham), Josepine Fernow (Uppsala University), Reinhard Fescharek (CSL Behring), Bruno Flamion (Actelion), Patricia Fleuranceau-Morel (Merck KGaA), Emuella Flood (AstraZeneca), Emily Freeman (Abbvie), Patrick Frey (Amgen), Elisabeth Furberg (Uppsala University), Domenico Galetta (Istituto Tumori Bari), Martin Gebel (Bayer), Pravin Girase (Roche), Charis Girvalaki (European Cancer Patient Coalition), Markus Goese (Roche), Grainne Gorman (Newcastle University), Heiko Götte (Merck KGaA), Matthias Gottwald (Bayer), Karin Groothuis-Oudshoorn (Uppsala University), Oliver Guenther (Merck KGaA), Paolo Guiddi (European Institute of Oncology), Florian Gutzwiller (Novartis), Tarek Hammad (Sanofi), Mats Hansson (Uppsala University), Sarah Harding (Takeda), Brett Hauber (Pfizer), Stephen Heaton (Bayer), Richard Hermann (AstraZeneca), Esteban Herrero-Martinez (Abbvie), Ceri Hirst (Bayer), Kimberley Hockley (Amgen), Anna Holm (Uppsala University), Axel Hueber (University of Erlangen), Isabelle Huys (KU Leuven), Telba Irony (Janssen), Mariam Jamali (Bayer), Ellen Janssen (Janssen), Rosanne Janssens (KU Leuven), Sandra Jeleazcov (University of Erlangen), Cecilia Jimenez-Moreno (Evidera), Byron Jones (Novartis), Mathieu Jouannin (Sanofi), Juhaeri Juhaeri (Sanofi), Eva Katz (Janssen), Mazzocco Ketti (European Institute of Oncology), Mona Khalid (Takeda), Ulrik Kihlbom (Uppsala University), Plamen Kozlovski (Novartis), Hilke Kracker (Actelion), Jürgen Kübler (CSL Behring), Bennett Levitan

Advisory board members

Patient advisory board members

Emma-Jayne Ashley (Cure Myotonic Dystrophy UK Charity), Eva Johansson (Swedish Rheumatism Association), Rachel Ward (patient representative)

Regulatory and HTA advisory board members

Juan-Antonio Blasco (Health Technology Assessment Area-AETSA, Andalusian Public Foundation Progress and Health-FPS), Jacoline Bouvy (NICE), Karen Facey (University of Edinburgh), Simon Fifer (CaPPRe), Inaki Gutierrez-Ibarluzea (Basque Foundation for Health Innovation and Research), Michelle Mujoomdar (Canadian Agency for Drugs and Technologies in Health)

Other

Nicholas Brooke (PFMD), Carol Mansfield (RTI Health Solutions), Kevin Marsh (Evidera), Tommi Tervonen (Evidera), Elise Quik (Zorginstitut Nederland / National Health Care Institute)

Medical writers

Susan Bromley (EpiMed Communications, funded by Bayer), Diane Shoda (Greyscaling, funded by Pfizer)

Funding

The PREFER project received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 115966. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and the European Federation of Pharmaceutical Industries and Associations (EFPIA).

As a public-private collaborative research project, members of PREFER were from academia, biopharmaceutical companies, patient organisations, Health Technology Assessment bodies and small and medium-sized enterprises.

Public partners of the PREFER project received funding from the European Commission. Private organisations provided funding of their work as in-kind. Beyond in-kind, several companies have contributed financially to the conduct of meetings and workshops.

Additional financial contributions have been provided by Bayer and Novartis in support of the EMA qualification process with regard to regulatory fee payments, medical writing, and consulting support. Both public and private partners conducted and collaborated in research efforts during the PREFER project. Both public and private partners of the PREFER project, as well as public and private entities outside of the PREFER consortium, may use the Recommendations in future research efforts and medical product decision-making.

A1 Annex for Section 1

A1.1 Patient preference studies conducted in the context of regulatory and HTA body/payer decision-making

Table A1-1 contains a non-exhaustive list of patient preference studies that involved regulatory bodies or HTA agencies, or which were performed by industry with the aim of informing regulatory benefit–risk assessment and/or payer decisions. See **Section A4.2** for a discussion of further preference studies that included regulatory involvement.

Table A1-1. Patient preference studies that have been conducted in the context of decision-making by medical product regulators and HTA bodies/payers.

Title of the patient preference study	Methodology used	Additional details
Individual Trade-Offs Between Possible Benefits and Risks of Cancer Treatments: Results from a Stated Preference Study with Patients with Multiple Myeloma	Multicriteria decision analysis and swing weighting	EMA regulators and the cancer charity Myeloma UK were involved in the set-up of this study, which aimed to illustrate how preference data can be gathered and used to estimate patients' acceptance of new oncology treatments.(1)
Incorporating patient preferences into medical product development and regulatory decision making: Results from a quantitative pilot study with cancer patients, carers, and regulators	A short online questionnaire, ordinal statements regarding the desirability of different outcomes in the treatment of advanced cancer	EMA regulators were involved in this study, which aimed quantify preferences regarding the desirability of treatment outcomes from advanced cancer patients, carers, regulators, and healthcare professionals.(2)
The VALUE study to quantify the relative value of different outcomes in the treatment of multiple sclerosis.	Ratings-based conjoint analysis	The EMA was actively engaged in and supported this VALUE study, which was conducted in collaboration with the Multiple Sclerosis Society in the United Kingdom (UK) and used the MACBETH software programme.(3)

Title of the patient preference study	Methodology used	Additional details
The Tubes Under Local Anesthesia (Tula) System study, to determine performance threshold for use as the primary endpoint	Threshold technique	The FDA recently approved a tympanostomy delivery system in which a patient preference study determined the performance threshold to use as the primary endpoint in the pivotal clinical trial for the procedure.(4, 5)
Endpoint identification in COPD, to understand patient-relevant symptoms	Qualitative patient insights and quantitative DCE	Novartis used a multi-phase approach (literature search, social media listening study, online bulletin board followed by quantitative patient preference study to understand the symptoms that patients with COPD think are most important to treat.(6) In the UK, the National Institute for Clinical Excellence (NICE) provided scientific advice on the design of the study, which may provide support for including additional patient-centred endpoints in clinical trials of COPD treatments.(7)
Maestro Rechargeable System for Obesity, 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	DCE	The FDA used the results of an FDA-sponsored patient preference study (8) 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.(9)
Spavato Nasal Spray for Treatment Resistant Depression, a preference study to assess patients' trade-off preferences for key benefits and harms associated with treatments for treatment-resistant depression	DCE	Janssen presented the results of a patient preference study to an FDA advisory committee as part of the FDA approval process.(10) Physicians on the committee indicated that the patient preference information helped them understand the patient voice.(11)

Title of the patient preference study	Methodology used	Additional details
NxStage Solo for Home Haemodialysis, a study to understand patients' views on the acceptability of trade-offs between treatment characteristics.	TT	NxStage and a patient advocacy group conducted a patient preference study to determine the MAR of death resulting from needle dislodgement that patients would accept to have home haemodialysis.(12) The results of the study were used to support an expansion of the indications for use to allow patients to use home haemodialysis without having a care partner present. (13)
Dexcom G5 Continuous Glucose Monitoring, a survey to understand the concerns of patients and parents about the safety of an insulin pump	Qualitative study	The FDA conducted qualitative interviews with patients and parents to understand their perspectives on the safety of using an insulin pump (FDA Center for Devices and Radiological Health [CDRH] Patient Engagement).(14) As a result, the FDA and the company developed risk 132itigateion measures designed to prevent unintended insulin boluses.
Rituxan HYCELA Labeling, a study to understand patients' views on intravenous compared to subcutaneous formulation. The same approach was taken for Herceptin (15):	Patient preference questionnaire in an interventional study	Genentech commissioned an interventional patient preference study using a crossover trial design to compare subcutaneous to intravenous administration of rituximab in blood cancers.(16) 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.(17)

In addition, as of 2021, other known preference studies informing the FDA's regulatory decision-making are ongoing. Particularly, the CDRH mentions collaborations in the area of medical devices for obesity,(18-22) Parkinson's disease,(18, 21) amputation,(19, 22) minimally invasive glaucoma surgical devices,(20) and a project with the Kidney Health Initiative.(23) There are also known applications of preference studies by the Center for Biologics Evaluation and Research (CBER); the 2018 FDA director's report (24) mentions that CBER launched three patient preference studies for prospective CBER products in the disease areas of osteoarthritis, sickle cell disease, and brittle diabetes.(25)

A1.2 Out-of-scope of PREFER

A1.2.1 Patient perspective data other than patient preference data

The collection and use of patient perspective data other than qualitative and quantitative data is out of scope. Patient preferences are a particular type of patient perspectives as illustrated in **Figure 1-1** of the PREFER recommendations.

A1.2.2 Healthcare professional preferences

Although the description of the PREFER framework focuses on how it could be used to collect patients' (and potentially caregivers') views, it could also be used to design, conduct, and analyse a study of healthcare professional preferences. While this information could add value and provide further evidence to inform regulatory decision-making, considering the specific issues related to healthcare professional preferences is beyond the scope of the proposed framework. The importance of the patient's voice in the medical product life cycle is clearly acknowledged, and the main objective of the PREFER framework is 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 healthcare professional preferences can certainly add an important and potentially different perspective, PREFER does not advocate that patient preference studies should always be accompanied by parallel healthcare professional preference studies.

A1.2.3 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 project. Although the framework is not focused on the use of preference data in shared decision-making, the preference data obtained within the framework can nevertheless help inform shared decision-making. Additionally, many insights gained from patient preference studies 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 the 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.

A1.2.4 HTA and payers' decisions based on cost-per-QALY calculations, obtained using public preferences for generic health outcomes

Although QALYs can, in principle, be calculated based on patient preferences (using a utility measurement method), they are usually based on public preferences, to take societal preferences into account in resource allocation decisions for healthcare as a whole. The rationale for using societal preferences in such decisions is that the general public finances healthcare as a taxpayer and its preferences should therefore matter in the decision-making process. Moreover, for system-wide decisions, generic health outcome measures enable comparisons between different diseases.

PREFER recognises the importance to many HTA agencies and payers of health economic evaluations performed from the societal perspective using generic outcome measures; however, it does not see these as topics that it would be able to usefully cover within the PREFER framework.

A1.2.5 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 such guidelines.

A2   Annex for Section 2

There is no annex content for Section 2.

A3   Annex for Section 3

This annex to Section 3 of the PREFER recommendations is intended as stand-alone content on the PREFER framework for patient preference studies and includes:

- **additional detail** on some framework material presented at a very high level in **Section 3**

- framework-related content that is not covered in **Section 3**
    - additional content on applications of preference data to inform medical product decision-making in Sections **A3.4.1.3**–**A3.4.1.6**
    - **technical methods** for the application of preference data in **Section A3.4.2**
    - suggestions on incorporating preference information into industry, regulatory, and HTA/payer documents in **Section A3.4.3**
    - further discussion of the PREFER framework in **Section A3.5**, including a review of how the PREFER framework supports scientific integrity and credibility of patient preference studies (**Section A3.5.4**)
    - an overview of the (minor) differences in the framework content included in the Qualification procedure with the EMA and the framework content presented here (**Section A3.5.5**).

Sections A3.3.2.1, **A3.3.2.5**, and **A3.4.3** include quotes from the CHMP Qualification Opinion about the importance of registering patient preference studies, choice profiles, and the use of preference data in industry, regulatory, and labelling documents, respectively. Further CHMP comments on specific aspects of the PREFER framework can be found within the Qualification Opinion.

A3.1   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.

The proposed design of preference studies intended to inform regulatory and/or HTA decisions should be discussed with the relevant decision-makers before initiating the study. The aspects of the framework that are especially helpful to discuss during a scientific advice process are described in **Section 4.2** of the main PREFER recommendations.

The objectives of the PREFER framework for patient preference studies are to:

- inform study research teams 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 medical product decision-making

- support discussions between industry, regulators, and HTA bodies and payers about preference studies intended to inform medical product decision-making.

The framework builds on much of the existing work in this area – such as the overview of patient preferences in the MDIC 2015 report,(26) the ISPOR report on good research practices for conjoint analysis,(27) the FDA's 2016 guideline about submission of preference data,(28) more recent work on patient preference research,(29) and the overview of preference study stages and steps from van Overbeeke and colleagues.(30) It also synthesises foundation work of the PREFER project that included systematic literature reviews, comprehensive stakeholder interviews, and case studies.

The PREFER framework is intended to cover all types of patient preference study, irrespective of the method used and is thereby applicable to both qualitative and quantitative studies. It has three broad components that can be mapped to the stages of a preference study (**Figure A3-1**).
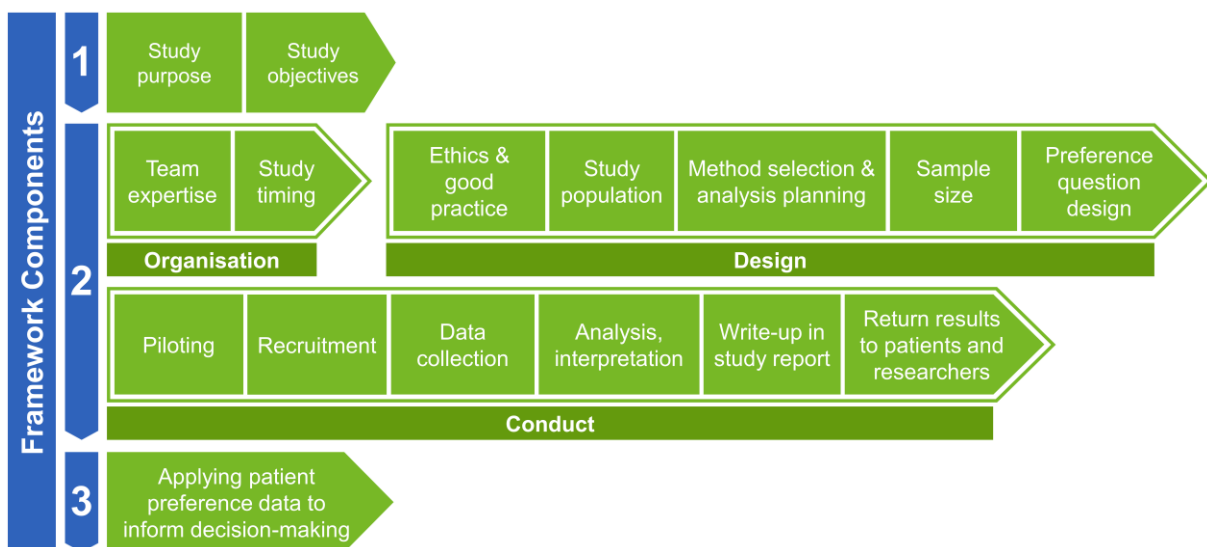
Figure A3-1. PREFER framework structure, aligned with stages and steps of patient preference studies. *Adapted from van Overbeeke et al, 2019.*(30)

A3.2 PREFER framework component 1: the preference study purpose and objectives

This section describes the suggested structure to characterise how the preference data will inform decision-making by describing the preference study purpose (**Section A3.2.1**), and the consideration of preference study objectives (**Section A3.2.2**).
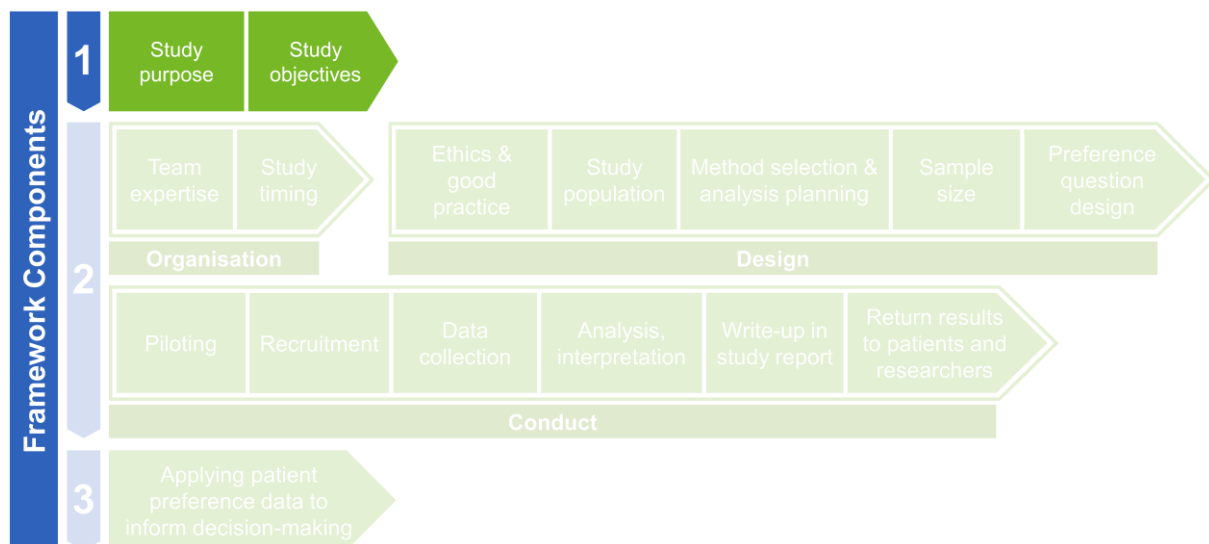
Figure A3-2. Stages of PREFER framework component 1.

A3.2.1 Framework component 1: the preference study purpose – what decision will be informed by the preference study

The preference study purpose should explain how the preference data will inform medical product decision-making. A patient preference study should only be performed if it is clear from the study purpose that it addresses a preference-sensitive situation, 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 and colleagues,(27) the study purpose should include information about:

- the decision and decision-makers, i.e. what decisions and by whom (industry, regulator, HTA body) will be informed by the results from the patient preference study, together with information about the relevant decision contexts

- how this decision is preference-sensitive

- whose preferences are of interest.

These topics are discussed in more detail below.

The decision and decision-makers, and the relevant decision context

A description of the decision should include information about the decision-makers and decision context. The study purpose could 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. Examples of how preference studies can inform industry, regulatory, HTA and payer decisions are described in **Section 2.1** of the main PREFER recommendations.

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.

Decisions relating to patients' views on trade-off or uncertainty situations will generally involve the combined use of preference and clinical data. This requires especially careful consideration of the decision context to ensure that preference data can be suitably aligned with clinical data, as is required for the technical application of the preference data. This technical application could be either preference data *used in parallel* with clinical or other data (**Section A3.4.2.2**), or preference data *combined* mathematically with clinical or other data (**Section A3.4.2.3**).

Points to consider about the decision context that are especially important when planning to use preference data together with clinical data include:

- The other treatment options relevant to the applicable decision makers. 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.

- The key benefits and key risks relevant to the decision (as discussed further in ICH M4E(R2),(31) bearing in mind that:

    – there may be key risks associated with the other treatment options that are not associated with the new treatment and vice versa

    – the choice of key benefits and key risks may vary by decision-maker (e.g. the regulatory choice of key benefits could differ to the HTA choice of key benefits) – note that 'key benefits' can include aspects such as convenience (e.g. mode of administration)

- The known or expected magnitude of the clinical effect of the new medical product relative to the other treatment options for these key benefits and key risks (**Section A3.3.2.5**).

The first two points above help ensure that the preference data intended to support the development of a new treatment can be aligned not only with its own key benefits and risks, but also the key benefits and risks associated with the other treatment options. For example,

suppose that the current standard of care for indication X is a treatment associated with a side-effect of angioedema, and that the new treatment is associated with a side-effect of hypertension. A preference study intended to inform decision-making related to the new medical product may well need to include features relating to both hypertension and angioedema, as well as features relating to the benefits relevant to the indication in question.

Point 3 helps ensure that the preference data intended to support the development of a new treatment reflects patients' preferences about the clinical effect size relevant to the decision. For example, suppose that the new treatment for indication X is 5 units better than the current standard of care. A preference study intended to support decision-making about the new treatment should be able to show the importance placed by patients on a 5-unit difference (and it would, for example, be problematic if the largest example effect size included in the preference study was only 3 units).

Describing how this decision is preference-sensitive

As discussed in **Section 2.1** of the main PREFER recommendations, and consistent with the reasons for using patient preferences described by van Overbeeke *et al*,(32) preference-sensitive decisions include situations where there is a need to understand patients' views when:

- it is unclear which are the most important outcomes or attributes for a disease or medical product

- multiple treatment options exist (including status quo and standard of care) and there is no option that is clearly superior for all patients

- the evidence supporting one option over others is considerably uncertain or variable, and patients' tolerance for this uncertainty may impact their decisions.

Describing whose preferences are of interest

The description of whose preferences are of interest should be appropriate to the decision identified in the study purpose. In a situation where a preference study is intended to support the development of a specific medical product, this will typically be aligned with the target population for the medical product. A description of whose preferences are of interest would usually include a description of the disease and the associated population., This would include the target age-group (e.g. children, adults), the degree of experience of the patients with the disease (e.g. at risk, newly diagnosed), the severity of disease (e.g. mild, moderate, severe), the current stage the patient is in the treatment journey (e.g. taking first-line, second-line treatment). For example, the description of patients whose preference are of interest could be: 'Patients with type 2 diabetes aged ≥18, or 'Patients with asthma aged ≥18 requiring add-on therapy'.

The study purpose should clearly describe whose preferences are of interest even if this group's preference cannot be collected directly, such as young children, individuals with cognitive problems like Alzheimer's disease.

Further discussion of the study population, and when it may be appropriate to survey caregivers instead of patients, is available in **Section A3.3.2.2**.

A3.2.2 PREFER framework component 1: preference study objectives – how the preference study will inform this decision

Once there is alignment on the preference study purpose, the study team should identify appropriate study objectives and the preference study endpoints associated with each objective. It is especially important that the study team ensures alignment with the decision-makers on this aspect of the preference study (**Section 4.2** of the main PREFER recommendations).

It is helpful to develop the primary objectives of a preference study with the end use of the preference data in mind, which will be to inform the decision as per the study purpose. The connection between preference study objectives and end use can be made by considering the link from study objective to preference study endpoint, and from preference study endpoint to application of preference data to inform the medical product. Examples of these links are shown in **Figures A3-3**, **A3-4** and **A3-5**.
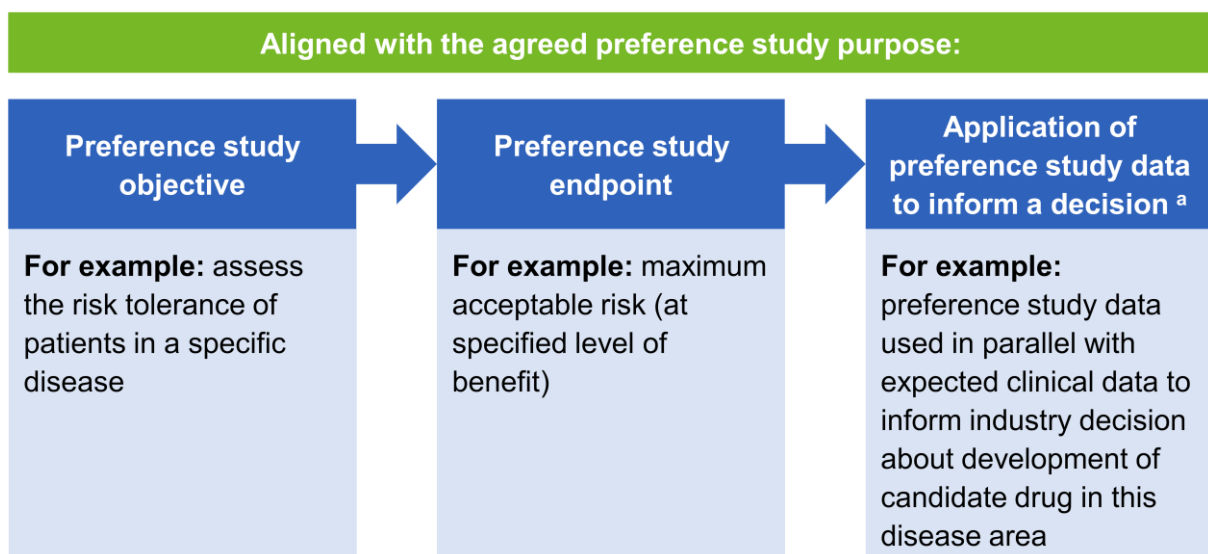
Figure A3-3. Link from study objective to study endpoint to application: example 1.

*ᵃ The application of preference data to inform decisions can involve the use of preference data in isolation, preference data in parallel with clinical data, and/or preference data mathematically combined with clinical data (**Section A3.4.2**).*

Aligned with the agreed preference study purpose:		
Preference study objective	**Preference study endpoint**	**Application of preference study data to inform a decision** [a]
For example: assess the preferences of patients with disease X for first-treatment options when choosing between new drug A and existing treatment option B	**For example:** preference weights for attributes aligned with the key benefits, key risks of treatment options A, B	**For example:** preference study data used in combination with clinical data (on drug A vs. drug B) to obtain preference share information for new drug A, to inform **regulatory decision** about approval

Figure A3-4. Link from study objective to study endpoint to application: example 2.

[a] *The application of preference data to inform decisions can involve the use of preference data in isolation, preference data in parallel with clinical data, and/or preference data mathematically combined with clinical data (**Section A3.4.2**).*

Aligned with the agreed preference study purpose:		
Preference study objective	**Preference study endpoint**	**Application of preference study data to inform a decision** [a]
For example: understand which aspects of disease Y are important to patients	**For example:** preference weights for different disease Y symptoms (attributes)	**For example:** preference study data used in isolation to inform industry decision (and industry discussion with regulator) about choice of endpoints in a submission clinical study
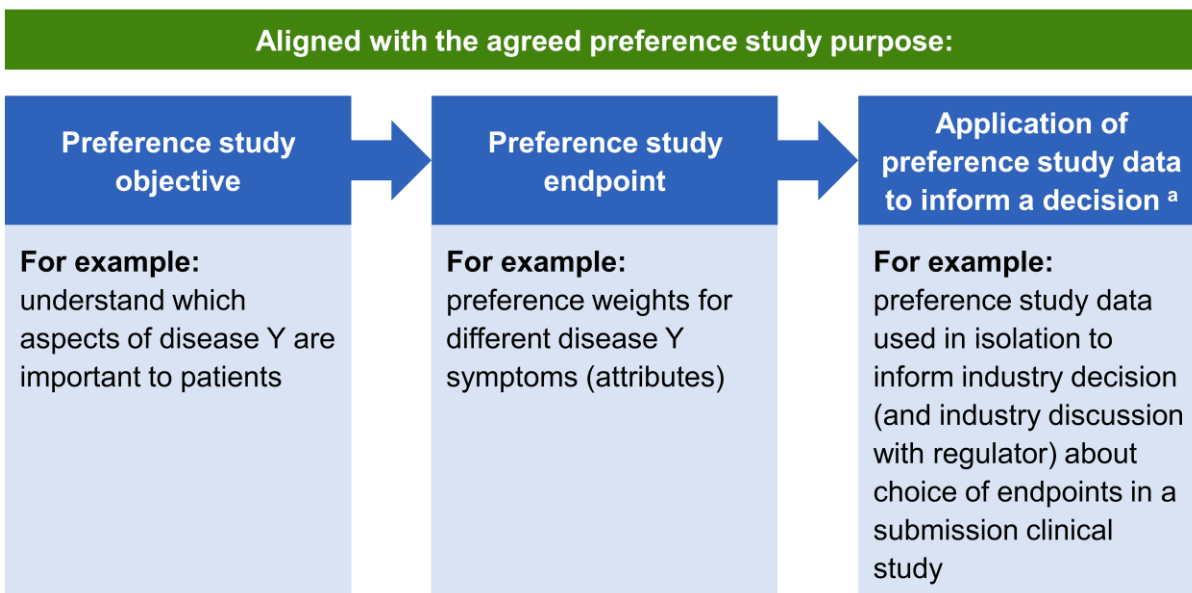
Figure A3-5. Link from study objective to study endpoint to application: example 3.

[a] *The application of preference data to inform decisions can involve the use of preference data in isolation, preference data in parallel with clinical data, and/or preference data mathematically combined with clinical data (**Section A3.4.2**).*

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, such as patients with varying time since diagnosis of the disease, patients with differing severity of disease, and patients who do or do not have experience of a specific adverse event.

- investigating whether patient preferences are associated with specific characteristics, such as socio-demographic characteristics, psychological constructs, and disease state characteristics.

A3.3   PREFER framework component 2: the preference study organisation, design, and conduct

A3.3.1   Framework component 2, organisation
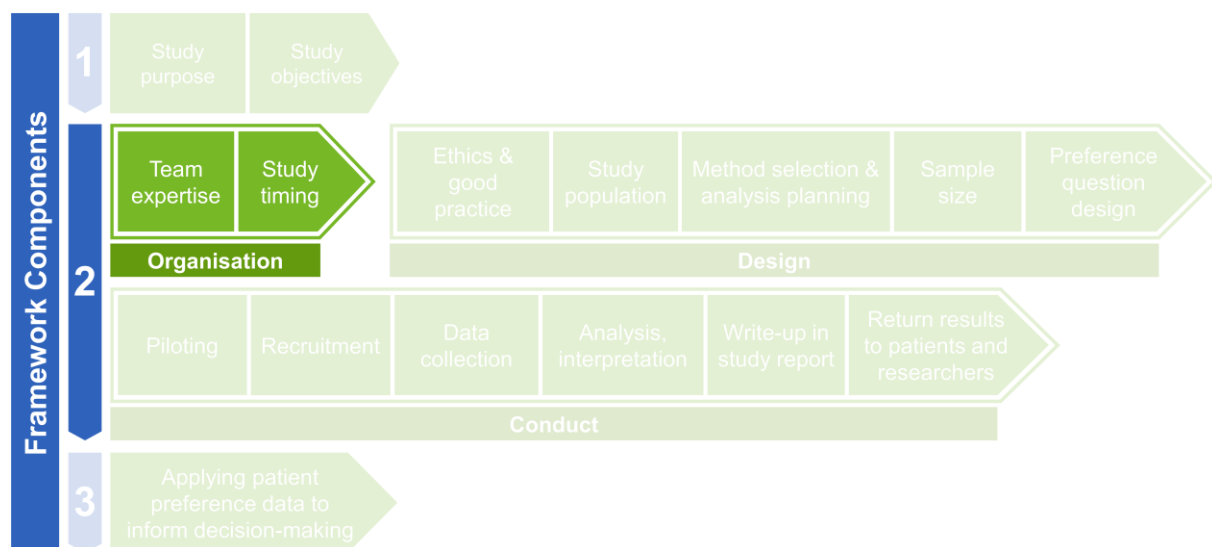
Figure A3-6. Stages of PREFER framework component 2: organisation.

A3.3.1.1   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. As with any team effort, clarity on the role and responsibilities of each preference study team member, as well as collaboration and communication with the preference study team, are critical to success.

Expected areas of expertise

Patient involvement in patient preference studies is critical – as stated in the FDA Patient Preference Information from 2016,(28) and also by Van Overbeeke and colleagues,(30) 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 A4.1** for further discussion of this topic. In addition to working with patient research partners, the team responsible for a preference study should include members with expertise in:

- medical aspects of the disease and its treatments

- statistics used in preference study design and analysis

- the conduct of preference study.

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:

- patient engagement

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

- HTA/reimbursement activities (for a preference study intended to support HTA decision-making)

- patient-reported outcomes (PRO expertise can be particularly valuable for a qualitative preference study, because concept elicitation for PROs is very similar to components of qualitative preference studies; PRO expertise can also be valuable if the attributes in the survey are PRO-based)

- medical product development

- psychological constructs (if planning to include them in the survey)

- patient educational material.

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 or an advisory board). A preference study may be conducted by, or in collaboration with, an external partner (e.g. external consultant, patient association) and/or 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).

The extended team working on a preference study would typically include members with expertise in further areas, including medical writing and legal and/or compliance activities. For a preference study intended to inform decisions on a specific medical product, the preference study team may benefit from including members of the team working on the product (e.g. epidemiologist; observational study expert, communications manager, person leading the cross-functional team working on the product).

A3.3.1.2 Framework component 2, organisation – preference study timing

Preference study timing requires consideration of whether to perform a qualitative and/or quantitative study.

Qualitative studies:

- can collect descriptive data by consulting or observing participants (or phenomena), and/or examining their subjective experiences and decisions

- can help provide insights into aspects of a disease or treatment that matter most to patients

- can be conducted in conjunction with quantitative studies, e.g. where the results from the qualitative study can identify attributes to be included in the quantitative study.

Quantitative studies:

- can collect quantifiable data that can be evaluated through statistical inferences or analysis.

Preference studies that aim to identify patient-relevant endpoints are typically composed of a *qualitative* phase to identify appropriate attributes of relevant existing or desired treatments, followed by a *quantitative* phase to understand patients' views on the relative importance of these attributes. Preference studies that aim to understand patients' views on the acceptability of trade-offs between treatment characteristics would typically be a quantitative study that includes a qualitative element (in the 'pre-testing' phase) to check whether the survey's content are clear and understood in the intended manner by the intended study population.

The timing of the preference study should be such that it allows the results to be available in a timely manner to inform the decision described in the study purpose. For a preference study related to a decision about a specific medical product, the timing should also be aligned with the availability of an appropriate level of knowledge about the associated medical product. 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 earlier decision.

Examples of preference study timing:

- **A preference study to inform the choice of patient-relevant endpoints.** The timing would typically be such that the results are available prior to the design of the regulatory submission study that would incorporate the chosen endpoints.

- An industry-sponsored preference study to provide information about a scenario involving the acceptability of trade-offs between treatment characteristics for regulatory decision-making. The timing would typically be such that information about which key benefits and key risks are expected to contribute to the 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. (The Clinical Overview is one of the documents written by industry for inclusion in a regulatory submission, and it includes the industry description of the benefit–risk assessment of the new medical product relative to the current treatment options.)
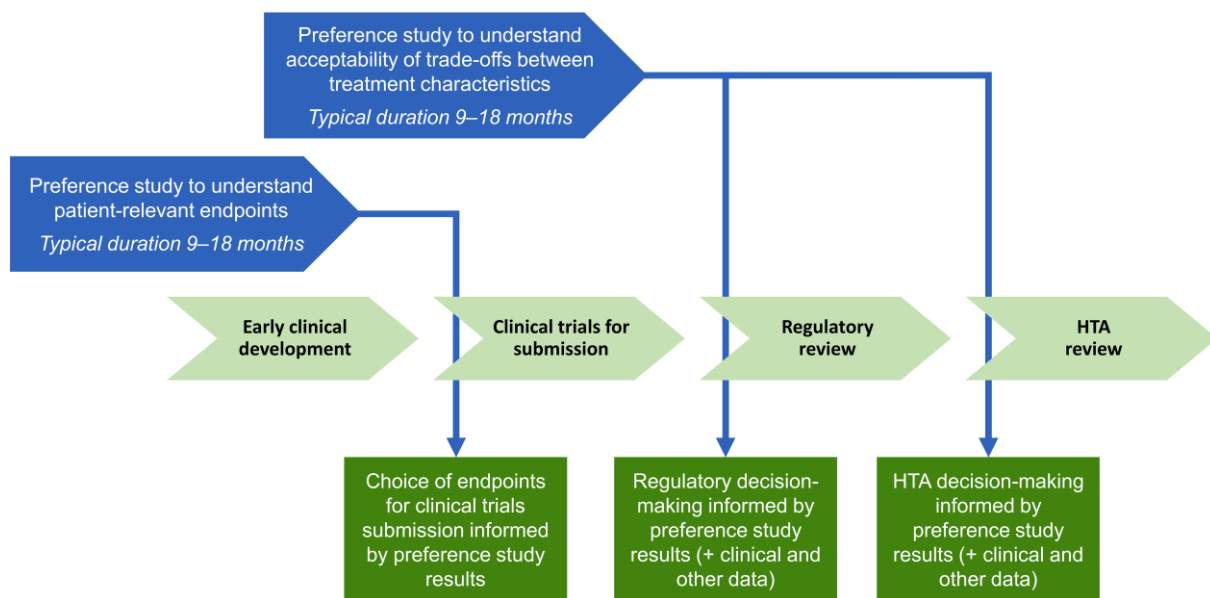
Figure A3-7. Examples of preference study timings relative to the medical product life cycle.

The timings in **Figure A3-7** are examples and can be adjusted according to the specific situation. For example, a preference study to understand the acceptability of trade-offs between treatment characteristics to inform the HTA review only could be run later than shown in this figure. If embedding the preference study within a clinical trial, a preference study to understand the acceptability of trade-offs between treatment characteristics to inform regulatory review might need to be designed prior to the start of the clinical trial for submission.

Depending on the situation, the development of a specific medical product could involve varying numbers of preference studies; for example:

One preference study:

- for a medical product in an indication without a good understanding of patient-relevant endpoints, but where the benefit–risk profile is expected to be straightforward – one preference study to understand patient-relevant endpoints

- for a medical product where the patient-relevant endpoints are well understood, but where the benefit–risk profile of the medical product is not clear-cut – one preference study to understand the acceptability of trade-offs between treatment characteristics.

Two preference studies:

- for a medical product in an indication without a good understanding of patient-relevant endpoints and where the benefit–risk profile is not expected to be clear-cut – one preference study to understand patient-relevant endpoints and another to understand the acceptability of trade-offs between treatment characteristics.

No preference studies:

- for example, if there is existing data on which endpoints are patient-relevant and the benefit–risk profile of the medical product relative to the treatment landscape is straightforward (e.g. the new medical product offers relevant additional benefit and minimal addition risk relative to the existing treatment options).

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 (**Section 4.2** of the main PREFER recommendations)
- financial resources
- study duration, including the time required for recruitment (which could be a critical factor in the overall duration of the preference study), arranging contracts with outside vendors/sub-contractors (if applicable) and adequate pilot testing of survey
- working with patients and patient representatives.

A3.3.2 Framework component 2, design
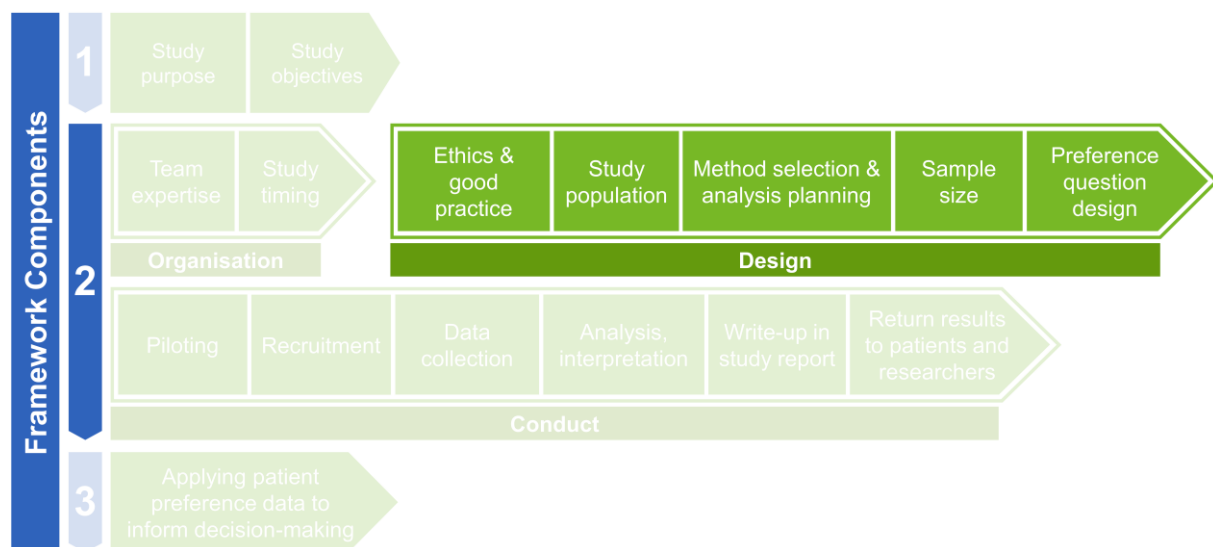
Figure A3-8. Stages of PREFER framework component 2: design.

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 (that is, the decision maker). The study purpose and objectives – as well as the planned approach to study design, conduct, and analysis – should be described in a study protocol (note that the **PREFER operational guidance** includes both a **protocol synopsis template** and a **protocol template**). Revisions to the protocol after study initiation may occur and can be captured using similar approaches to clinical trial protocol amendments.

The activities described in the 'design' section of component 2 are interrelated and hence the sequence of activities described in the figure could be adjusted as needed; for example, some aspects of analysis planning might only be feasible after work is completed on the preference question design.

A3.3.2.1   Design: ethics and good practice

As with clinical trials, patient preference studies should adhere to ethical principles, which include:

- ensuring the study was reviewed by or determined to be exempt by an appropriate ethics review board (*note:* some journals require IRB review and may not consider a study that was deemed exempt from IRB review)

- ensuring patients receive all the information they need to provide informed consent for all the planned occasions the data and their information will be used – one further option to consider is the creation of a distress protocol that describes how to assist a patient if the patient becomes emotionally or psychologically distressed. This is further discussed in **Section A3.3.2.5** Preference Question Design, Discussion Guide and Survey Design

- encouraging the study sponsor to plan a lay summary of the preference study results for participants (this aligns with current guidelines about informing trial participants about clinical trial results)

- the study sponsor should consider registering the preference study in a public registry at the protocol stage and publishing the study results; if a preference study will be conducted within clinical trials, the study sponsor should follow registration and reporting requirements for clinical trials. This point is also noted in the CHMP Qualification Opinion, which states that "The importance of transparency with regard to PPS is emphasised and it is generally recommended to publish PPS in a register ... even when the clinical trial in which the patient preference study might be embedded is already registered as clinical trial in EudraCT or clinicaltrials.gov."

Operational considerations include the:

- number of countries where ethics approval needs to be obtained and differences in requirements and timelines

- time to obtain ethics review board/IRB approval after submission of the preference study protocol

- 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.

A3.3.2.2 Design: study population

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 caregivers.

Points to consider when defining the study population

Alignment of the preference study population with the preference study purpose

See **Section A3.2.1** for high-level considerations about the definition of the preference study sample. The preference study protocol should describe further population inclusion and exclusion criteria as relevant. For example, 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 decision about a specific medical product, the preference study population should typically be aligned with the population for whom the product is intended, 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. This is to ensure that the results of the preference study can be generalised to the population of interest.(28, 33)

Consideration of the representativeness of the preference study sample

According to research across stakeholders conducted within PREFER, representativeness relates to the characteristics of the population for which the medical product is intended and the potential for extending results to the full population.(30) Characteristics related to representativeness include age groups, genders, cultures, ethnicities, geographical areas (continents, countries or regions depending on what decisions are aimed to inform), levels of education, time since diagnosis, stages and severities of disease, and treatment experiences. As for clinical trials,(34) no preference study can be totally representative of the target population given the diverse factors that may influence representativeness. Therefore, consideration should be given to what baseline participant or disease characteristics could inform the stated preferences and, by extension, interpretation of the results. It can be helpful to consider recruitment of patients from diverse sites or sources to help ensure a representative sample.(35)

Other points to consider when defining the study population

The level of knowledge of a patient (e.g. about a specific product, indication, or endpoints) would typically be irrelevant to the choice of preference study population. The design of a

preference study should include 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 A3.3.2.5**).

Consideration should be given as to whether a self-reported diagnosis of disease is acceptable, or whether a physician-confirmed diagnosis is required. A self-report approach may be more acceptable for some disease states (e.g. heart attack, type of cancer) than others with subjective measures (e.g. depression). A self-report approach should also include validation to confirm presence of the target disease state population. For example, recruiters could contact the clinic where a diagnosis was provided, or survey questions could be included to test the likelihood the person has a diagnosis, such as medications being taken, and ensure provided answers are consistent with the diagnosis.

Whether the choice is taken to have a physician-confirmed diagnosis or self-report, a clear rationale for the approach taken should be provided and the approach discussed with the stakeholder for which the data will inform their decision (**Section 4.2** of the main PREFER recommendations).

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

Depending on the preference study objectives and the decision the results will inform, there may be interest in embedding the preference study within a clinical trial for the medical product in question. In these cases – such as to directly inform preferences on the investigational product characteristics – participants can be recruited either completely within or completely outside of a clinical trial, or a combination of within and outside (e.g. the Janssen Research & Development esketamine submission).(10) Some pros and cons of each approach are described in **Table A3-2**.

Table A3-2. Pros and cons of recruiting preference study participants completely within or completely outside of clinical trials.

Approach to recruitment	Pros	Cons

Within clinical trial (e.g. if study will inform on preferences related to treatment characteristics)	• Straightforward to ensure alignment between preference study population and clinical trial population.
	
	• High degree of confidence that the preference population has the disease in question.
	
	• A large array of clinical information is available, facilitating the interpretation of preference data.
		• Participants within clinical trials are not likely to be representative of the broader patient population.
		
		• Participants 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.
		
		Operational issues
		• Requires careful consideration of the timing with respect to:
		  – preference study design, since pre-testing the components of a preference study must be done ahead of writing the clinical trial protocol if the preference study must be fully described within the protocol
		  – collecting data, as experience within the clinical trial could influence preferences; e.g. if conducted during or after the trial, those in the intervention arm may have current experience with the risks and benefits of the investigational product, hence resulting in actual experiences versus preferences.
		• Cost (possibly increased when conducting a preference study within the clinical trial)

Approach to recruitment	Pros	Cons
Outside clinical trial (e.g. if study will inform on outcomes to include in a future clinical trial)	• Potential to recruit a broader range of patients, rather than only the patients in a clinical trial. • Potential for more flexible timing. • Regardless of timing, all participants could be treatment naïve to the investigational product, hence elicitation of stated preferences for all participants.	• Self-report of diagnosis may be possible and so consideration on whether self-report or healthcare professional diagnosis is most appropriate for the intended study. • Potentially limited information on the clinical characteristics of the preference study participants. • More focus needed on how to align the preference study and clinical trial populations (only applicable if such alignment will be relevant to the decision to be informed by the preference study results).

Caregivers 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,(35) such as young children and patients with severe cognitive impairment. If the collection of caregiver preferences instead of patients is considered, it is helpful to discuss this approach upfront with the relevant decision-makers (**Section 4.2.2** of the main PREFER recommendations).

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. Caregivers' preferences can therefore be useful evidence to inform decision-making related to medical products. 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?) See the **NMD preference study** within PREFER.

• **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.(36-38)

A3.3.2.3 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. A **protocol template** and **statistical analysis plan template** are available within the **PREFER operational guidance**.

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 5** of the main PREFER recommendations for points to consider for method selection, **Table 5-3** for examples on how the different PREFER case studies approached method selection, and further discussion of this topic in **Section A5.6**.

Points to consider when planning analyses

As noted above, the PREFER framework is aligned with key principles of ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials.(39) Specifically, concepts include the use of detailed study objectives, the choice of method that aligns with these study objectives, and the selection of study metrics that align with estimand requirements appropriate to the sample and considering how the preference data will inform the decision (e.g. choice share, MRB, MAR).

- For both qualitative and quantitative methods:
  - Analyses must be pre-specified, describe the analytical approach (e.g. descriptive statistics, thematic analysis, modelling if relevant), include any planned statistical testing and explain its relevance to the study objectives.
  - Describe how responses to the survey will be used to support the study objective (**Section A3.2.2**), as well as how the objective will be applied and inform the decision (Framework Component 3).
  - The basic approach for the planned pre-specified analyses or assessments should be summarised in the protocol at the minimum; a statistical analysis plan for more detailed planning can be developed.
    - 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.

- 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 questions 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 (**Section A3.3.2.5**).

- 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 finalising 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).

– Describe whether, when, and how patient research partners will contribute to the analysis of data and interpretation of results of the patient preference study (**Section 2** of the main PREFER recommendations).

– As part of planning the analyses, data collection and a data management plan should be established and followed. All data should be handled consistently with that plan (note that a PREFER **data management template** is available).

- **For qualitative methods** – qualitative analyses should be prespecified and should follow best practices, e.g. Pope *et al*,(40) Lacey and Luff,(41) and Hollin *et al*, (42) and would include:

  – A description of which analytical approach will be used (e.g. thematic analysis, discourse analysis, grounded theory) (in the context of the current study.(43-45)

  – 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.

  – How data will be compiled, patterns identified, and results interpreted, such as who will be involved in the interpretation and how specific quotes will be selected. Describe if and how results will inform alternatives selection and associated descriptions for quantitative study (**Section 3.3.2.5**).

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

  – How variables will be defined and created.

  – A rationale for why the statistical tests and outputs are appropriate for the study.

– The approach taken for 'Assessment of survey materials' and (if applicable) piloting, including how necessary changes will be implemented and documented.

– An assessment of 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 (**Section A3.3.2.5**).

– 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 an approach on how to use such data.

– Plans and a rationale for conducting additional analyses following completion of pre-specified analyses to better understand potentially unexpected results from the preference study (**Section A3.3.3.4**). The approach to conducting such analyses, including adjustments to models used and assumptions made within those models, should be described as much as possible before beginning analysis.

An example of how PREFER studies planned the analyses in advance is shown in **Table A3-3** within a particular case study (rheumatoid arthritis); while all case studies could be examples, only one is presented for brevity.

Table A3-3. Example of rheumatoid arthritis case study approach to study analysis planning.

Approach to analysis planning	Details
Statistical analysis plans were created for both the qualitative and quantitative phases of the study, prior to study initiation and following alignment on study purpose	**Qualitative phase** 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. **Quantitative phase** Final survey content, including attributes from focus groups, were 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.

A3.3.2.4 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 sample sizing – (34) the number of participants in a preference study should always be sufficient to reliably address the study objective, including any objectives related to subgroup assessment,(28) and in light of the chosen method. If the preference study objectives include assessment of preferences by subgroups, the sample sizing should – as when assessing subgroups in a clinical trial setting – (34) include consideration of multiplicity adjustments.

Because the approach taken to justify the sample size is method-specific, see **Section 5** of main PREFER recommendations and **Section A5** for other considerations on sample sizing in the context of the certain methods.

A3.3.2.5 Design: preference question design

After the study purpose, objectives and planned application are defined, the preference question design development can begin as described in PREFER work.(32, 46) For the PREFER recommendations, preference question design comprises the totality of components within an interview or a discussion guide (for qualitative studies) or a survey instrument (for quantitative studies, for example). The core components are described below. Recommendations on these components are organised as follows:

- Background
    - Context description
    - Baseline characteristics
- Discussion guide and survey development
    - Description of alternatives
    - Developing levels
    - Inclusion of assessments (patient education and comprehension, internal validity, psychological constructs)
    - Question and exercise development
- Finalising preference question design
    - Considerations on cognitive burden and capacity
    - Assessment of the study materials
    - Translation of study materials

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 **Section 5** of the main PREFER recommendations and **Section A5**.

Background: context description

A critical part of preference question design is the study introduction because it orients the participant to the entire study(27) and can include the study purpose (**Section A3.2.1**). The introductory section includes the context description (sometimes referred to as a scenario or vignette), which has multiple purposes:

- Clearly describing the reason why the participant is being asked to take part in the study and what topics the preference questions will cover.

- Informing the participant about their role in 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.

- Fully describing 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 healthcare 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.

Background: baseline characteristics

As part of the preference question design, consideration should be given to what baseline participant or disease characteristics could inform the stated preferences and, by extension, interpretation of the results. Some characteristics may be defined in the sample definition, such as severity of disease (**Section A3.3.2.2**). The participant characteristic categories may include the following:

- demographics, such as region, ethnicity, gender, and age

- other background information, such as economic, educational, and employment status

- relevant medical history, such as degree of experience with the disease and severity of disease

- experience with types of treatments

- experience with adverse effects relevant to alternatives and their features, whether personal or second-hand knowledge if relevant.

Discussion guide and survey development: description of alternatives

Preference studies are used to understand the relative desirability or acceptability of treatments or the attributes of those treatments.(26) Treatments refer to medications, procedures, or other health-related services. Treatment attributes are the feature or characteristics of these treatments, such as clinical benefits (e.g. reduced chance of heart attack), clinical risks (e.g. increased chance of bleeding), convenience measures (e.g. dosing frequency), quality of life measures (e.g. activities of daily living), health status (e.g. SF-36), or any other property by which the alternative treatments differ. Stated differently, there are two levels at which a preference study can be conducted: the treatment level and the attribute level. Of note, the information included about treatment alternatives would take into consideration the individual decision-maker, bearing in mind that different decision-makers might have different views on which alternatives or features are important. **Section A4.2** provides considerations for study teams whose relevant stakeholder is a regulator agency or HTA body.

When developing treatment alternative and associated attribute descriptions, recommendations include:

- Care should be taken to minimise potential framing bias in the presentation and description of the attributes.

- Descriptions 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.

- Descriptions would be 'preferentially independent' – that is, the elements of one treatment alternative or attribute would not substantially overlap with another.

- Treatment attributes 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; out of pocket costs).

- Treatment alternative could be a real-world reference condition or opt-out/status quo description.

Other considerations are available in literature describing qualitative and quantitative study development.(27, 42, 47). This issue is also commented on within the CHMP Qualification Opinion, which notes that "Meaningful PPS results should be robust to variability in how choice profiles are set up (see e.g., Veldwijk et al., Value Health 2016; Vass & Payne, Pharmacoeconomics 2017 for critical discussion) or at least enable an understanding as to the magnitude and direction of potentially introduced biases due to the experimental design."

Two main approaches are used to identify treatment options or treatment attributes for inclusion in preference questions: 'top-down' and 'bottom-up'. **Table A3-4** presents a single example among the many PREFER case studies that used the bottom-up approach to develop treatment and associated attribute descriptions.

A **top-down** 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 as might expert interviews with healthcare professionals. Such approaches could be particularly relevant when preference data will be used with clinical trial data to support a medical product decision (**Section A3.4.2**). 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 that span the range of the adverse events that may potentially occur to 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 *et al*, 2019).(48) 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.

A **bottom-up** approach uses direct input 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 medial listening, and online bulletin boards. Inevitably, such conversations lead to more attributes than can be included in a quantitative survey so ranking exercises for all attributes of interest are helpful to identify attributes of greater importance than others to inform future research.

The approach taken can be a mix of both and is dependent upon the study purpose and objectives (**Sections A3.2.1** and **A3.2.2**), as well as the amount of background knowledge available to the research team. Examples of how attributes were identified in the PREFER project are provided in the **study reports**, with one example highlighted in **Table A3-4**. Regardless of the approach used, all alternative descriptions should be piloted with study participants prior to full study implementation (as described below).

Table A3-4. PREFER rheumatoid arthritis prospective case study showing bottom-up approach to treatment and attribute development.

Plan	Details
A bottom-up approach was planned. **Phase 1: Literature search plus focus groups in three countries, separate focus groups for the two sample populations used.** **Phase 2: Final content of the quantitative survey was developed in collaboration with patient research partners**	**Phase 1: Literature** A total of 3954 abstracts were reviewed. Twenty studies met the criteria for inclusion, including 15 studies in patients with rheumatoid arthritis, two studies in the general population, and three studies in first-degree relatives who are at an increased risk of rheumatoid arthritis. **Phase 1: Focus groups** Due to the Covid pandemic restrictions in 2020, 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. **Phase 2** 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. The alternative descriptions were preferentially independent. These also included non-health related items (how the medicine is taken and how often).

Discussion guide and survey development: development of levels

For some attribute-based quantitative methods (e.g. DCE, swing weighting, threshold technique) the selection of levels for each attribute is required. Selection of the type, number, and way of presenting associated levels for each attribute depends upon the study objectives and method selected, the clinical relevance, and the clarity to patients. See **Section 5** of the main PREFER recommendations and **Section A5** for method-specific considerations. However, the following are some general considerations across all methods:

- determine if absolute level vs. improvement from current status (incremental) approach can be taken

- determine if categorical or continuous variables are required

- ensure what is described is generally realistic or perceived as such; this could be based on what is anticipated from the medical product in development and/or published data for the treatment choices 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; when the preference study results will be applied together with clinical data, this includes ensuring that the range of levels encompasses (or is expected to encompass) the range of the clinical study outcomes

- use good health literacy practices when presenting levels; for example, absolute values are typically easier to understand than relative measurements or use images to illustrate the different probabilities that a side-effect will occur

- ensure the range and number of levels for each attribute are developed to minimise unintended bias towards one attribute over another.(29)

Discussion guide and survey development: inclusion of assessments

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 questionnaire development phase (see **Section 2** of the main PREFER recommendations).

Including comprehension questions in the discussion guide or survey provides an objective way to assess participants' understanding of the treatments or associated attribute descriptions, the exercises themselves, or any associated numeric or probabilistic concepts. Such comprehension questions, often positioned before presenting the formal questions/exercise tasks, can also inform the interpretation of study results (**Section A3.3.3.4**). Additional recommendations are detailed in **Section 7** of the main PREFER recommendations.

Psychological constructs to potentially explain differences in preferences

Evaluating the psychological characteristics of patients, including personality traits, and emotional and socio-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.(49) Recommendations related to the use of psychological measures within a preference study are found in **Section 6** of the main PREFER recommendations.

Internal validity assessments

These are a reflection of the degree to which results from a preference study are trustworthy and meaningful; many methods (**Section A5**) incorporate specific internal validity assessments, the range of which is well documented.(50, 51) Addressing issues of internal validity in preference studies should be pre-specified in the protocol and statistical analysis plan (**Section A3.3.2.3**). The chosen internal validity assessments will depend on the study

objectives, the preference method, the sample size, and the length and cognitive burden of the study. Follow best practices as outlined in the literature.(51) Additionally, the PREFER case studies used approaches to assess validity of the quantitative phases of the study and are described within their respective **study reports.**

Discussion guide and survey development: question and exercise development

Developing questions or exercises requires multiple considerations, and this can be supported by including relevant experts in the development process (**Section A3.3.1.1**; **Section A4**). The approach taken should be documented and a rationale should be provided on how the design answers the research question. Further, as with many aspects of study design, method-specific considerations are necessary (**Section A5**). Overall, however, when developing such questions or exercises, best practices for qualitative (41) and quantitative (27, 29) studies should be followed and include:

- avoid framing bias as much as possible

- avoid 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 making a decision based on the level of that single attribute only)

- for qualitative studies, be appropriately open-ended and unbiased

- for quantitative studies, use the appropriate number of questions 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 (**Section 5** of the main PREFER recommendations and **Section A5**).

Finalising preference question design: 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, health literacy, numeracy) length of the interaction/survey, and medium of conducting the study (e.g. computer-based, face-to-face). For example, a computer-based only format may impede the ability of the intended participants to complete the study, as shown in the NMD case study, **protocol** and report. Assessment of the study materials (see below) in an iterative fashion through face-to-face interviews is critical for verifying the appropriate design is in place to enable successful data collection.

Finalising preference question design: assessment of study materials

It is recommended that study materials are assessed by interaction with patients prior to full study initiation – this is 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, if applicable, whether levels are sufficient to induce trade-offs. This assessment also provides insight into the amount of cognitive burden on participants. Depending on the complexity of the study objectives or context, a subsequent pilot in a small portion of the full sample may be desirable (see **Section A3.3.3.1**).

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; for example, resulting from descriptions of outcomes/harms of the medical product or disease. At the preference instrument design stage, the risk of this harm can be minimised 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. Such information can be used to guide changes to the study materials or to inform the development of distress protocols or other mitigation efforts.

Finalising preference question design: translation of study materials

For patient preference study materials that require translation, it is recommended that the ISPOR Principles of Good Practice for translation of patient-facing material (52) is followed. Initial translations could be made by suitably translation companies qualified (e.g. International Organization for Standardisation certification). However, having patient review of these translations would be necessary for some study components (e.g. interview questions) and helpful for some other study materials, to affirm the translated versions are understandable to patients in their local language.

A3.3.3   Framework component 2, conduct

During the 'conduct' stage of a preference study (**Figure A3-9**), teams should continue to apply the principles of ethics and good practice as outlined in **Section A3.3.2.1**.
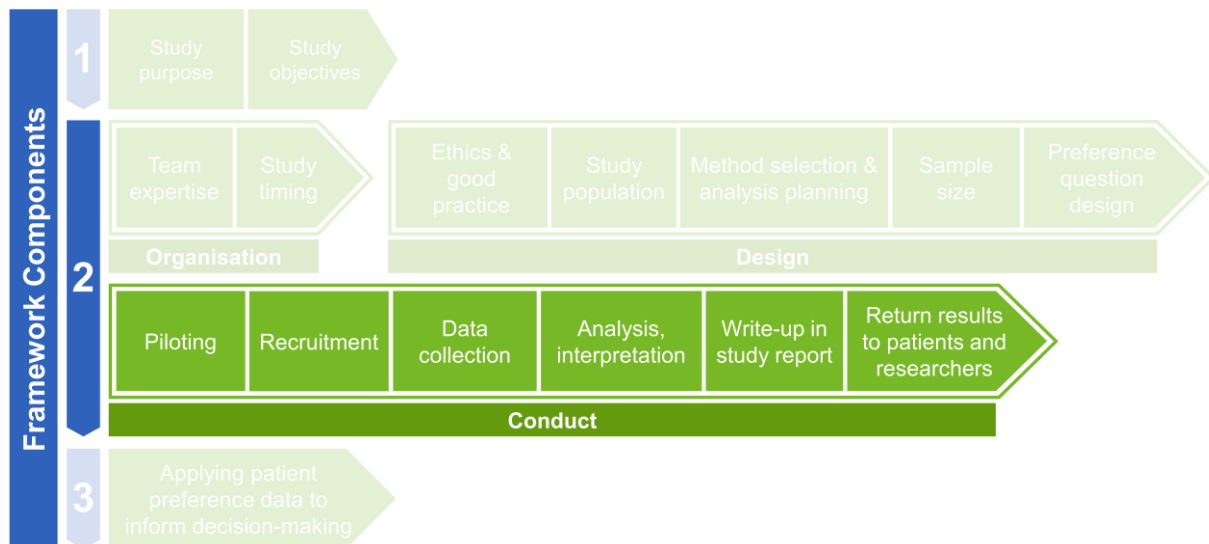
Figure A3-9. Stages of PREFER framework component 2: conduct.

A3.3.3.1   Conduct: piloting

In 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' (**Section A3.3.2.5**), which is good practice across both qualitative and quantitative studies.

Before launching the quantitative survey into the full sample, piloting and associated assessments can be performed to test if the programming and encoding of the survey instrument was correct (results as expected), to assess the validity and reliability of the preference 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 final survey into the full sample population. Additionally, piloting can be used to obtain priors for optimising the experimental design. Examples from PREFER case studies are shown in **Table A3-5**.

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

Table A3-5. Examples of how case studies updated a questionnaire based on piloting with patients.

PREFER study	Information about the pre-testing and impact on the final questionnaire	Details
Prospective study of preferences for patients with history of myocardial infarction	Objective of the quantitative pilot interview: number of patients included – 40 chronic patients with myocardial infarction	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.
	Outcome of the quantitative pilot interviews: confirmation that the design was inducing trade-off behaviour; update to the main survey to address unexpected responses on one parameter	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 myocardial infarction) where patients valued risk reduction in cardiovascular disease, ischaemic stroke and intracranial haemorrhage 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 myocardial infarction, the DCE experimental design was updated."
Prospective rheumatoid arthritis study	Objective of the quantitative pilot: to inform the final experimental design and optimize statistical efficiency.	A survey pilot was conducted with 100 members of the general public in the UK. Previous information on the importance of the attributes was be based on previous literature and best guesses for a pilot study and outcomes of initial analysis (conditional logit) of pilot data for the main survey.

A3.3.3.2 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, is described in **Section A3.3.2.2**). Below are some higher-level considerations when planning for participant recruitment. Additionally, details on the approach taken and learnings from recruiting activities can be found in reports of the **PREFER case studies**.

Operational considerations:

- time taken to recruit patients (which can often be longer than anticipated), especially when recruiting patients with a rare disease and/or patients meeting complex inclusion and exclusion criteria

- for preference studies that include the use of online surveys – ensuring that the survey links are secure and only accessible to the intended participants

- costs associated with recruitment activities and/or reimbursement of patient expenses

- with a view to reporting study results, track the number of participants who were invited to participate and who agreed/were eligible to participate.

A3.3.3.3 Conduct: data collection

All preference studies must include a data management plan describing how the data from the preference study will be collected, protected, and stored. A clear data collection plan is critical to the quality and success of the study; as such, PREFER has developed a general **data management template** for use. While many details need to be considered for successful and quality data collection, some general considerations include:

- Develop collection plans that address elements in the preference survey and/or discussion guides (**Section A3.3.2.5**), and relevant patient characteristics, such as information about demographic, psychological, and disease characteristics (e.g. disease duration, severity).

- Data collected 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) and described in the protocol or associated document (e.g. data collection plan).

- Final 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 randomisation of research questions or multiple blocks/sets of questions.

A3.3.3.4 Conduct: analysis, interpretation

The analyses conducted in a preference study are primarily those specified within the statistical analysis plan and/or protocol (**Section A3.3.2.3**). However, if unplanned analyses are needed to better understand potentially unexpected results, one should carefully explain and describe the implications of these additional analyses or deviations from planned analyses when interpreting and reporting the preference study results.(34)

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:

- as appropriate, topics of heterogeneity identified within study objectives and planned analyses would be described (**Sections A3.2.2** and **A3.3.2.3**, respectively); these include, for example, how certain observable characteristics could impact preference outcomes and/or latent class analyses

- having input from patients or patient representatives in this phase may lead to meaningful interpretations; if involved, describe how patient research partners informed the interpretation of study results (**Section 4.1** of the main PREFER recommendations).

For quantitative studies, other considerations may include:

- understanding whether participants responded in the anticipated way – this could include formal responses to comprehension questions (e.g. testing the interpretation of displays of probability;(53) or results from validity assessments (e.g. choice consistency and face validity; see 'Discussion Guide and Survey Development' in **Section A3.3.2.5**); and whether this may impact or inform interpretation of results.

- assessing and discussing 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 modelling, if required for the method selected) – hold true once data are collected.(27, 29)

- interpreting findings related to internal validity assessments is complex and research continues into how best to do this (see **Section A5** and **Section 8.3.3** for methodological considerations). Apparent violations of internal validity can be the result of behavioural considerations such as learning, fatigue, or simplifying heuristics used by patients to minimise 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:

167

- Subjects may infer additional information beyond what is presented in the study materials or use heuristics such as assuming higher costs reflects higher quality.(29) (54)

- in choice-based studies, such as those based on DCE or BWS case 2 or 3, a patient's response pattern may indicate a lack of attention to the task; that same pattern 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 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.

A3.3.3.5 Conduct: write-up in study report

Considerations when reporting study results

PREFER developed a **study report template** to assist with creating a study report. Recommendations when writing a study report (for both qualitative and quantitative studies) are as follows:

- include an explanation of the patient preference study purpose that covers the decisions and decision-makers, the type of preference-sensitive situation, and whose preferences are of interest

- include a description of the study objectives

- include an appropriate description and rationale for the method, instruments, questionnaire design, and experimental design used – this should include information on how the instrument was developed, tested, and revised

- describe the number of people invited to participate in the study, the number who agreed to participate, and the number who completed the study

- in the presentation of results include a description of the sample population, a description of missing data and/or impact on interpretation of results, and summaries of completed analyses; if unplanned analyses were conducted and presented, the rationale for these should be clearly summarised

- in the discussion of results include a transparent consideration of study limitations (including potential bias) and applications (including generalisability and/or representativeness), if applicable

- include conclusions on what was learned from the study and future implications of the research.

A3.3.3.6 Conduct: return results to patients and researchers

The final steps when conducting a patient preference study are the return of results to study participants, and publication in a peer-reviewed journal to make the preference study results available to other researchers.

Returning results to patients

Returning results to patients should be done in plain language in accordance with best practices on plain language.(55-58) These plain language principles are summarised in the PREFER **plain language summary template***.*

Additional recommendations on returning results to participants include:

- before study initiation, plan how patient friendly summaries will be provided back to participants at the end of the study; this then allows advance communication to study participants where they can access final study results upon study completion

- make every effort to collaborate with patient research partners and include translation steps of the plain language summary into the participants' native language.

Making preference study results available to researchers

PREFER recommends that preference study teams should make every effort to publish the study results; guidelines on reporting practices have been published relevant to qualitative (41, 42) and quantitative studies.(27)

Additionally, if the study has been entered into a registry, results should be provided in that forum as per the registry's guidance.

A3.4 PREFER framework component 3: applying preference data to inform medical product decision-making
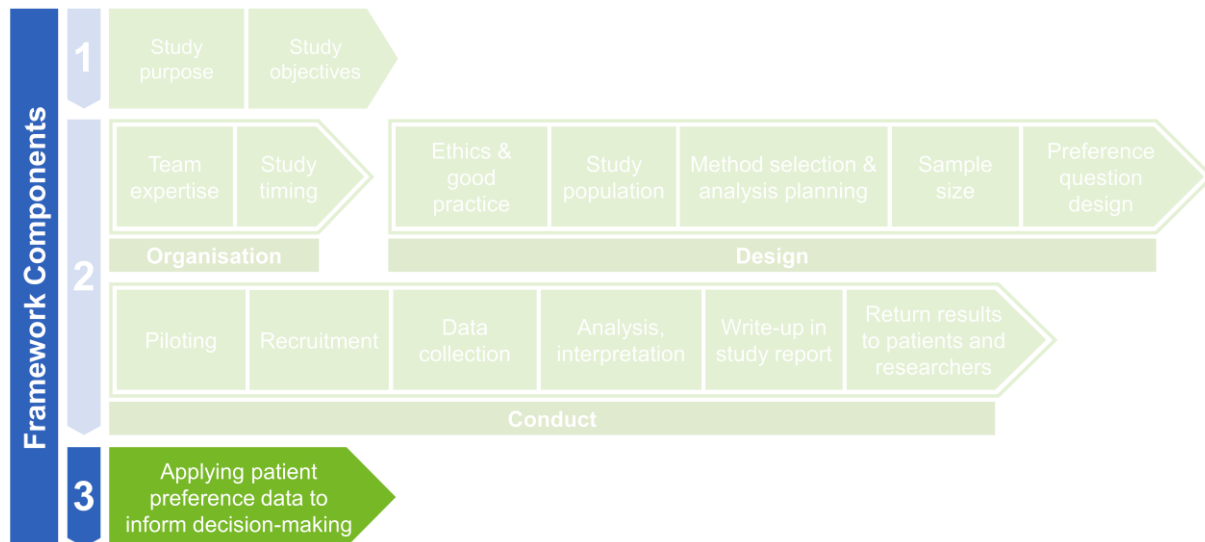
Figure A3-10. Stages of PREFER framework component 3.

A preference study is conducted to inform a decision, with the study purpose specifying the decision-maker and the decision being informed. 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** of the main PREFER recommendations. This section is not intended to be a complete overview of all possible approaches or methodological details.

The discussion in this section includes three sub-sections. **Section A3.4.1** describes common ways in which preference data can be applied to inform medical product decisions. **Section A3.4.2** describes the technical methods behind these applications. **Section A3.4.3** 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 [EPAR], 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.

A3.4.1 Applications of preference data to inform medical product decision making

There are many potential applications of preference data to inform medical product decision-making.(30) This section discusses various potential applications consistent with the preference-sensitive decisions described in **Section 2.1** of the main PREFER recommendations, with cross-references to the more technical aspects that are discussed in **Section A 3.4.2**.

A3.4.1.1 *Industry and regulatory decisions* about choice of patient-relevant endpoints, HTA decisions about endpoints to consider in REAs

The types of preference study purposes discussed in this section are:

- **an industry decision** about the choice of patient-relevant endpoints for development studies and/or interest in understanding the importance to patients of a non-health benefit (e.g. a more convenient mode of administration). This might be prompted by:

  - a lack of established endpoints to study in this indication (e.g. medical product development for an indication without approved treatments)

  - an interest in re-assessing established 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)

  - an interest in understanding the value to patients of a non-health benefit (e.g. the importance to patients of an oral formulation vs. an injection).

- **a regulatory decision** about the choice of patient-relevant endpoints when making a benefit–risk assessment

- **an HTA/payer decision** about which patient-relevant endpoints to consider when making an REA.

All these decisions can potentially be informed by a single patient preference study, as shown in **Figure 2-1** of the main PREFER recommendations.

Application of preference data to inform these decisions

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

- In a situation where there are no established endpoints for an indication, a typical approach would be to base the decision on the preference weights (i.e. preference data in isolation – **Section A3.4.2.1**). This approach can also be used to understand the importance to patients of a non-health benefit such as convenience or mode of administration.

- In a situation where the preference study is used to re-assess presumed 'established' endpoints for an indication, one approach could be the use of choice share information (**Section A3.4.2.3**) 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 regulators, and between industry and HTA bodies/payers about the choice of endpoints to include in a study for submission (see also **Section A3.4.1.4** for information on industry decisions about a meaningful effect size).

After using a preference study to inform the industry choice of endpoints, the **regulatory decision** about the choice of patient-relevant endpoints would generally be driven by the endpoints used in the clinical study. As stated in the EMA Day 80 assessment report template,(59) 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 patient-relevant endpoints:(31) "Information about the patient perspective may be considered when describing the...benefits". This could include the patients' views on non-health benefits (e.g. convenience), which can be considered as key benefits.(31)

Similarly, 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 REA. For instance, patient preference data can be directly used in HTAs to:(60)

- define/confirm the relevant clinical outcomes to be included in the REA

- quantify patient preferences towards non-clinical outcomes (e.g. improved convenience of a new treatment, impact on activities of daily living).

Preferences for disease-specific health outcomes add to the information from preferences for generic outcomes in an REA (e.g. overall survival, health-related quality of life measured by means of a generic health-related quality of life [HRQoL] instrument such as the EQ-5D), 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 enables decision-makers to assess the extent to which the new treatment meets the patients' needs better than the existing treatments, leading to a more accurate and patient-centred REA. It is a normative choice the HTA body needs to make about whether disease-specific information is considered relevant or not for the decision-making.

A3.4.1.2 How a preference study about the acceptability of trade-offs between treatment characteristics can inform industry, regulatory and HTA/payer decisions

The types of preference study purpose discussed in this section are:

- **industry decisions** about the acceptability of trade-offs between treatment characteristics to inform submission decisions

- **regulatory decisions** about the acceptability of trade-offs between treatment characteristics to inform initial marketing authorisation, follow-up indications, and line extension decisions

- **HTA/payer decisions** about the hypothetical uptake of a new treatment to inform budget impact calculations and organisational decisions.

As shown in **Figure 2.1**, all these decisions can potentially be informed by a single patient preference study.

Of note, the decision context for a regulatory decision could be different to the context for an HTA/payer decision; for example, the regulatory decision might rely on a different assessment of the treatment landscape to the HTA body/payer.

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

- **industry decisions** about the acceptability of trade-offs between treatment characteristics to inform development decisions

- **regulatory decisions** about the acceptability of trade-offs between treatment characteristics post-authorisation, in the event of a post-marketing safety signal that prompts a renewed evaluation of a product's benefit–risk profile

- **HTA/payer decisions** about reimbursement revisions when the treatment landscape changed compared with the time when the initial reimbursement request was submitted.

These types of decisions would typically be supported by a specific preference study, the timing of which would be such that the study results are available to inform the relevant decision (**Section A3.3.1.2**).

Application of preference data to inform these decisions

An **industry decision** about submission **and a regulatory decision** about the acceptability of trade-offs between treatment characteristics at the time of submission could be informed by:

- data displays combining both preference and clinical data (**Section A3.4.2.2**), *and/or*

- a side-by-side approach to preference and clinical data to support discussions on MAR for a specific level of benefit / MRB for a specific level of risk (**Section A3.4.2.2**) – this approach is only applicable 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 or multi-criteria decision analysis (MCDA) (**Section A3.4.2.3**).

An **HTA body/payer assessment** about the hypothetical uptake of a new treatment could be informed by the choice share (**Section A3.4.2.3**). Of note, the decision context relevant to the HTA decision could be different to the decision context relevant to the regulatory decision (**Section A3.2.1**). Differing decision contexts could mean that the clinical data appropriate to the choice share estimate for an HTA body/payer decision could differ to that for a regulator. For example, the choice share estimate for an HTA body could be based on clinical data associated with the new medical product vs. standard of care A, whereas the choice share estimate for a regulator could be based on clinical data associated with the new medical product vs. standard of care B. However, both estimates can be based on the same preference study provided that the relevant features of all relevant treatments (standard of care A, standard of care B, and new treatment) are included in the study.

An **industry decision** about the acceptability of trade-offs between treatment characteristics 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 (**Section A3.4.2.2**).

A **regulatory decision** about the acceptability of trade-offs between treatment characteristics post-marketing, in the event of a post-approval safety signal, could be informed by the same technical approaches as described for a regulatory decision about marketing authorisation.

HTA/payer decisions about reimbursement revisions require information on the REA of all treatments for a specific indication and could be informed by patient preference studies covering relevant characteristics of all treatments. Data displays covering both clinical and preference data as well as information on choice share will help in the assessment, potentially leading to reimbursement revisions.

A3.4.1.3   Direct elicitation of patients' preferences to support cost-effectiveness analyses

The types of preference study purposes discussed in this section are:

- a **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 (e.g. EQ-5D) could be used in a patient preference study to enable 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. The counter-argument is that patients might have very different preferences for generic HrQoL attributes depending on their condition, and hence QALYs calculated using patient preferences may not be comparable across patient groups. 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.(61) 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 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. Time trade-off technique and standard gamble 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.(62) 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.(63) The discussion of the appropriateness of one approach or another to cost-effectiveness analyses is beyond the scope of PREFER.

Application of preference data to inform these decisions

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 (**Section A3.4.2.3**). An example is shown in **Table A3-6**. 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 A3-6. Example of how a preference study informed cost-effectiveness analysis.

Study	Mühlbacher & Sadler 2017 and Mühlbacher *et al* 2017
How the study used preferences to inform cost-effectiveness analysis	The DCE on patient preferences for antiviral therapy of chronic hepatitis C allowed the derivation (by weighting of multiple outcomes of the antiviral therapy of the disease) 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 benefit was derived.
Outcome	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 patient preferences and available clinical data.

A3.4.1.4   Industry decisions about meaningful effect size

The type of preference study purpose discussed in this section:

- An **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 PROM instrument.

Application of preference data to inform 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 A3.4.2.1**). 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.

176

A3.4.1.5 PRO instrument scoring

The type of preference study purpose discussed in this section:

- An **industry decision** 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 (**Section A3.4.1.4**).

Patient preference information can also be associated with PROs for further reasons. 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 (see further discussion of this point in **Section 8** of the main PREFER recommendations). Using preference data to inform decisions about the acceptability of trade-offs between treatment characteristics 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 in **Sections A3.4.1.1**, **A3.4.1.2** and **A3.4.1.4**, respectively.

Application of preference data to inform these decisions

Typically, this activity will be based on the use of preference data in isolation (**Section A3.4.2.1**).

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 A3-7** 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. For example, 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.(64, 65)

Table A3-7. Example of how a preference study could inform a decision about PRO instrument scoring.

Study	As described by Johnson *et al*,(65) 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).
How the study used preferences to inform a decision about PRO instrument scoring	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.

A3.4.1.6 How a preference study about the acceptability of uncertainty can inform industry, regulatory and HTA/payer decisions

The type of preference study purpose discussed in this section:

- an **industry decision** about the acceptability of uncertainty, to inform submission decisions

- a **regulatory approval decision** relating to the acceptability of uncertainty

- a **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 2-2** of the main PREFER recommendations).

An 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. This decision takes different elements into account, including the incremental cost of the new treatment, the budget impact, and the possible organisational, ethical, and legal impact of (temporarily) reimbursing the treatment for which the evidence is insufficiently strong to make a definitive reimbursement decision. In other words, HTA bodies and payers will make their own assessment of the acceptability of uncertainty from a societal point of view. Nevertheless, a preference study about the acceptability of uncertainty from patients' point of view can be useful to assess the hypothetical uptake of a new treatment, such as to inform budget considerations in the context of temporary reimbursement (i.e. coverage with evidence development).

Two potential sources of uncertainty are:

- **statistical uncertainty**, e.g. wide confidence intervals. This might apply to estimates of efficacy and safety for a medical product developed for an orphan indication, where relatively few patients were included in the clinical trials for the product.

- **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, such as when choosing between drug A with moderate efficacy, tolerable safety and 5 years of post-approval use vs. drug B with good efficacy, tolerable safety, and no post-approval use.

Patients' views on either type of uncertainty can be assessed via a preference study that includes 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* (66) 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* (67) 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 *et al* (68) included an attribute of 'how long the medication has been studied', with levels of 1, 3 and 6 years. Hauber *et al* (69) 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: 'the problem can be treated successfully' and 'the problem cannot be treated successfully').

Application of preference data to inform 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 (NCB) or MCDA/stochastic multicriteria acceptability analysis (SMAA) based on the mathematical combination of preference and clinical data (**Section A3.4.2.3**).

An 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 A3.4.2.3**).

A3.4.2   Technical methods for the application of preference data

Section A3.4.1 described common ways in which preference data can be applied to inform medical product decisions. This section describes the technical methods used for 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 A3.4.2.1**)
- preference data applied in parallel with clinical or other data (**Section A3.4.2.2**)
- preference data mathematically combined with clinical or other data (**Section A3.4.2.3**).

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 benefit–risk assessments during marketing authorisation, or HTA and reimbursement decisions – preference data, 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.

A3.4.2.1   Preference data applied in isolation

There are 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 MAR / MRB.

Preference weights

These are defined as quantitative measures of the relative importance of the attributes included in the study.(26) 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 **Table A3-8**, 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.(8) 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 standard gamble, the weights are in 'normalised 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 normalised 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, the importance of a change from 0% risk of stroke to 5% risk is intuitively 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 A3.4.2.3**, value functions are assessed in MCDA to explicitly characterise this non-linearity. In methods like DCE and PTT, the part worths implicitly combine both the weight and value functions. In the example by Ho *et al*,(8) 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 A3-8**). 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 A3-8. Estimate of preference scores by attributes and levels. *Adapted from Ho et al.*(8)

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 (0.15)
	0 months	Reference level
Side-effect duration	0 months	– 1.0 (0.11)
	1 months	– 2.0 (0.09)
	12 months	– 3.2 (0.31)
	60 months	Reference level
Chance of side-effects requiring hospitalisation	None	Reference level
	5% chance, no surgery	– 0.2 (0.39)
	20% chance, no surgery	– 0.5 (0.35)
	5% chance, with surgery	– 0.6 (0.36)
Dietary restrictions	Eat ¼ cup of food at a time	Reference level
	Wait 4h 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 [a]	Eliminate need/risk	+3.2 (0.37)
	50% dose/risk	+2.2 (-.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)

[a] *Average reduction in dose of prescription drugs for the current primary comorbid condition or chance of getting the most feared comorbid condition at the lower weight.*

Clinically meaningful effect size

For novel endpoints, there may be little experience from which to base a minimum difference in that endpoint that is clinically meaningful. Preference studies may allow 'borrowing' information on effect sizes from well-characterised 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 MAR or MRB (**Figure A3-11**). 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 A3.4.2.3**).
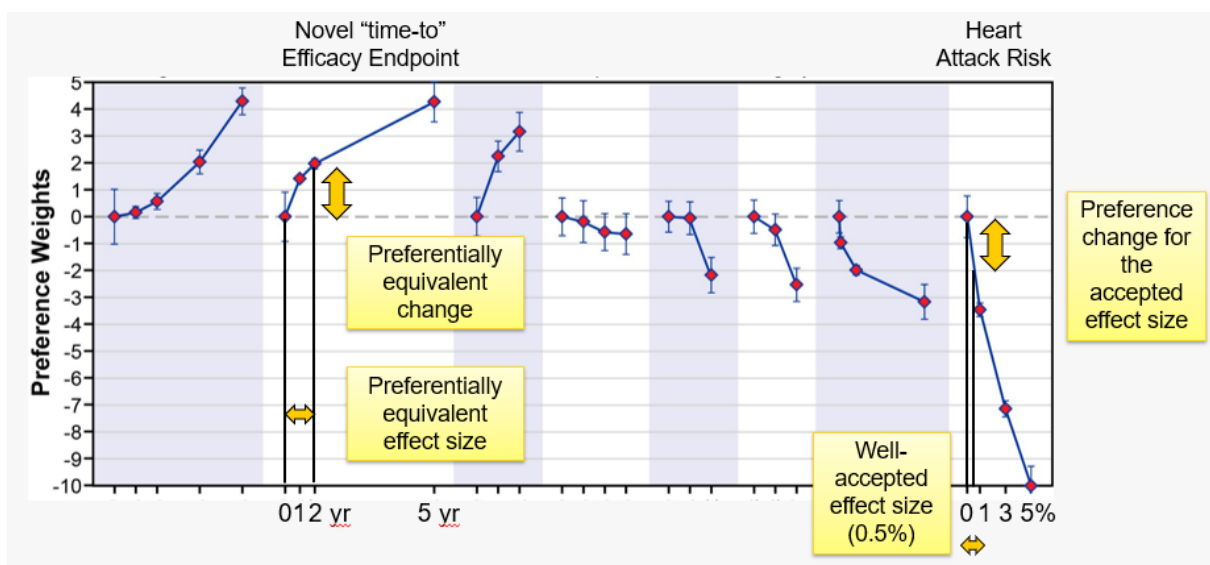
Figure A3-11. Mock example of using a preference study to develop an effect size for a novel endpoint. *Adapted from Ho et al, 2015.*(8)

Maximum acceptable risk / minimum required benefit

Maximum acceptable risk is defined as "the greatest increase in probability or magnitude of a harm a patient would accept to achieve or realise a given benefit" (adopted from the

definition in the MDIC 2015 report).(26) 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.(8, 50)

The general approach to calculating a MAR is similar to the approach described in **Figure A3-11** 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 MAR 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 A3-12**).(70)
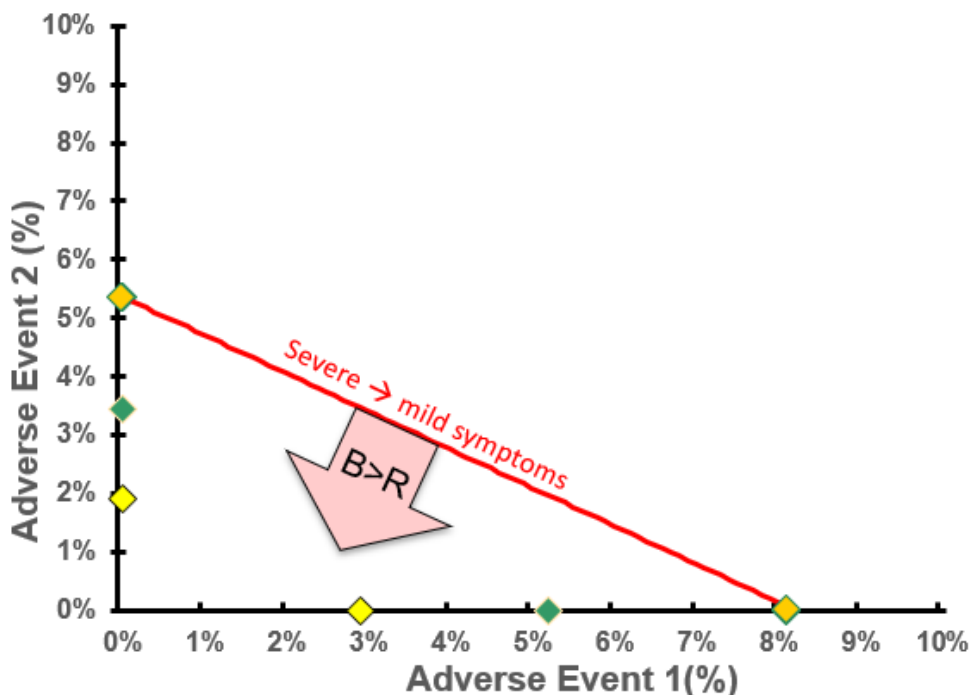
Figure A3-12. Example of maximum acceptable combination of risks for a given level of benefit. *Adapted from Fairchild et al, 2020.(70)*

Figure A3-12 shows that, for a treatment that reduces severe symptoms to mild symptoms, the MAR for adverse event 1 is 8.2%. The MAR 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 A3-12**, the red line shows the joint MAR curve for adverse events 1 and 2; benefits exceed 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 it can take on more complex forms. 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 benefits exceed the joint adverse event incidences.

Minimum required benefit 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 the MDIC 2015 report).(26) 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 MAR

Results of a preference study can be expressed as a MAR for a specific treatment benefit. For example, in a study of preferences of patients with multiple sclerosis,(71) 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 MAR 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 an MRB, 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.

A3.4.2.2 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 visualisations, followed by an integrated interpretation of both the clinical and preference 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 displays, such as:

  – effects tables or other tables in which clinical endpoints and preference data are shown in separate columns (**Figure A3-13**)

  – forest plots, in which the endpoints are shown in ranked order of decreasing or increasing preference weight (**Figure A3-14**)

  – forest plots, in which the location of the bars depicting the endpoints are shown at locations proportional to their preference weight (**Figure A3-15**)

  – parallel bar plots, in which one plot depicts some measure of between-treatment differences and an adjacent plot depicts preference weights (**Figure A3-16**).

  This type of approach is sometimes referred to as deliberative or qualitative MCDA, in which the decision-maker mentally integrates the clinical data and preference data, rather than a quantitative integration with MCDA (see **Section A3.4.2.3** for a description of quantitative MCDA).

- Comparison of a new medical product's effect on risks and benefits to the MAR / 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 A3-13** 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.(59)

	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

Figure A3-13. Example table showing clinical data side-by-side with preference weights (mock data and weights).

A different approach is a figure showing clinical data ordered by patients' importance weights. **Figure A3-14** shows data for the rate differences of benefits and harms of a 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,(72) 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.
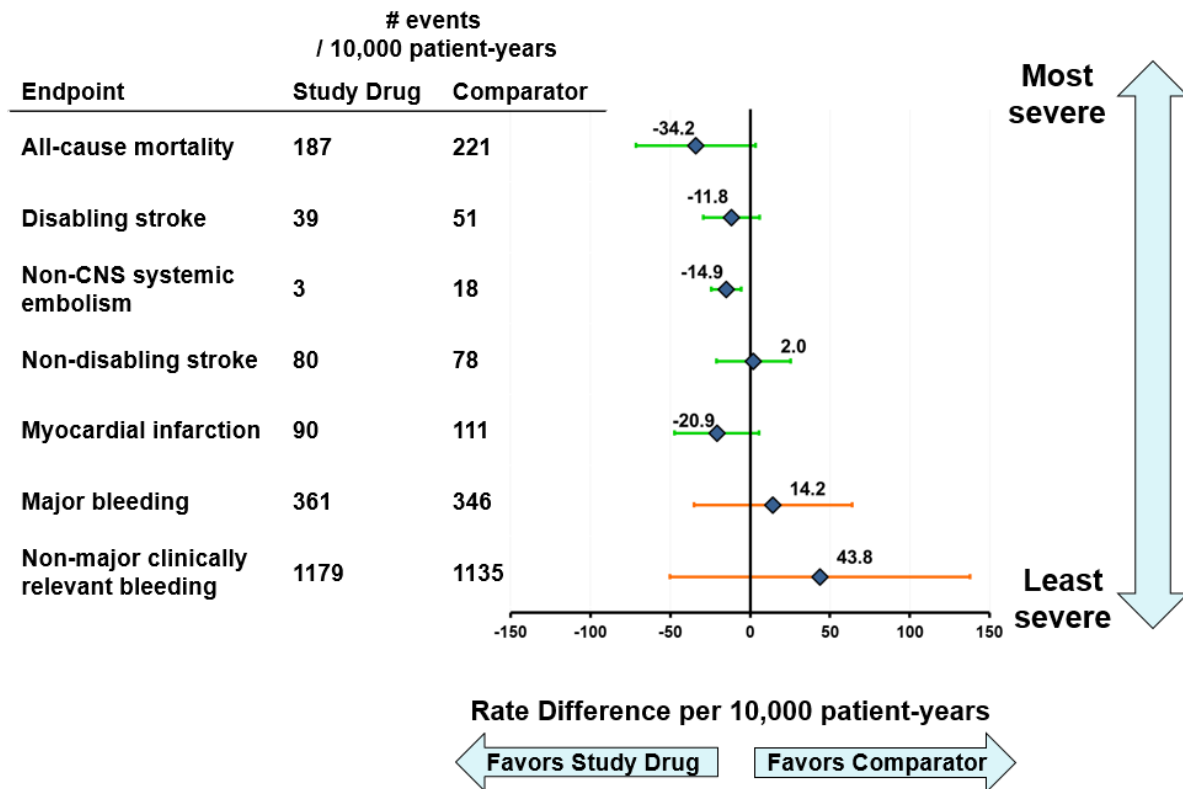
Figure A3-14. Example of a forest plot with endpoints in order of decreasing health state utility.

Figure A3-15 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 A3-14**. 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 A3-14** hides the sizes of the difference in preferences between endpoints, it makes for a much simpler figure than the proportional positioning approach in **Figure A3-15**. 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 A3-15**, 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 visualisation approach best serves the needs of the decision-makers.
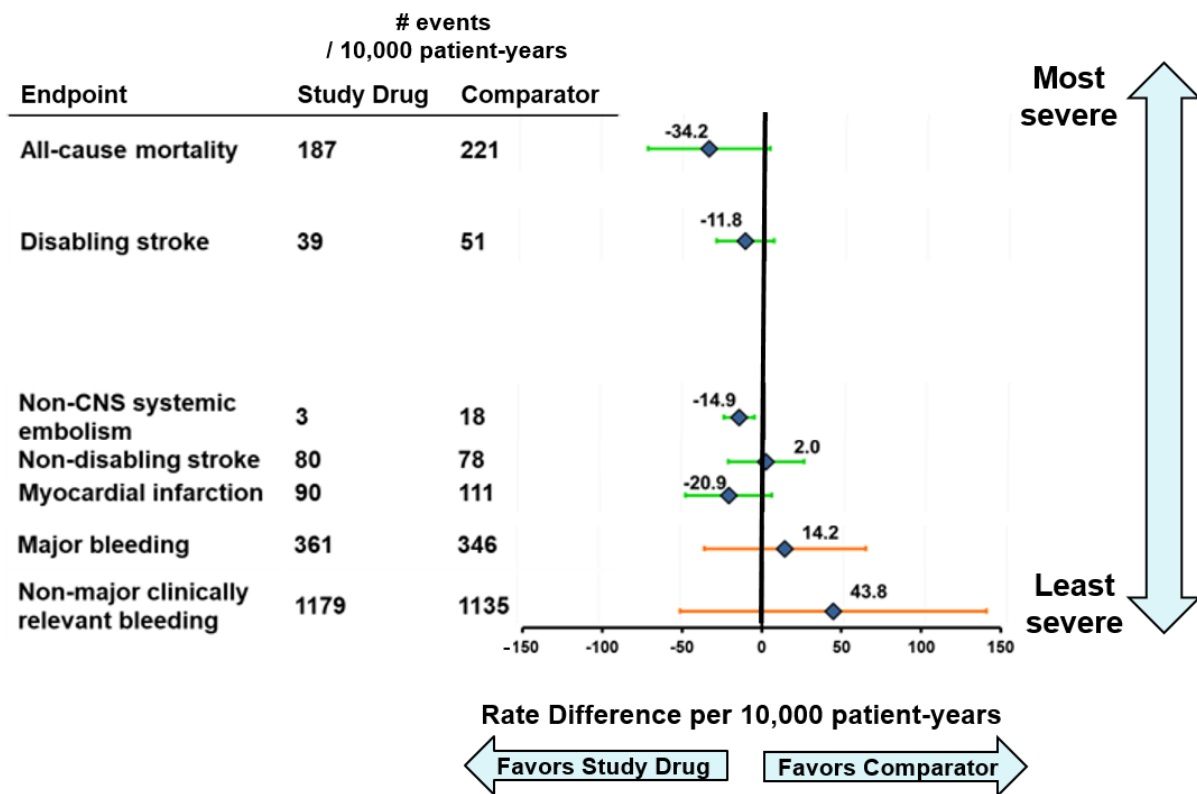
Figure A3-15. Example figure showing location of bars at locations proportional to their preference weights.

It is also possible to combine the figure and effects table approaches. **Figure A3-16** 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).

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)		1.000
Disabling stroke	38.5	50.2	-11.8 (-29.31, 5.81)		0.811
Non-CNS systemic embolism	4.5	19.4	-14.9 (-24.29, -5.53)		0.420
Non-disabling stroke	78.8	76.7	2.0 (-21.16, 25.19)		0.359
Myocardial infarction	90.8	111.7	-20.9 (-47.33, 5.59)		0.275
Major bleeding	359.6	345.4	14.2 (-35.30, 63.73)		0.267
Non-major clin relevant bleeding	1180.4	1136.6	43.8 (-50.23, 137.77)		0.159

Figure A3-16. 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 medical product and red bars favouring the comparator. The green bars reflect preference weight magnitude, with longer bars showing those endpoints are of greater importance to patients.

A3.4.2.3 Preference data combined mathematically with clinical or other data

Approaches that mathematically combine preference data 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
- NRB
- MCDA/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 among treatments and can be used to assess whether a substantial portion of the population for a treatment would choose to use that treatment (see example of use of choice share in Section A4.2.2.3 of this NICE assessment),(73) while market share is generally used for commercial decisions and some HTA/payer 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.(8)

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. Of the five methods addressed in **Section A5**, 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 question.

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.

NCB

There are a variety of metrics that could be classified as a type of NCB. Here, NCB is defined as a weighted sum of between-treatment differences in two or more dichotomous endpoints.(74-76) 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, although these are not covered here.(77, 78) 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; that is, 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% CIs 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 in which 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. Visualisations (e.g. tornado plot) are typically used. This is similar to the weight sensitivity analysis conducted in MCDA.

- 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.

MCDA / SMAA

Multi-criteria decision analysis 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.(79, 80) The focus here is on the patient-oriented application of MCDA. Stochastic Multi-Criteria Acceptability Analysis 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 normalised (e.g. 0 to 1) scale, and weights that are applied to these normalised 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.(81) 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.(82) 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.(83) (66

- **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

QALYs are frequently used as an effectiveness measure in cost-utility analyses. They 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.

A3.4.3 Incorporating preference information into industry, regulatory, and HTA/payer documents

PREFER suggests 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 product 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. In the Qualification Opinion, the CHMP agrees that "In principle, information on PPS may be included in the Clinical Overview or the EPAR and other relevant documents. This would pertain to cases for which the information was either relevant to the regulatory decision and the benefit-risk assessment, and/or where PPS data are relevant to inform prescribers and users of the medicinal product. The decision will be made on a case-by-case basis." (See also a note in **Section A3.4.3.3** about the Qualification Opinion view about incorporating preference data in labelling information.)

Most of the proposed approaches align with current processes documented either in ICH M4(R2),(31) EMA Day 80 Assessment Report,(59) or FDA patient preference and benefit–risk guidance documents.(31, 84, 85) Suggestions related to placement within labelling reflect the approach taken in the FDA Patient Preference Information guidance,(28) 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 trade-off decisions.

A3.4.3.1 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:

- a preference study report could be included within Section 5.3.5.4. of the eCTD (electronic Common Technical Document)

- high-level preference study results: within the Clinical Overview (eCTD Section 2.5).

Preference study results intended to support a marketing authorisation 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 in Section 2.5.6.1 'Therapeutic context' to characterise the unmet need.

- furthermore, as described earlier in this document (**Section A3.4.1.1**), 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 Clinical Overview Section 2.5.6.2 'Benefits'

- the results of a preference study assessing patients' views on the acceptability of trade-offs between treatment characteristics, or the acceptability of uncertainty issue, could have an influence on the choice of benefits in Clinical Overview Section 2.5.6.2 'Benefits' and the choice of risks in Clinical Overview Section 2.5.6.3 'Risks', as well as being included in Clinical Overview Section 2.5.6.4 'Benefit–risk assessment'

- a preference study report could be included: within eCTD Section 5.3.5.4 'Other study reports'.

For a descriptive assessment of trade-offs between treatment characteristics the clinical overview could include:

- a description of the importance of the observed effects, based on the results of the preference study, *and/or*

- a description of the MAR for a given level of benefit, based on the results of the preference study.

For quantitative assessments, an explicit relative weighing of trade-offs (e.g. via MCDA analysis) could be included in the clinical overview.

A3.4.3.2 Inclusion of preference study results in the 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 suggests that preference study results could be included in the appropriate section(s) of the EPAR.

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 in **Section A3.4.1.1**, 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 trade-off between treatment characteristics 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".

Before inclusion in the EPAR, 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. Three approaches for the evaluation of trade-offs between treatment characteristics are foreseen there: basic, descriptive or quantitative assessment of trade-offs. Such an assessment in the EMA Day 80 report (59) could include the three following aspects:

- a critical appraisal of the preference study

- the reviewer's view on the relative importance of the observed effects and/or view on MAR for a given level of benefit, informed by the results of the patient preference study

- the extent to which this informs the reviewers thinking on the acceptability of the benefit–risk assessment

For a quantitative benefit–risk assessment, an EMA Day 80 report (59) could include the three following aspects:

- a critical appraisal of the preference study

- the reviewer's view on the acceptability of the quantitative analysis in the Clinical Overview

- the extent to which this informs the reviewers thinking on the acceptability of the trade-offs between treatment characteristics.

A3.4.3.3   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.) In the Qualification Opinion, the CHMP notes that "the value of conveying information on group-level preferences to individual patients in relevant documents would have to be carefully considered for situations where individual choice is paramount (i.e., for prescription or administration/use). If the primary intent was to reflect and justify the decision processes considered at the time of clinical programme planning and during MAA assessment, the EPAR would appear a more appropriate place for PPS-related descriptions and/or data. As said, a final decision by CHMP would only be possible at the time of an assessment of a MAA on a case-by-case basis, taking into account the validity and robustness of the data".

Should it be appropriate to include preference results in the SmPC, this could be approached 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 (86) 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.

A3.4.3.4 Inclusion of preference study results in published information about HTA body / payer decisions

When HTA bodies/payers publish documentation relating to their decision-making, PREFER suggests that the preference study results could be included in such published documentation with explanation of how useful the research was, or not.

In particular for HTA, an additional assessment element on patient preferences could be added to the HTA Core Model (87) 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 including a specific assessment element in the clinical effectiveness domain, to allow assessors to describe how patient preference studies informed REAs in a more specific way, in addition to the relevance of patient preferences for the other HTA domains.

A3.5 The PREFER framework and methodological, operational, preference heterogeneity, and scientific integrity issues

This section describes, when applying the PREFER framework, whether specific framework elements do or don't require further consideration of preference-method-specific details (**Section A3.5.1)**, how the framework addresses operational issues (**Section A3.5.2**), an overview of how the framework addresses preference heterogeneity (**Section A3.5.3**) and how the framework supports scientific integrity and credibility of patient preference studies (**Section A3.5.4**).

A3.5.1 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 consideration of preference-method-specific details. **Table A3-9** describes which elements of the framework are fully method-independent and which elements of the framework require method-specific considerations.

Table A3-9. Which aspects of the PREFER Framework do/don't require method-specific considerations.

Framework component	Whether this framework component does or doesn't require method-specific considerations
Component 1 – preference study purpose & objectives (**Section A3.2**)	PREFER framework considerations for component 1 are preference method-independent
Component 2 – organisation, design and conduct (**Section A3.3**)	PREFER framework component 2 includes some steps which do not require any method-specific considerations, and some steps which do require consideration of method-specific concepts. Aspects of PREFER framework component 2 that are **preference method-independent**: **Organisation** • Team expertise • Study timing **Design** • Ethics and good practice • Sample definition **Conduct** • Piloting • Participant recruitment • Data collection • Write-up • Returning results to patient participants

Framework component	Whether this framework component does or doesn't require method-specific considerations
	Aspects of PREFER framework component 2 that require **preference method-specific** considerations:
	Design • Method selection and analysis planning • Sample size • Preference question design
	Conduct • Analysis, interpretation
	See **Section 5** for further information on preference study methods.
Component 3 – applying preference data to inform medical product decision-making (**Section A3.4**)	PREFER framework considerations for component 3 are preference method-independent.

A3.5.2   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:

- highlight operational issues which may need further consideration when working on a preference study, *and*

- 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. Some information on the approach to these operational aspects for the PREFER case studies can be found in the associated **case study reports**.

A3.5.3   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.(88) Such preference heterogeneity can be explained in that preference studies measure individuals' preferences - which are by

nature subjective – and hence it is to be expected that 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.(28, 30)

Some of these differences between individuals' preferences 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 (**Table A3-10**). 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.

Table A3-10. Example of a case study that evaluated preference heterogeneity.

Study	Way in which the case study addressed heterogeneity	Details
Patients with a history of myocardial infarction	Specific study objectives relating to heterogeneity	The primary study objective was to compare preferences for patients at 2 different stages of disease: acute (≤1 year of hospitalisation) and chronic (>1 year after hospitalisation) 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.

In the regulatory context, preference heterogeneity is typically assessed based on pre-specified subgroup analyses using observable characteristics (e.g. demographic, clinical) that can potentially be tied to a labelling claim. See further discussion of this point within the FDA CDRH guideline on voluntary submission of preference data.(28) The PREFER framework covers issues of population preference heterogeneity in several manners (**Table A3-11**).

Table A3-11. How the PREFER framework covers issues of population preference heterogeneity.

Framework section	Advice relating to heterogeneity issues
Section A3.2.2, preference study objectives	Consider the need for study objectives that investigate preference heterogeneity across the patient sample.
Section A3.3.2.3, method selection and analysis planning	Consider to what extent a preference method enables investigation of preference heterogeneity, and the *a priori* planned approach to any analyses assessing patient heterogeneity (i.e. analyses linked to the study objectives relating to heterogeneity).
Section A3.3.2.4, sample size	If a study objective relates to a specific subgroup of patients, consider the need for the study to include sufficient patients in all subgroups of interest.
Section A3.3.2.5, section on collection of baseline data	Consider the need to collect data on baseline characteristics, disease characteristics or any other characteristics of the anticipated patient population that may influence their response choices.
Section A3.3.3.4, analysis, interpretation	Consider if/how patient heterogeneity influences the interpretation of results.

Of note, the impact of heterogeneity in patients' views differs depending on the type of decision to be made (**Table A3-12**).

Table A3-12. How heterogeneity in patients' views could impact specific types of decisions.

Type of preference-sensitive decision	Impact of heterogeneity in patients' views
Understanding patients' views on the relative importance of issues relevant to their disease or treatment	Heterogeneity in patients' views could affect the degree to which a choice of patient-relevant endpoints is applicable across a population.
Acceptability of trade-offs and acceptability of uncertainty	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.

A3.5.4 How the PREFER framework supports scientific integrity and credibility of patient preference studies

One concern of stakeholders is whether results of patient preference studies are unbiased.(89) As with clinical trials,(90) 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. Areas of the PREFER framework that specifically address concerns related to scientific integrity and to the credibility of preference study results are described in **Table A3-13**.

Furthermore, the use of scientific advice is encouraged so that sponsors can discuss preference study proposals with regulators and/or HTA bodies. Such discussion is expected to help everyone gain experience with, and expertise in, acceptable approaches to patient preference studies, and can also ensure that information from preference studies will meet the needs of decision-makers. Topics that can be particularly helpful to cover in a scientific advice process are described in **Section 4.2**.

Table A3-13. Major scientific needs that are addressed by the PREFER framework.

Scientific need	High-level recommendations	Section of the framework with further discussion of this recommendation
Pre-specification	As for clinical trials, an analysis plan should be written prior to the results becoming available.	**Section A3.3.2.3**, method selection and analysis planning
Appropriate choice of method	The validity and reliability of a method should be considered when selecting a method, or ways to establish validity and reliability should be examined.	**Section A3.3.2.3**, method selection and analysis planning
Alignment between patient sample and study purpose; potential selection bias when recruiting patients into the preference study	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.	**Section A3.3.2.2**, sample definition
Appropriate choice of attributes	The choice of attributes or scenarios should take into consideration the information that is relevant to the associated decision.	**Section A3.3.2.5**, preference question design
Clear description and framing of attributes	The description and framing of attributes should be defined with as much precision and clarity as possible.	**Section A3.3.2.5**, preference question design
Handling of missing data	Approaches to missing data should be described at the analysis planning stage and at the stage of writing up preference study results.	**Section A3.3.2.3**, method selection and analysis planning **Section A3.3.3.5**, write-up
Alignment with key stakeholder(s) (e.g. regulator, HTA body) on the proposed design, conduct and analysis of the preference study.	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 can help ensure that information from preference studies will meet the needs of decision-makers.	**Section 4.2** of the main PREFER recommendations

A3.5.5 Differences in framework content presented here vs. framework content presented in the EMA qualification

Much of the framework content in this section was included in the Qualification procedure with the EMA. Differences in the content presented here versus the content presented in the EMA qualification are described in **Table A3-14**.

Table A3-14. Overview of differences between framework as presented here vs. framework as included in the EMA qualification.

Framework section	Difference between framework in this document vs. framework in the EMA qualification	Details
Component 1: preference study objectives	Difference in description of how to define preference study objectives	Based on reviewer feedback, the version in this document more clearly explains the expected link between preference study objectives and endpoints, and application of results.
Component 2: Overall	Removal of most examples	Example tables throughout were replaced with cross references to PREFER case study reports, training materials, and/or operational guidance
Component 2: design	'Sample definition' changed to 'study population'	Terminology has been changed for better alignment with standard terminology used in clinical trials.
Component 2: Preference question design	Reorganisation of content presentation and truncated text	The content was re-organised to make it more readable and in a logical flow; this included places subsections into 3 categories. Minimised text presented elsewhere in the recommendations, specific additions of cross references.
Component 2: write-up in study report	In the framework presented in the EMA qualification, this chevron was titled 'write-up'.	There is more emphasis on writing up the preference study results in a study report.
Component 2: returning results to patients and researchers	In the framework presented in the EMA qualification, this chevron covered only 'returning results to patients'.	Previous content is contained in a plain language summary template delivered as part of the operational guidance. Added a cross reference to that template and included relevant literature references from that template into this section. This section was also edited to recommend that the results be published in a peer-reviewed journal.

A4 Annex for Section 4

A4.1 Working with patients as research partners in patient preference studies

Tables A4-1 to **A4-6** describe the specific roles patients fulfilled within the various patient preference **case studies** conducted within the PREFER project; and particularly in defining the study purpose (**Table A4-2**); in formulating the questions patients are asked in patient preference studies **(Table A4-3**); in defining the study's attributes and levels (**Table A4-4**); and in the survey presentation (**Table A4-5**) and how they contributed to the different steps of the rheumatoid arthritis patient preference study (**Table A4-6**).

Table A4-1. Examples of patients as research partners in prospective PREFER case studies.

PREFER case study	Role of patient	Details/comments
Multiple myeloma	Study team member	• involved in study protocol development • mentioned as co-investigators in the ethics committee protocol • provided input into the study objectives • co-designed the recruitment procedures • helped develop the attributes, levels and their descriptions (including visuals) • helped develop the focus group guide and survey questions • pre-tested the focus group guide and survey • involved in interpretation and dissemination of study results
PAVING	Advisory board member	• advisor for protocol development
COPD	Advisors/consultants	• advisor on study design, discussion guides and survey questionnaires • involved study results interpretation

PREFER case study	Role of patient	Details/comments
Rheumatoid arthritis	Study team member	• helped decide the clinical objectives • designed the recruitment procedures, content of focus group guide and survey instrument • informed the selection and framing of attributes and levels • supported the identification of participants' educational needs and framing of the information in the educational material • informed the survey design (including images and pictograms), pre-testing, and management • involved in the analysis, interpretation, write-up, and dissemination of the study results
NMD study	Transitioned from advisory board member to study team member	• involved in study protocol development • involved study results interpretation
Lung cancer	Consultants	• involved in qualitative and quantitative study protocol development

Table A4-2. PREFER case study examples on patient involvement in defining the study purpose.

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

Table A4-3. PREFER case study examples of patient involvement in formulating questions to ask patients in patient preference studies.

PREFER case study	Problem	Context	Contributions
Multiple myeloma	Initial proposed version of the swing weighting 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 finalised focus group discussion questions and survey
Rheumatoid arthritis	Public perceptions of rheumatoid arthritis 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 due to patient partner validation of the focus group study results	Inclusion of specific examples of ways in which the early symptoms of rheumatoid arthritis might impact patient's daily activities
COPD	Suitability and patient-friendliness of the language and phraseology	The survey was reviewed by a patient network	Improvement of survey

Table A4-4. PREFER case study examples of patient involvement in defining attributes and levels.

PREFER case study	Problem	Context	Contributions
Multiple myeloma	Patients highlighted the importance of considering the timeframe in which a certain negative treatment effect would take place What levels should be included for the attribute 'expected additional life years'	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
Rheumatoid arthritis		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 A4-5. PREFER case study example of patient involvement in survey presentation.

PREFER case study	Problem	Context	Contributions
Multiple myeloma	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	The survey was shortened and simplified Visuals were combined with text, and wording was made consistent Patient research partners proposed specific text to include in the invitations and communication to patients, clarifying the study's purpose and benefits

Table A4-6. Patient involvement across the rheumatoid arthritis patient preference study.

Study phase where patient partners were involved	Specific activity where patient partner input was needed	Relevant experience/expertise of patient partners
Definition of research question and target population	• Determine research objective and define research questions based on prior experience with rheumatoid arthritis project related to biomarker development	• Insight into patient priorities for rheumatology research • Involvement in a previous research study in a related area, in particular qualitative data analysis
Selection of methods and instruments	• Development of study protocol • Development of focus group schedule • Development of survey instrument • Development of study documents for participants • Development of disease background information and communication to study participants – this step is critical to inform treatment preferences, as participants do not have direct experience of having the disease, and public perceptions of rheumatoid arthritis are often inaccurate • Ensure study procedures and survey instrument are appropriate for use in different European countries, and appropriate for members of the public to complete	• Experience of living with rheumatoid arthritis • Experience of rheumatoid arthritis treatment • Insight into public perceptions of rheumatoid arthritis • Experience of taking part in surveys • Representation from different European countries (UK, Germany, Netherlands, Sweden) • Experience of involvement in related cross-European projects (91-93)
Patient recruitment	• Ensure recruitment strategy takes into consideration the needs of the target population – first degree relatives were indirectly recruited via patients with rheumatoid arthritis, therefore recruitment procedures needed to appeal to, and be sensitive to, the needs of patients with rheumatoid arthritis	• Experience of living with rheumatoid arthritis • Experience of rheumatoid arthritis treatments

Study phase where patient partners were involved	Specific activity where patient partner input was needed	Relevant experience/expertise of patient partners
Piloting	• Participate in virtual survey pre-test session with researchers to share perspectives on the usability of the survey	• Experience of living with rheumatoid arthritis • Experience of rheumatoid arthritis treatments • Experience of completing surveys
Analysis, interpretation and dissemination of data	• Coding and interpretation of qualitative data • Identification and prioritisation of attributes for quantitative study • Development of lay summaries of study results for public dissemination	• Analysis of qualitative data • Experience of living with rheumatoid arthritis • Experience of rheumatoid arthritis treatments

A4.2 Interactions with regulators and HTA bodies on patient preference studies

This section provides an analysis of the current uptake of patient preference research by regulators and HTA bodies. The most relevant recommendations and best practices have been distilled from this analysis and can be found in the main PREFER recommendations document, while full details and evidence behind the six overall recommendations are described below.

There is alignment that patient information informs medical product development and healthcare decision-making

In a recent stakeholder survey, policymakers/regulators affirmed their role in ensuring effective patient engagement, and payers affirmed their role in including the patient voice into decision-making. Overall, expectations were raised that policymakers/regulators take a role in driving patient engagement, create a framework, provide guidelines of good practice, and connect stakeholders, while this expectation was not shared as strongly by policymakers/regulators themselves.(94)

Despite these expectations, awareness and willingness to conduct and use patient preference research as a systematic process to gather patient insights is limited. As such, stakeholders would have to first address the current limitation of available expertise in the system, engaging in capacity building. Because public financing is scarce, it is anticipated that decision-making bodies will also, in future, rely on sponsors investing in patient preference research.

Current experience with patient preferences in the regulatory/HTA environment

Regulators and HTA bodies could encourage patient preference research by developing guidance – based on PREFER recommendations – as this would create more certainty about the acceptance of the research results. Currently, there is no requirement for patient preference studies to be included in marketing authorisation applications to regulatory authorities, and the role and methodology for such studies have yet to be agreed upon.

In the last decade, the EU regulatory network has gained experience with MCDA and other methodologies in the framework of their benefit–risk project.(95) In this context, the need to gather the patient perspective to inform benefit–risk decision-making has become explicit. It has been postulated that preferences in benefit–risk decisions "should more often be quantified and communicated explicitly" to enable better decisions and improve transparency.(96, 97)

The EMA has recognised that patients are end users of approved medical products, and agreed to continue seeking their involvement in benefit–risk discussions to consider their values and preferences when making regulatory decisions. The input of patients and healthcare professionals is expected to play an increasing role in improving the way benefits and risks are communicated in the product information.(96)

The EMA is actively engaged in and supporting research in this area.(98) For example, the VALUE study, which benefited from collaboration with the UK Multiple Sclerosis Society, used novel software (MACBETH) to elicit patient preferences for different outcomes in the treatment of multiple sclerosis, and assign weightings that could be used to quantify the relative attractiveness of those outcomes. The study was able to identify factors that influenced patient preferences and willingness to risk adverse effects (notably, severity of disease and ability to walk) and this input could be used to build decision models for actual treatments. The study showed that a decision analytic technique, expressed via a tool like MACBETH and that uses a qualitative preference elicitation procedure, can be easily integrated into a pre- or post-authorisation setting and the results included in the regulatory approval process.(3, 99) Additionally, the EMA has collaborated with academia in a pilot study to gain experience on how the collection of individual preferences can inform the regulatory review.(2)

In parallel, the IMI PROTECT project has provided similar insights into benefit–risk modelling and highlighted the need for weighing uncertainties and identifying the preferred trade-offs by individual stakeholder groups. It concluded that patients and the public can provide justification for which favourable and unfavourable effects should be evaluated in a benefit–risk assessment. The results from IMI PROTECT outline that it is important that patient and public knowledge and expertise is not disregarded; hence decision-makers would be well advised to account for all sources of evidence appropriate to the decision problem, and that this includes patient knowledge as well as other forms of objective scientific data.(100)

The most prominent case study for modelling benefit–risk assessment in the regulatory arena is natalizumab, for the treatment of relapsing–remitting multiple sclerosis. Although the drug received authorisation, the benefit–risk balance was re-assessed later due to the occurrence of progressive multifocal leukoencephalopathy in some patients. Authorisation was maintained with risk minimisation measures, and a framework was subsequently developed to guide the application, reporting, and evaluation of patient and public involvement in benefit–risk assessments. Feasibility of the framework was tested with protocols designed to elicit patient and public preferences on the benefits and risks of natalizumab. This natalizumab case study provided evidence that preferences can successfully be elicited from patients and the public and used to determine the benefit–risk balance of a medicine. An early methodological and operational foundation to guide the application, reporting, and evaluation of patient preference information in the benefit–risk assessment of medicines to improve the legitimacy, transparency, and quality of regulatory decision-making is provided by Hockley.[101]

The FDA is equally involved in patient preference research and its use in benefit–risk decision-making, particularly relating to medical devices. One example is the collaboration between the FDA, Johns Hopkins University Center of Excellence in Regulatory Science and Innovation (CERSI), University of California San Francisco-Stanford CERSI, and the American Glaucoma Society, which has been established to collect patient preference information on benefits and risks for MIGS devices. Another example is the FDA approval of a 510(k) submitted by Nxstage based on patient preference research to compare the willingness of patients undergoing home haemodialysis to perform solo haemodialysis during waking hours given the benefits and risks compared with in-centre haemodialysis.[13]

Attempts and initiatives to integrate patient preference studies in decision-making have been systematically reviewed.[102] In a recent review, 20 noticeable endeavours have been identified. Most of these (n=13) were undertaken either at the European level (n=8) or in a European country (n=5), as has been the case in Germany, Finland, and the UK, with six initiatives identified in the US. Interestingly, the disparity was attributed to the difference in decision-making processes between Europe and the US. Market access pathway of new therapies in Europe is conditioned by a centralised two-step process: marketing authorisation granted at the European level by the EMA followed by country-specific market access resulting in HTA, pricing, and reimbursement negotiations. In the US, the centralised process is a one-step evaluation by the FDA followed by decentralised negotiations with individual health insurance vendors. Only one initiative has been identified in Australia in the context of the HTA process.[102]

At the European level, the INTEGRATE-HTA project, co-founded by the EU Commission, was dedicated to improving HTA methodologies to close gaps that EU regulators had identified, in building patient-centric solutions.[103] It highlighted the current shortcomings of patient preference inclusion in relation to HTA and economic evaluations (e.g. cost-effectiveness), and acknowledged that patient preference elicitation, assessment, and

integration in decision-making may also empower acceptability of health policy decisions and enhance the transparency of the decision-making processes.(103)

<div style="background-color:#d7e4bc; padding:10px;">

Recommendation 1

To promote patient preference research and investment, PREFER would welcome the development of regulatory and HTA body guidelines outlining their expectations on the design and use of patient preference studies.

</div>

Global convergence is needed because medical product development is a global activity. In the regulatory environment, such harmonisation processes have been widely developed, and considerations on patient engagement / patient preference research are envisaged at the ICH level. Specifically, an ICH reflection paper has postulated that information about patient perspectives may be considered when describing the therapeutic context, benefits, risks, and the benefit–risk assessment.(104) Such information could include descriptive information on patient attitudes and preferences, including information obtained directly from patients or indirectly from other stakeholders using qualitative, quantitative, or descriptive methods.

Such global harmonisation seems more of a challenge in the HTA environment because HTA is a national competency in most countries and not linked to the ICH process. However, EUnetHTA members and HTA bodies have been involved in research in this area, and collaboration is recommended with HTA bodies to develop such guidance at the level of ISPOR, HTA conferences or through the Cochrane Collaboration.

<div style="background-color:#d7e4bc; padding:10px;">

Recommendation 2

The alignment of stakeholder needs through global convergence – for example, at the European/EU or global/ICH level – would support progress of patient preference research and give patients a stronger voice in medical product development.

</div>

Sponsors should look for any available guidance

Regulatory guidance for the use of patient preferences is still in development. Guidance currently available include:

- ICH MC-endorsed Reflection Paper on Patient-Focused Drug Development (104)

- FDA CDRH and CBER guidance on patient preference information – '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 Labelling'.(28)

- The MDIC project report – this was established by a consortium in the US and was intended to improve the understanding of industry, the FDA, and others of how the patient perspective might be incorporated into the regulatory approval process.(26)

The PREFER recommendations were established to close this gap for medical products until more detailed guidance was available; however, until such time as more detailed guidance becomes widely available, or if such guidance is conflicting, it is strongly recommended to seek scientific advice with the relevant decision-making stakeholder. If the perspectives of different stakeholders cannot be aligned, it may be necessary to perform several studies to satisfy individual needs.

Recommendation 3

When planning a patient preference study, all available guidance by relevant regulatory and HTA bodies should be sought and considered. To complement this, and in the absence of detailed patient preference guidance, the PREFER recommendations and EMA Qualification Opinion on PREFER can be considered.

Broaden the expert base in the regulatory and HTA environment

As postulated by Bauer and König,(105) uptake of new scientific methodology into regulatory decision-making is often hampered by the need for capacity and capability building in new areas. Similar uncertainties due to unfamiliarity may exist in relation to patient preference studies.

Recommendation 4

Regulatory and HTA bodies may choose to include scientific experts in patient preference elicitation into the scientific advice process because preference studies use methodologies that differ from those used in clinical trials or observational studies and are more comparable to those used in utility studies. Protocol development advice may require experts who can assess the design and results of a preference study.

Aligning with stakeholders relevant for context of use

Many regulators and HTA bodies have structured involvement processes that are exemplified in the following for EMA, EUnetHTA and FDA scientific advice processes.

The EMA states in their guidelines that scientific advice helps to ensure that developers perform the appropriate tests and studies so that no major objections regarding the design of the tests are likely to be raised during the evaluation of the marketing authorisation

application. It also helps avoid patients taking part in studies that will not produce useful evidence.

Not all national HTA agencies in the EU have individual scientific advice processes. Where possible, regional or even cross regional advice procedures are more efficient and provide broader acceptance across jurisdictions than advice from several individual nations. For example, EUnetHTA joint scientific advice (or a follow up procedure thereof) would be preferable. The EMA and EUnetHTA also offer parallel joint scientific consultations. This enables sponsors to obtain feedback from regulators and HTA bodies on their evidence-generation for decision-making relating to marketing authorisation and reimbursement at the same time.

The FDA provides similar advice: "If you are considering collecting patient experience data, FDA encourages you to have early interactions with FDA and obtain feedback from the relevant FDA review division." Although this guidance presents methods and approaches for collecting patient experience data, it does not fully address methods for collecting and analysing COAs or patient preference information.(35) Such scientific advice can be sought in the CDER's framework of interaction meetings for the development of new medical products.(13)

Alternatively, interactions can be conducted within the framework of critical path innovation meetings (CPIMs), which involves meeting with the CDER to obtain general advice and discuss how a proposed methodology or technology might enhance medical product development. Other FDA centres participate in CPIM meetings when cross-cutting issues arise that involve a broader audience. Through this program, the FDA expects to become more familiar with prospective innovations in medical product development and broaden its regulatory perspective. The discussions and background information submitted through the CPIM are product-independent and are nonbinding for both the FDA and CPIM requestor.

The FDA encourages medical device manufacturers and other stakeholders to consult with the CDRH early on when considering patient preference studies. Interested groups should email CDRH-PPI@fda.hhs.gov and consider requesting a pre-submission meeting, which is a type of Q-Submission meeting.

Patients should be involved in the scientific advice process; for example, to ensure alignment on the need for patient preference studies in the proposed context and to discuss the general direction of the approach.

Scientific consultation should occur early and iteratively. Consultation with regulators and HTA bodies should begin as early as possible when considering a patient preference study. However, given that such involvement requires an investment of resources, the time of involvement needs to be chosen carefully. The ideal point should be once the study concept has been outlined but when there is still enough time to change the study design. If further consultation is needed (so-called 'follow-up advice' with the same body), this usually take less time.

Ideally, regulators and HTA bodies would best provide continuous input as demonstrated in the PREFER case studies, but such mechanisms do not exist in their current scientific advice processes, and it may be difficult to realise outside of the public–private partnership environment.

Medical product development requires rapid decision-making based on a continuously growing set of data. Consequently, a rapid and continuous review of development plans is critical, and agile adaption of plans is unavoidable.

Recommendation 5

A good vehicle for engagement with the regulators and HTA bodies is scientific advice. This can be provided by individual regulators such as the FDA, Committee for Human Medicinal Products/EMA and individual HTA bodies, or through future convergence mechanisms of EU joint HTA scientific advice or parallel EMA/EUnetHTA joint scientific advice.

To foster partnerships when creating patient preference studies, patients should be involved as research partners, as outlined in **Section 4.1** of the main recommendations. In line with these principles, regulators and HTA bodies could be encouraged to include patients more often into scientific advice procedures because their perspectives can complement the scientific rationale for conducting the study.

In situations where patient preference studies will be used to inform decision-making by regulators and HTA bodies, scientific advice from these stakeholders should be sought as early as possible in the study planning period. This is particularly the case for quantitative studies, whereas exploratory qualitative studies may be undertaken prior to scientific advice.

A more rapid, flexible, and preferably continuous scientific advice process at the EMA for patient preference studies is desirable to receive stakeholder input, both across regions and across decision-makers. This could be facilitated by the creation of centres of excellence at the EU member state level.

Topics for scientific advice

Regulators and HTA bodies act within regional and national frameworks and have clear objectives, beliefs, and experiences that may not easily be aligned. So far, only the CDRH and NICE have issued guidance encouraging scientific advice on patient preference study protocols. Because there is some similarity with utility study designs frequently used in health economic decisions, HTA bodies have already developed some expertise in this area.

Not all stakeholders have advanced their considerations to this level. In general, regulators would measure the robustness of study designs against the standards of confirmatory

clinical trials with standard statistical methodologies (such as the magic $p<0.05$ hurdle) being applied.

As outlined by Bauer and König,[105] familiarity with new methodologies is critical for the willingness to embrace new research and technology in decision-making.

> **Recommendation 6**
>
> High-quality guidance, in the form of a 'briefing book', which explains the methodology and context of the research question in sufficient depth and transparency is a critical prerequisite for successful scientific advice with regulators and HTA bodies. Any limitations of the patient preference study should be disclosed and contextualised.

What matters to regulators and HTA bodies has been investigated in PREFER, and the **D2.6 final report** provides the results of the stakeholder interviews. The interviewees included both regulators and HTA body representatives, along with a range of experienced members of academia and industry. 'Estimating trade-offs' and 'quantifying heterogeneity' were identified as the two most important aspects by six of the participants.

In both the late-stage phase 3 and post-marketing settings, 'establishing validity and reliability' was one of the highest ranked criteria, mostly driven by the desire for internal validity and external validity. For the post-approval setting, 'establishing heterogeneity' was also prioritised, and the most important sub-criterion was being able to quantify heterogeneity rather than calculating risk attitudes. Additionally, being able to estimate trade-offs between attributes was weighted higher than the number of attributes that can be explored or being able to estimate weights for the attributes (**D2.6 final report**).

The results show the main areas of methodological concerns that need to be addressed in scientific advice. In summary, the context of the use of the results is critical. The use cases identified in **Section A3** (1, inform on uncertainty acceptable to patients; 2, selection of endpoints for clinical trials; and 3, identify views on trade-offs to inform B/R decisions, provide information on uncertainty acceptable and views on trade-offs in the label) will require different level of reassurance by regulators and HTA bodies on the validity of the results.

History of regulatory decision-making shows that the results of quality of life (QoL) questionnaires have received increasing attention by decision-makers for approval of new cancer therapies based on feedback received from patients. Consideration of these data were first provided in the assessment reports, subsequently in the public assessment reports, and is increasingly more reflected in the approved labels (product characteristics) of the medical products as the relevance for decision on therapies has meanwhile been acknowledged. While similar progress can be expected over time for patient preference

information there are currently very few examples of inclusion of patient preference data in medical product labels.

Based on the above findings, the experience with the various case studies, and prior experience of the PREFER patient preference experts, a checklist was developed that could serve as a best practice tool for preparing scientific advice with regulators and HTA bodies (see **Box 4.3** in **Section 4.2.2** of the main body of the recommendations). It should be acknowledged that the EMA/EUnetHTA Qualification Advice and the subsequent EMA Qualification Opinion on the IMI PREFER framework has supported the project with many valuable insights into regulatory and HTA body perspectives, and contributed to development of this comprehensive best practice.

Three examples of stakeholder engagement from IMI PREFER case studies

- Important evidence for the usefulness of scientific advice has been gathered by PREFER from work with NICE on the **COPD case study**. The scientific advice clarified the stakeholder's expectations and most critical questions (such as appropriate patient population, intra-country differences, extrapolation criteria), and enabled the research question to be refined.

- Critical input was provided by HTA bodies for the **PAVING study**, in which 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.

- Relevant expert regulatory input was sought as part of the **rheumatoid arthritis study by Uppsala University**, which aimed to estimate the MAB and to explain preference heterogeneity.

A5   Annex for Section 5

A5.1   Discrete Choice Experiment (DCE)

A5.1.1   Background to DCE

The discrete choice experiment (DCE) has increasingly been used to quantify patients' preferences for health outcomes, health services, and medical treatments.(106-108) Hauber *et al* (109) has described a conceptual framework for applying these methods to benefit–risk decisions. The FDA CDRH has demonstrated how data from a DCE can be developed to quantify patient preferences in a format that can be used for regulatory approval decisions (8) and these data were used to support the regulatory approval of a device to treat obesity.(110) In addition, DCE results can be used to support HTA and payer decisions.(29, 111)

The DCE is a utility-theoretic method for eliciting preferences for medical interventions.(109) It is based on the hedonic principle that products or services are evaluated based on their attributes and how well 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 healthcare 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 (112) that allows the estimation of a unique preference parameter for each attribute level and potentially interactions among attributes.

Best-worst scaling case 3 (BWS 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 BWS 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 BWS case 3 are constructed and analysed similarly and yield similar types of results. For the purpose of this document, we treat the DCE and BWS case 3 as equivalent.

Each preference parameter indicates the relative contribution of each attribute level to the probability of choosing an alternative with that attribute level from among the set of all

possible combinations of attribute levels. McFadden 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.(113) 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 the probability that respondents will choose a profile with a given set of attribute levels when faced with a choice between this profile and a set of alternative profiles (for example, see Ho *et al*).(8)

These measures can thus inform:

- 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 (114, 115)

- the minimum level of treatment benefit patients would require to accept a given set of treatment-related risks (116)

- 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). By scaling up to the population of interest these can be considered as choice shares.

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.(117)

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.(118) This option may, for example, be 'neither', '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.(113) The conditional multinomial logit model, however, 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.(106, 108, 119) 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 characterised 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 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 patient preferences. These results can then be used to determine the extent to which the preference measures generated by the DCE differ for different subgroups within the sample.

A5.1.2 DCE design: method selection and analysis planning

A5.1.2.1 What type of research question is appropriate for a DCE?

A DCE is typically used in decisions where the alternative treatment options are characterised by multiple attributes, where some attributes favour one treatment and other attributes favour other treatments, and clinical judgment is insufficient to defensibly assess the benefits and risks amongst them. The output from a DCE can quantify the relative importance of these attributes and what 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 and colleagues.(120) A DCE was used to evaluate patient' 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 trade-offs between benefits and risks.

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 cited 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 *et al.*(121) The research question there 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 *et al.*(27)

A5.1.2.2 When is a DCE an appropriate method?

Factors that suggest a DCE would be appropriate include:

- When it is of interest to obtain patient preferences for treatments or services that are possible but not yet available. DCEs naturally enable the use of benefits and risks that are not relevant to available treatments but are critical for those under development or being considered. DCEs also enable the levels of these benefits and risks to include those expected for future treatments. For the same reason, DCEs enable benefit–risk assessments to be computed between current and future potential treatments.

- 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 patients in a population vary in terms of their preferences (i.e. there is preference heterogeneity) and it is important to identify both decision-relevant subgroups within which there is relative homogeneity of preferences, as well as the heterogeneity within these subgroups. Preference subgroups can be particularly important to identify subgroups 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 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.

- 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).

Considering these capabilities, a DCE can be considered as one of the most complete methods currently available to assess preferences (106) in the sense that it can essentially incorporate all 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.

A5.1.2.3   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, generally faster and are less cognitively burdensome on responders)

* too little is known about the specific benefits or risks in the treatment alternatives (in which case it may be a better option to use 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 of a specific individual are required.

A5.1.2.4   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 trade-offs between treatment characteristics 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.

A5.1.2.5 Points to consider regarding the analysis planning for DCE studies

- The ISPOR Task Force report (119, 122) gives 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 *et al* (119) provides a checklist of questions to consider when justifying the choice of analysis method, describing the analysis and interpreting the results.

A5.1.2.6 Expected timeline for conducting a DCE

This strongly depends on contracting between the relevant parties, how long it takes to determine the attributes and their levels (which often involves a pre-period of qualitative research, such as online and structured interviews), and how long it will take to recruit subjects and collect the DCE survey data. Th could be from a matter of weeks for simple qualitative studies to up to more than 12 months for joint qualitative/quantitative studies. Therefore, overall timelines can be highly variable for each individual study, and this 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.

A5.1.2.7 Points to consider for DCE internal validity testing

There are numerous methods that are informative on internal validity in preference studies. Some tests for validity and reliability primarily apply to DCEs and similar survey-based methods,(50) and include:

- **Stability:** Repeating a choice task and assessing whether the respondent chooses the same alternative.

- **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 generalisation of the within-set dominated-pairs test that is based on two choice tasks.(50)

- **Transitivity:** A test in which, if alternative X is preferred to alternative Y and alternative Y is preferred to alternative Z, then alternative X must be preferred to alternative Z.

- **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:** When a respondent always chooses the alternative in the same location in choice tasks, it suggests the respondent is not paying attention to the survey.

'Failing' these tests is not always a definitive indication that a survey failed. For example, the alternatives in a repeated choice task may have very similar utilities, leading 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, leading respondents to again 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 behaviour. At times, these issues can be identified in advance by good pretesting or a pilot survey, though not in all cases.

A5.1.3   DCE design: sample definitions – justifying the sample size

A5.1.3.1   Points to consider regarding the sample size of a DCE

The sample size depends on the complexity of the DCE experimental 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, and when it is 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.

Methods for constructing a DCE experimental design are given in Johnson *et al.*(112)

- If a formal sample size calculation is needed to test a particular hypothesis, the methodology given by De Bekker-Grob may be used;(123) 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, as 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 (for example, see Marshall *et al* (124) and Bridges *et al.*(120) 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 *et al* (125) give a meta-analytic review of sample sizes used in 32 DCE patient preference studies. They consider more than just hypothesis testing and, 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 size of 250 subjects is probably not untypical.

A5.1.4 DCE design: instrument design – consideration of the appropriate number of attributes and attribute levels; patient burden issues

A5.1.4.1 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 the 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

- what is important to patients.

If the DCE is to be conducted to inform a benefit–risk evaluation of a specific product, the choice of attributes must recognise 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.

A5.1.4.2 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 attribute) and types of level. There are three types of attribute level:

- numeric (e.g. time, probability) that are continuous and naturally ordered

- categorical and naturally ordered (e.g. mild, moderate, severe)

- 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 of levels, 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.

A5.1.4.3 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 (112) and several software packages to generate good experimental designs (e.g. SAS).

A5.1.4.4 Why DCE?

Soekhai *et al* (126) identified 32 patient preference qualitative (exploration) and quantitative (elicitation) methods. From this taxonomy of methods, criteria with associated weights were developed for three key stages in the medical product life cycle to characterise and appraise which methods would be most likely to meet decision-makers' needs throughout the medical product life cycle. This assessment also considered publication frequency and reported theoretical issues. These assessments identified 13 (out of the 32) elicitation and exploration methods as most likely to meet most decision-makers' needs during all stages of the medical product life cycle. 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 of DCE as appropriate for qualification among the trade-off elicitation methods is its increasing use in quantifying preferences in health research, within regulatory benefit–risk related decisions, and applicability of results to HTA and payer decision-making.

A5.2   Swing weighting

This section provides a structure to guide a preference study sponsor through key issues when designing, conducting and analysing a swing weighting preference study and provides a guide for decision-makers when assessing and using swing weighting results to inform decision-making. In addition, this section supports the discussion between industry, regulators and HTA bodies/payers about swing weighting preference studies intended to support medical product decision-making.

A5.2.1   Background to swing weighting

Swing weighting is a preference elicitation method that obtains respondents' trade-offs for changes between attributes. The trade-offs are elicited directly from individuals in a complete format, which enables the analysis of individual-level preferences. This contrasts with some other preference elicitation methods such as 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.(127) While individual-level predictions (from model estimates) can be obtained from DCEs using choice models that account for heterogeneity,(128) it does not allow for individual-level trade-off data for each individual in the sample as can be obtained from swing weighting.

The typical swing weighting 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. In the second stage, 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 (between 0 and 100) of the second highest ranked attribute swing as compared to the highest ranked swing. 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.(129-133) There are also other approaches to obtain the weights, such as *point allocation* (distributing 100 points among the swings based on desirability).(134)

The basic swing weighting procedure only captures the trade-offs respondents make between attributes. It is often paired with a scoring procedure to capture preferences for changes within attributes; that is, potential non-linearities of the partial value functions.(135) In scoring, respondents make value judgements on the incremental changes in each attribute to determine the partial value function.(130)

Swing weighting and multiple-criteria decision analysis (MCDA) are often used interchangeably, although there is a difference between the two approaches. While swing weighting refers to the elicitation of trade-offs, MCDA is a decision-making framework that

enables individuals and groups to reach a consensus by assessing multiple benefits and risks by combining judgements and data. Additionally, MCDA applies decision theory to decisions with multiple objectives, enabling the appraisal of treatments with multiple attributes and combining them into a single overall appraisal.(136) Swing weighting and data from other preference elicitation methods capturing trade-offs, such as DCE, can be used to inform MCDA assessments.(137)

Multi-criteria decision analysis methods, including swing weighting, have been widely used in the public and private sector, such as for policy decision-making in areas such as transport, education, environment (138) The original swing weighting technique, 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;(133, 139, 140) MCDA adoption has been slower in healthcare.(137, 141): a 2014 literature review of MCDA within the healthcare industry showed that only 7.3% of MCDAs conducted used swing weighting to elicit weights.(137)

A5.2.2   Use of swing weighting within the healthcare sector

Within healthcare, swing weighting has been used for a range of purposes that can have an impact from a medicines regulatory and access perspective (137, 142-144) as well as from societal and health policy perspectives (145, 146) and for eliciting patient preferences.(145, 146)

In the regulatory context, swing weighting has been used to inform quantitative benefit–risk assessments (BRAs) (97) with stakeholders such as regulators, experts, and clinicians, and the EMA has concluded that MCDA is a suitable framework for such assessments.(142-144)

In the HTA context, swing weighting 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.(129, 147, 148) For example, MCDA was used to evaluate an integrated care programme for people with multi-morbidities in eight European countries.(149) Ordinal swing weighting was used to elicit weights for the different assessment criteria from five 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 project, aimed at improving person-centred care for people with multi-morbidities, and which uses both swing weighting and DCEs to elicit weights.(150)

Swing weighting can also be used to inform public health strategy by eliciting values from expert groups.(151, 152) 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.(151) This assessment has been replicated in the EU with similar findings.(153) Similarly, an international expert panel used swing weighting to assess the different types of harm from nicotine-containing products.(154) Swing weighting has also been used to elicit benefit–risk preferences of treatments for physicians and other experts,(145, 146) which may have practical implications such as helping prescribers make objective decisions about appropriate treatments to recommend to individual patients.(146)

There has been a paucity of published studies using swing weighting 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.(127) Nevertheless, there are some studies that have applied online, swing weighting-inspired procedures.(1, 2, 155, 156)

Multi-criteria decision analysis has also been used in clinical practice as personal decision support tools to facilitate shared-decision making.(157, 158) Other than the standard approach described above, there are also variations of the swing weighting method, including imprecise swing weighting and choice-based matching. Imprecise swing weighting enables decision-makers to provide imprecise value estimates using a range estimate rather than a single point estimate to express their perceived value.(159) This approach aims to account for the behavioural biases from elicitation techniques that result in a single exact weight.(160) Choice-based matching(1, 2) also known as adaptive swing weighting,(155, 156) involves a ranking exercise followed by a thresholding approach for preference elicitation.

A5.2.3   Guidance on swing weighting

A range of guidance documents are available on swing weighting, which are predominantly focused on its use of 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.(81, 135) In this guidance, the weight and scoring of different decision criteria are discussed as one of the MCDA steps, and the swing weighting technique is mentioned as one approach to eliciting weights from stakeholders.

Tervonen and colleagues (159) have also provided guidance on how MCDA can be incorporated into a BRA; in addition to the methods and weight elicitation using swing weighting process, there is also guidance on the inclusion of imprecise/incomplete data into a BRA. Phillips (161) has also outlined an eight-step framework for constructing an MCDA

model with best practice principles for use in healthcare decisions, including the weighting of criteria; for example, by using a swing weighting technique.

Beyond the healthcare domain, the UK Department for Communities and Local Government has released an MCDA manual for the appraisal of policy options, which also explains the swing weighting process.(130)

A5.2.4 Swing weighting design: method selection and analysis planning

A5.2.4.1 What type of research question is appropriate for swing weighting?

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

A5.2.4.2 When is swing weighting an appropriate method?

There are instances where swing weighting may be more appropriate such as when it is useful for understanding individual patients' preferences. Swing weighting provides individual-level preference data that does not require any modelling, unlike DCE/BWS.(127) For example, Postmus and colleagues used swing weighting to explore the distribution of individual preferences for multiple myeloma treatments.(1) Marsh *et al* (155) also used swing weighting to conduct a patient-centred BRA of aortic stenosis treatments, while SriBhashyam and colleagues used an adapted swing weighting task to establish the MAB of treatment in exchange for the treatment risks.(156)

Swing weighting may be appropriate when it is only feasible to obtain small-to-medium sample sizes (5 to 50 respondents), such as in studies of rare diseases. (127, 156) Swing weighting may also be suitable for studies that have complex attributes that would benefit from an interviewer being present, such as attributes (and their implications) that are difficult to understand, or unfamiliar to respondents.(127)

Swing weighting may additionally be appropriate 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, swing weighting enables the construction of attributes and elicitation tasks during the workshop. 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 final attribute list and preference elicitation to be developed within one workshop session.

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

A5.2.4.3   When is swing weighting not an appropriate method?

There are situations where swing weighting would not be appropriate, such as studies where trade-offs (i.e. preference data) are not required, or studies with large sample sizes or large populations; a standard swing weighting study is resource intensive because it requires interviewer-led elicitation.(127) Swing weighting in small samples also may not be representative of patient population preferences in conditions with a large, heterogenous patient population.

Further, swing weighting may not be appropriate where there is a lack of experienced interviewers or where it is not feasible to train interviewers for the elicitation. Swing weighting requires considerable expertise because the interviewers need to understand the method, ask confirmatory questions to validate respondents' preference statements, and moderate the workshop.(162)

Swing weighting appears to be more cognitively burdensome than DCE for respondents.(163) Some participants may not be able to take part in long swing weighting workshops (e.g. those who may be critically ill) or have difficulty processing the swing weighting task or providing the relevant numerical responses (e.g. those with cognitive difficulties).(131)

Standard swing weighting typically assumes an additive value model, which is based on the assumption that the different attributes are preferentially independent, i.e. improvements in one attribute do not affect preferences for change in other attributes.(136, 164) Although swing weighting can be applied with non-additive (multilinear and multiplicative) preference models,(136, 142) they are difficult to apply in practice.

A5.2.4.4   Points to consider regarding the analysis planning for swing weighting studies

Analysis of swing weighting data is relatively simple and can be done with basic descriptive statistics.(131) It does not require complex modelling because preference parameters are directly elicited from respondents in the required format.(127) Population preferences can be estimated from swing weighting data using Dirichlet regression,(165) which also enables the evaluation of the impact of respondents' characteristics on their preferences.

The robustness of swing weighting can be assessed using various techniques such as the SMAA, which enables the quantification of uncertainty of a benefit–risk decision due to imprecise swing weight estimates.(165-168) There is also software available that specifically

supports the elicitation and analysis of swing weighting data, including ADDIS (169, 170) and HiView,(171) although simple Excel sheets are typically sufficient for swing weighting elicitation workshops.

A5.2.4.5   Expected timeline for conducting a swing weighting study

There are many factors that affect the typical timeline for a swing weighting study. These include the time for contracting between the relevant parties, 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 swing weighting 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 swing weighting studies with no qualitative research or piloting, to longer than 12 months for more extensive studies. Analysis time may also vary depending on whether analyses are conducted at the population or individual level. The variability of timelines from study to study should be considered when planning to incorporate swing weighting into the overall clinical program.

A5.2.4.6   Points to consider for swing weighting internal validity testing

Swing weighting elicitation tasks are cognitively demanding and therefore swing weighting is typically conducted with individuals in a workshop or focus-group setting, facilitated an interviewer. Some studies have also implemented online swing weighting via a survey.(149, 163, 172) It is important to explain to respondents how to respond to the swing weighting tasks prior to the main preference elicitation tasks. Respondents should also 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 allows the interviewer to gauge whether the respondents' understanding of elicitation tasks is consistent with how their responses will be used.(135) Consistency checks should also be conducted, whereby the interviewer reports back their interpretation of the respondents' responses in a different format for confirmation.(130, 161, 173)

A5.2.4.7   Points to consider for sample size

Sample size is not typically a concern in swing weighting given the small sample needed. However, a larger sample size is required if there is a desire to establish population preferences, although sample size requirements for establishing population preferences with swing weighting are likely to be significantly lower than for similarly complex DCEs.(165)

This may pose an implementation challenge given the labour-intensive nature of the preference elicitation.(127)

A5.2.4.8 Points to consider when setting up the experimental design

There are several points to consider in setting up the experimental design of a swing weighting study, such as the number of attributes, mode of elicitation, implementation of swing weighting tasks and minimising bias. While swing weighting can account for a greater number of attributes than some other preference elicitation methods, too many attributes may be cognitively burdensome for respondents.

Although swing weighting is typically conducted as interviewer-facilitated focus groups or workshops they may also be conducted as individual interviews if a group setting is not feasible. A workshop or focus-group setting facilitates knowledge sharing, enabling respondents to clarify the tasks being posed, and also facilitates discussion between respondents.(162) Some swing weighting variants have used an online survey without interviewer facilitation. Results from the PREFER case study on glucose monitoring devices for diabetes suggest that online swing weighting using the standard procedure without the best practices of interviewer facilitation and confirmatory questions is likely to generate low quality data. The experimental design should also consider whether the study objective is to obtain individual-level preference data or population-level consensus data.

Swing weighting tasks can be implemented in various ways; for example, the second phase of swing weighting elicitation can be implemented as comparisons in rank order (first ranked compared to second ranked, second ranked to third, etc), or by using the most important attribute as a comparison for all other attributes. Attributes should also be designed and selected to be 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 swing weighting exercise represent the importance of the attribute swing, the range of the swing should also be considered. Scale ranges that are too large or too small may not elicit meaningful preferences.(161) 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.

If swing weighting is combined with other techniques such as scoring to account for non-linearity, 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.(174, 175)

The potential for bias from the facilitator should also be considered. To minimise the risk of this, facilitators should understand the objectives of the exercise and be thoroughly trained on the facilitation approach.(161)

A5.2.5 Why swing weighting?

Overall, swing weighting may be more appropriate for studies where a small sample size and trade-off data is desired. Swing weighting is thought to provide the same precision of population preference estimates as a DCE with smaller sample sizes.(165) However, it has a higher responder burden (127) and risks eliciting lower quality choice data in studies where interviewer-led confirmatory checks are not carefully implemented.(176) Past studies have indicated preferences elicited with swing weighting are stable even when the elicitation process was replicated with new respondents in a different country.(151, 153)

A5.3 Best-Worst Scaling (BWS) case 1

A5.3.1 Background to BWS case 1

This section provides an overview of best-worst scaling (BWS) case 1, also known as the object case, to guide a preference researcher or reviewer through key issues that may occur when conducting a BWS case 1 preference study. The section introduces BWS and its origins in market research, considers recent reviews and methodological developments, and then provides practical considerations including a guide to when to use—and when not to use —the method for different research questions. Principles of instrument design and analysis are also presented. The overview provides guidance for decision-makers seeking to review a BWS case 1 study or to use the results to inform regulatory or payer and HTA decisions for medical products.

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.(177, 178) 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,(179) but is arguably easier for respondents than conducting the rating or ranking of options required by conjoint analysis methods. 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.(180)

In BWS case 1, a list of objects, which may be 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 sometimes called items, criteria, and/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 maximisation) and the least utility (the least preferred or worst). Typically, respondents are asked to select their most and least preferred, although variations may include best and worst, most and least important, or most agree with and least agree with from a subset. The framing of the choice question will be 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.

The subset of objects presented in the choice tasks are created using an experimental design. As with DCEs, an experimental design enables 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)(181) 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 said to be balanced.

Approaches for analysing 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 analysed sequentially (respondents chose the best from a set, then the worst from the remaining objects) using discrete-choice models. Alternatively, maxdiff models assume an individual makes a choice by simultaneously selecting the option with the biggest difference in utility; these are rarely implemented in practice and will not be described in detail in this section.

Best-worst scaling rating scores can be used to understand the following:

- the most and least relevant treatment outcomes to patients (182)

- the relative importance of benefits and risks of treatment (183)

- priorities for research into new health technologies (184)

- concerns for health and non-health consequences of risky activities (185)

- the importance of different adverse events on physicians' treatment decisions.(186)

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 through 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 ([Section 2](#) of the main PREFER recommendations).

A5.3.2   BWS case 1 design: method selection and analysis planning

A5.3.2.1   *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 which 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, which 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.*(187) A BWS case 1 study was conducted to understand the perceived bother of psoriatic disease manifestations and to compare patients' and physicians' perceptions. 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, which was 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 among 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.*(188) 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 that patients with anaemia placed on avoiding seven potential problems of a blood transfusion. Respondents received seven choice sets asking them to select the most and least bothersome from a subset of three transfusion attributes. Analysis of the BWS case 1 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 centre to receive a transfusion.

A5.3.2.2 When is BWS case 1 an appropriate method?

Factors suggesting that a BWS case 1 would be appropriate include the following:

- 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).

A5.3.2.3 When Is BWS case 1 not an appropriate method?

Because 'best' is not a synonym for 'acceptable', and 'worst' does not mean 'unacceptable', BWS case 1 (and, to a lesser extent, case 2) is limited in its ability to look at thresholds such as MAR, WTP, 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.

A5.3.2.4 *Points to consider in choice of preference elicitation method* (i.e. why BWS case 1 and not something else?)

Most applications of preference elicitation methods tend to involve studies where the research question requests a ranking or relative importance of many objects or attributes (e.g. Silverman *et al* (189) 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.

A5.3.2.5 *Points to consider regarding the analysis planning for BWS* case 1 studies

There is limited guidance for BWS case 1 best practice from task forces, policy makers, or other established bodies, and guidance for conducting DCEs or other preference-based methods may not be applicable. However, the textbook 'Best-Worst Scaling: Theory, Methods and Applications' (178) may be useful. For experimental design, popular software programs can convert BIBD into BWS case 1 questions (see, for example, the supporting BWS R package).(190) The choice of modelling 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* (191) provides a useful illustration of five analytical methods including count analysis, multinomial logit, random parameter logit, and latent class models,(192) as well as hierarchical Bayes estimation.

A5.3.2.6 Expected timeline for conducting a BWS case 1

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 BWS case 1 studies conducted by sponsors in collaboration with an academic group or a consulting company take approximately one year.

A5.3.2.7 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. For example, a BWS case 1 study can incorporate tests for the following:

- **stability** with a repeated choice set to test if the same objects are consistently chosen as best and/or worst

- **transitivity,** by testing if object X is worse than object Y, and object Y is worse than object Z, that object X is selected worse over object Z in a subsequent choice set

- **straight-lining or flatlining** by testing if respondents make a selection based on object location rather than information.

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* (183) 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.

A5.3.3   BWS case 1 design: sample definitions – justifying the sample size

A5.3.3.1   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 fewer than 100.(193-196) 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. Studies with many objects therefore 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,(124, 125) 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.(197)

A5.3.4 BWS case 1 design: instrument design – consideration of the appropriate number of attributes and attribute levels; patient burden issues

A5.3.4.1 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 my 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, and it could be advantageous to remove or combine items when possible.(178)

A5.3.4.2 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. For some questions, it may be more reasonable to ask about the 'most preferred' and 'least preferred' or the 'most important' or '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 end points 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 to make a choice. However, in some instances it may be that a forced choice is unrealistic, and BWS case 1 is not an appropriate method and researchers should choose an alternative method to allow for indifference (e.g. threshold technique) or the opportunity to opt-out (e.g. DCE).

A5.3.4.3 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.(181) 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.(198) For certain numbers of objects, a BIBD may exist. Alternatively, OMEP, as used in DCEs, can be used to create blocks of objects to present in the choice tasks.

A5.3.5   Why BWS 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 can be analysed using basic count models to give 'real-time' individualised results, which may be useful for integrating into shared decision-making tools.(199) Reviews have shown BWS case 1 has risen in popularity during recent years.(29, 181) These studies are also very transparent and are 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 complex research questions. As such, BWS case 1 is frequently used as a complementary method. For example, Mansfield *et al* (200) used BWS case 1 alongside a DCE to understand the preference 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.

A5.4   Best-Worst Scaling Case 2

A5.4.1   Background to BWS case 2

This section will guide a preference study sponsor through key issues when designing, conducting, and analysing a BWS case 2 preference study and provides a guide for decision makers when assessing and using BWS case 2 results to inform decision making. The section should support discussions between industry, regulators, and HTA bodies and/or payers when conducting or reviewing a BWS case 2 study.

The section starts with an overview to BWS case 2 (also known as a 'profile case') and describes the parallels and contrasts with two related methods: BWS case 1 and a DCE. The section then lays out key points of consideration when designing and analysing data from a BWS case 2 study. Specific and suitable research questions (as well as questions that are not suitable) are also outlined.

Best-worst scaling case 2 has some 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, determined by an experimental design, from which they must state which attribute level is

best or worst.(178) As with BWS case 1, BIBD designs are possible but most studies typically use designs similar to a DCE (e.g. fractional factorial designs).(181) A review of BWS studies found two-thirds of case 2 examples used OMEP. This is because introducing attribute levels means there are many more possible combinations than in BWS case 1, and BIBD can therefore quickly become too complex.

The analysis of BWS case 2 also has parallels to both BWS case 1 and DCE. A researcher can conduct simple account analysis, 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.(181) As with BWS case 1, data can also be analysed 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 later.

One of the first applications of BWS case 2 in health (although described as a 'maximum difference conjoint analysis') was a study eliciting preferences for health states.(201) 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 patients 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.(202-204)

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.(179) Some studies suggest that BWS case 2 and DCE are not significantly different,(205) whereas others found they yielded different preference estimates with poor performance of mid-ranked attributes/levels.(206)

A5.4.2   BWS case 2 design: method selection and analysis planning

A5.4.2.1   What type of research question is appropriate for a BWS case 2 study?

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* (207) 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).

A5.4.2.2 When is BWS case 2 an appropriate method?

The following factors suggest situations where BWS case 2 would be appropriate.

- BWS case 2 provides information on the relative importance of levels, as well as attributes, thus offering more information than case 1.

- Because respondents see a single profile for some samples (rather than the multiple profiles they would see in a DCE), it 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.(203)

- When an alternative utility frame is proposed, Coast *et al* (203) 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.

- Best-worst scaling case 2 provides researchers with the opportunity to tailor the study to make it more sophisticated as the research question develops. A well-designed study can be analysed with either simple or complex models, depending on the final audience.

- When real-time individualised preference data is sought (e.g. for shared decision making), count analysis can provide an immediate ranking of attribute levels.

A5.4.2.3 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, the researcher is interested in the relative importance of attributes and attribute levels, in addition to demand, MAR, or WTP. 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, MAB, WTP, 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/No'). In these instances, including a follow-up question raises two issues: firstly, how to model these responses with the best-worst data, and secondly, why not use a method more adept to modelling unconditional demand.

- Although there is evidence to the contrary,(208) there is also evidence that some respondents find discrete choice easier and selecting between best and worst sometimes conceptually difficult.(209)

A5.4.2.4 Points to consider in choice of preference-elicitation method (i.e. why BWS case 2 and not something else?)

Best-worst scaling case 2 offers more insights than the ranking of objects in BWS case 1 but can still be designed in a way that is simpler than a DCE (e.g. with BIBD designs and count analysis). For this reason, BWS case 2 can be particularly useful when the researcher (or the participants) 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 analysed 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.

A5.4.2.5 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).(178)

- As with BWS case 1, the choice of modelling approach for case 2 also depends on the underpinning psychological theory about how individuals are believed to have made choices, either sequentially 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 ISPOR 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 approximately 15% of articles.(181) Weighted least squares regression is an alternative to estimating a conditional logit model with maximum likelihood estimation.(178) In a BWS case 2 study estimating preferences for dermatology consultations, data were analysed 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.(210) It has been suggested that weighted least squares may be a useful first step for identifying outliers (i.e. attribute levels at the extremes).(178)

A5.4.2.6 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 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. need for qualitative investigations), and contracting between the relevant parties.

A5.4.2.7 Points to consider for BWS case 2 internal validity testing

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:** looking at the consistency across choices by repeating a profile to test if the respondent selects the same attribute levels when asked again.

- **transitivity:** testing whether attribute level ordering is maintained across profiles. Holding all other attribute levels constant, if an attribute level X is determined to be best (or worst) in the presence of another attribute level Y, and attribute level Y is chosen to be best (or worst) in the presence of another attribute level Z, then attribute level X should be selected as best (or worst) in a profile with attribute level Z.

- **trading behaviour:** testing for dominant preferences by measuring the times an attribute is selected as best or worst. For example, Ryan and colleagues (211) conducted a BWS case 2 study for health state valuation and created a score to identify lexicographic behaviour, 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 (179) also found that several BWS case 2 studies used self-reported measures of difficulty to assess understanding.

A5.4.3 BWS case 2 design: sample definitions – justifying the sample size

A5.4.3.1 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 (212) and with much larger samples (more than 1,000 respondents in some health state valuation.(213) Because no sample size calculations exist specifically for case 2, researchers can use the rules of thumb from the DCE literature.(123) 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.

A5.4.4 BWS case 2 design: instrument design – consideration of the appropriate number of attributes and attribute levels; patient burden issues

A5.4.4.1 Points to consider regarding the choice of attributes

As with other stated preference methods, the attributes describing the profiles will be determined by the research question. As for 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, it is arguable that a BWS case 2 study can 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, firstly, the features of treatment important to the decision maker (e.g. a regulator, a HTA body, a physician), and secondly, 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 researchers may seek to consult with experts or conduct qualitative research to identify or reduce attributes.(214, 215)

A5.4.4.2 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 behaviour (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 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.

A5.4.4.3 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, although these 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 the respective DCE section. Indeed, most BWS case 2 studies used orthogonal main effects plans (OMEP).(181)

A5.4.5 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 BWS method, with slightly more applications than case 1 and many more applications than case 3 in health.(181, 216)

A5.5 Threshold technique

A5.5.1 Background to threshold technique

Threshold-based approaches to preference elicitation are a type of indifference method that aims to find combinations of attributes (e.g. benefits and risks) that offer the same level of utility.(126) Typically, these methods vary the value of one attribute in an option until the participant is indifferent to the alternatives. In addition to threshold technique, threshold

approaches include the standard gamble, time trade-off, and contingent valuation.(126) 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.

This section provides an overview to threshold technique and will guide researchers, sponsors, and reviewers through key issues when designing, conducting, and analysing a threshold technique study. The section also serves as a guide for decision makers when evaluating or seeking to use the results of a threshold technique study. In addition, this section should facilitate discussions between industry, regulators, and HTA bodies/payers about threshold technique preference studies and their role in medical product decision making.

In a threshold technique 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 and is usually the 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, respondents are required to make a trade-off when choosing their preferred alternative.

The alternative in a threshold technique, like 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).(217)

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 threshold technique 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 where the respondent is indifferent to the reference and target options. As each threshold technique series is focused on eliciting the threshold for a particular attribute, to understand thresholds for other attributes or other target treatments, the threshold technique exercise should be repeated. Therefore, to some extent, threshold technique 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:

- 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.

A5.5.2   Threshold technique design: method selection and analysis planning

A5.5.2.1   What type of research question is appropriate for a threshold technique study?

Typically, the threshold technique is used when there are a few attributes or trade-offs of interest. In instances when there are many attributes, a DCE (if fewer than approximately eight 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 threshold technique cannot incorporate many attributes as is possible with BWS case 1, responses indicate demand rather than the relative order of preferences alone.

A good example of a multi-attribute threshold technique is described in a paper by Devereaux and colleagues.(218) The authors conducted a probabilistic threshold technique study with patients and physicians to understand 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.*(219) The study investigated the trade-offs patients with haemophilia were willing to make when choosing between prophylactic factor replacement therapy (PFRT) and a new gene therapy. The survey design incorporated three series of threshold technique 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.

A5.5.2.2   When is threshold technique an appropriate method?

The following factors may suggest that a threshold technique would be appropriate:

- When pairwise trade-offs are all that are 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 threshold technique 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, very young children). There is emerging evidence suggesting that the threshold technique may be more suitable in samples of patients who are unable to make the complex trade-offs required by DCEs or the ranking needed in a BWS study.(220) 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 threshold technique is a direct elicitation method,(109) 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 threshold technique 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.

A5.5.2.3 When is threshold technique not an appropriate method?

The following are examples of situations where threshold technique would be inappropriate or insufficient:

- When the research question lends itself to a more complex method because there are many key attributes or trade-offs of interest that would result in too many threshold technique exercises for any individual to complete.

- When interactions between the attributes are important. A threshold technique study cannot easily accommodate interactions (i.e., 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 which can simultaneously estimate preferences for multiple attributes and interactions maybe more appropriate. Discrete-choice experiments and related methods are most adept at incorporating designs to estimate between attribute interactions.

A5.5.2.4 Points to consider in choice of preference elicitation method (i.e. why threshold technique and not something else?)

Most applications tend to involve estimating thresholds for benefit–risk assessments, particularly to understand MAR and MAB.(217) Although studies have incorporated multiple thresholds (e.g. Tomlinson *et al*),(221) a review showed that most have focused on one or two key attributes of interest.(217) The relative simplicity of a threshold technique 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.

A5.5.2.5 Points to consider regarding the analysis planning for threshold technique studies

As of 2020, there exists no specific guidance for the analyses of threshold technique 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 bound and upper-bound 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.

Multicollinearity 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.

A5.5.2.6   Expected timeline for conducting a threshold technique

As with all preference methods, the expected timeline for conducting a threshold technique study depends on practical and logistical constraints, including contracting between parties, recruitment of the sample, need for ethical approval, and the extent of engagement with decision makers. Generally, the timeline of a threshold technique 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.

A5.5.2.7   Points to consider for threshold technique internal validity testing

Previous sections noted techniques for understanding the validity of a preference study, and many of these could be informative in a threshold technique study (including stability, straight-lining, or flatlining). For threshold technique specifically, some 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 threshold technique series.

- **Anchoring effects.** In a threshold technique 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 threshold technique may be more susceptible to starting point biases than DCEs, but the evidence is varied.(222, 223) 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.(217)

- **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 threshold technique studies have also investigated the test–retest reliability of responses by repeating the threshold technique exercises with the same sample a few weeks or months later.(224-226) 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 interlude could be affected by external shocks, experiences, or simply the act of thinking about the choices made in the original exercise.(227)

It may also be useful to explore the face validity of response, 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.

A5.5.3 Threshold technique design: sample definitions – justifying the sample size

A5.5.3.1 Points to consider regarding the sample size of a threshold technique study

There is no specific power calculation to determine sample size in threshold technique studies without knowing the expected threshold value *a priori*. Most threshold technique studies are conducted with 100 or fewer respondents, and substantially smaller samples (between 20 and 42 respondents) have been used successfully in previous studies.(228-231) Although there is a lack of clear guidance on sample size estimation for threshold technique studies, it is generally assumed that a minimum of 50 responses per threshold technique choice set would be needed to estimate a threshold value in each threshold exercise.

A5.5.4 Threshold technique design: instrument design – consideration of the appropriate number of attributes and attribute levels; patient burden issues

A5.5.4.1 Points to consider regarding the choice of attributes

Two key points of consideration in a threshold technique are (1) the target and reference treatments and (2) the attribute(s) to be varied.(232) 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 which differentiate the target

treatment and provide information about the thresholds of interest, and this is typically the attribute to be varied.

Although a threshold technique study can include multiple attributes, this results in more question series. Too many series may induce respondent fatigue, resulting in poorer quality preference data.

A5.5.4.2 Points to consider in the range of levels

The levels in a threshold technique 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 maybe asked to state their threshold directly.

If the level of a fixed attribute is 100%, respondents in the threshold technique 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.(223, 233)

A5.5.4.3 Points to consider in the set-up of choice sets (experimental design)

The experimental design refers to the specific attributes and levels used in the choice tasks that responders complete. Because each threshold technique series is typically varying 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 minimise potential biases. When using threshold technique to estimate preferences for multiple attributes, researchers should randomize the threshold technique series to avoid ordering effects.

A5.5.5 Why threshold technique?

The threshold technique is becoming an increasingly popular approach for quantifying preferences, and a review of the method found 43 examples published between 1991 and 2016.(217) Probabilistic threshold technique is particularly suited to estimating individuals' MAR and MAB, and the method is therefore promising for benefit–risk decisions, specifically. The FDA CDRH has used threshold technique to quantify preferences for health technologies. For example, data from a threshold technique study quantifying preferences for ear tube placement were used to support regulatory approval of the Tula ear system.(5) 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 anaesthesia with a success of 99%.

A5.6 Strengths, limitations, and uncertainties of method selection recommendations

One of the strengths of the recommendations is that they are based on perspectives from different stakeholders, including patients, healthcare professionals, and experts in the field. Criteria from the systematic review of the literature and from previous work (the MDIC) were also taken into account. At different stages of the PREFER project, various approaches to evaluating the preference methods against these criteria were employed, and the results were, in general, consistent, which further strengthened the recommendations. PREFER undertook a systematic literature review, Q-methodology, and AHP in WP2, and later used MCDA in WP3.

Method selection recommendations are partly based on Whichello *et al*,(234) and the design limitations of that study should be considered when interpreting the recommendations. For example, there was a lack of consensus among experts about the performance of elicitation methods on the method criteria. Furthermore, consideration should be given to the limitations in the design of medical product life cycle scenarios used in the AHP exercise and the sample. As only a limited number of PREFER case studies were used to further inform method selection recommendations, additional research is needed to support their findings.

Two sources were used to identify method criteria: work performed by PREFER in WP2 and work undertaken for the EMA qualification procedure (i.e. the briefing book). While there was some overlap between the two sources, several criteria originated from only one source. In WP2, a robust approach was taken to the identification and application of these criteria. While the evidence level for criteria originating solely from the briefing book was mainly based on expert opinion and experience, the criteria were also considered relevant and therefore included in the recommendations. Sources and criteria are as follows:

- Criteria originating from WP2 and the briefing book:
  - sample size ≤100
  - estimates weights (relative importance) for attributes
  - estimates trade-offs between attributes
  - quantifies heterogeneity in preferences
  - a low cognitive burden on patients.
- Criteria originating from WP2 only:
  - ≥8 attributes can be explored
  - ease with which new attributes can be added
  - calculates risk attitudes

- explores reasons behind a preference in qualitative detail

- internal validation methods can be incorporated

- establishes external validity

- public acknowledgement as an acceptable method to study preferences

- no interaction between participants (solitary exercise)

- group dynamic with participants

- low complexity of instructions to participants

- low cost

- quick sessions with participants (≤30 min)

- low frequency of sessions (<2 sessions)

- study duration (≤6 months).

- Criteria added from the briefing book only:

  - provides estimates at the level of the individual

  - provides estimates at the level of the sample/population

  - estimates preferences for individual attributes

  - estimates preferences over multiple levels of each attribute

  - simultaneous estimation of trade-offs between multiple attributes

  - pairwise estimation of trade-offs between attributes

  - can accommodate interactions between treatment characteristics

  - specific methodological expertise required for design

  - specific methodological expertise required for conduct

  - specific methodological expertise required for analysis.

Table A6-1. Group of available empirical evidence and consensus-based recommendations.

Psychological construct	Group of the empirical evidence available	Class of recommendation
Anxiety	C	III
Assertiveness	B	III
Autonomy preference	B	II
Behavioural inhibition and activation	C	III
Conservatism	B	III
Control preference	B	I
Coping style	B	III
Decision-making style	B	II
Depression	C	III
Dispositional optimism	B	III
Health anxiety	C	III
Health literacy	A	I
Health locus of control	A	I
Health numeracy	A	I
Health orientation	B	II
Illness perception	C	I
Mastery	C	III
Mood states	C	III
Need for closure	C	III
Need for cognition	C	III
Patient activation	B	I
Personality	C	III
Psychological well-being	C	III
Rational and experiential thinking styles	C	III
Resilience	B	III
Risk propensity	B	I
Self-efficacy	B	III
Sensation seeking	C	III
Sense of coherence	C	III
Social support	C	III
Treatment-related beliefs	B	I

Group A: psychological dimensions for which there are strong and consistent results regarding their associations with patient preferences and decisions. Group B: psychological dimensions that are theoretically linked to patient preferences even if related empirical evidence is not yet satisfactory. Group C: psychological dimensions for which evidence is not available, or there are inconclusive or inconsistent results on their role in influencing preference. class I: unanimity or high levels of agreement among experts; class II: agreement among a majority of experts; class III: disagreement among experts

A7   Annex for Section 7

To better understand the impact that educational tools can have on preference outcomes, the PREFER consortium conducted five case studies using enhanced educational materials. These case studies assessed preferences for treatments in five different disease and treatment contexts (haemophilia, lung cancer, NMD, diabetes, and rheumatoid arthritis), using five different preference assessment techniques, with respondents from 10 different countries who had varying levels of prior knowledge and experience with the disease and treatment options.

The content and aim of the educational materials varied between the case studies but covered disease and treatment information, attribute information, and instructions on how to understand and complete the choice tasks. Two of the PREFER studies (rheumatoid arthritis and diabetes) evaluated the direct effects on the validity and reliability of the preference assessment, using a split sample with random assignment of patients to either enhanced educational materials or written text. The other three studies included qualitative assessments of the value of the enhanced educational materials. An overview of the studies conducted within PREFER, along with the two previous studies assessing the impact of enhanced educational materials, can be found in **Table A7-1**.

Table A7-1. Summary of patient preference studies assessing enhanced educational material.

PREFER case study	PAVING-haemophilia	No split sample		Split sample	
		Lung cancer	NMD, Newcastle	Diabetes, Utrecht	Rheumatoid arthritis, Uppsala
Disease area	Haemophilia	NSCLC	NMD	Diabetes	Rheumatoid arthritis
Sample inclusion criteria	Self-reported[a] diagnosis of moderate/severe haemophilia A or B; age ≥18 years; living in Belgium at time of survey	NSCLC patients stage I–IV; age ≥18 years; able to read, speak, and understand Italian, Dutch, or French	Personal diagnosis of a mitochondrial disorder or myotonic dystrophy, or caregiver of someone with a mitochondrial disorder or myotonic dystrophy; age ≥18 years; or their caregiver	Self-reported diagnosis diabetes mellitus (type I, type II, MODY, LADA); age ≥18 years; able to read Dutch or Polish	(Self-reported) rheumatoid arthritis diagnosis; age 18–80 years; able to understand and answer the questions; able to read and understand Swedish without aid
Sample size	117	307	517	981	675
Preference elicitation technique	Threshold technique	DCE, swing weighting	DCE, BWS case 2, Q-methodology	DCE, swing weighting	DCE
Preference sensitive situation	Preferences for gene therapy or standard of care that are important to patients	Preferences for different NSCLC treatments attributes	Identification of missing medical treatments and desired symptoms	Preferences for attributes of glucose monitoring technology used in chronic disease management	Preferences for second-line treatment of rheumatoid arthritis

PREFER case study	No split sample			Split sample	
	PAVING-haemophilia	Lung cancer	NMD, Newcastle	Diabetes, Utrecht	Rheumatoid arthritis, Uppsala
Treatment context	Gene therapy (novel) vs. current standard of care (prophylactic clotting factor administration with high administration burden)	Hypothetical, acute treatments that are likely novel to the patient and have a large patient burden both regarding the treatment itself and the side-effects	Novel treatment as the study addresses hypothetical treatments as no treatment available at the time of the study	Treatments are chronic and not novel to patients; treatments are meant to reduce the chance of short- and long-term treatment effects	Second-line treatments not novel to patients
Patient experience in disease or treatment population	Heterogeneous; likely some prior knowledge/ experience	Generally older; likely no prior knowledge/ experience	Heterogeneous (mild and severe and patients and caregivers) with no prior knowledge/ experience regarding treatments	Heterogeneous; likely experienced with a lot of knowledge regarding the disease or treatment	Heterogeneous; likely some prior knowledge/ experience
Recruitment	Haemophilia reference centres, patient organisation	Lung cancer outpatient clinics in Belgium and Italy	UK Myotonic Dystrophy Patient Registry and the MRC MitoCohort	Online panel (SurveyEngine)	Online panel (n=162) (dynata.com), patient organisation (Swedish Rheumatism Association, n=228), and rheumatology clinic (Uppsala University Hospital, Sweden, n=285)

PREFER case study	No split sample			Split sample	
	PAVING-haemophilia	Lung cancer	NMD, Newcastle	Diabetes, Utrecht	Rheumatoid arthritis, Uppsala
Educational features of enhanced educational materials[b]	Icons, animations, interactive navigation, voice-over	Icons, animations, interactive navigation, voice-over	Icons, animations, interactive navigation, voice-over	Icons, animations, interactive navigation, voice-over	Icons, animations, interactive navigation, voice-over
Aim of educational tool	Explain context for gene therapy and attributes used in choice task	Explain how to complete choice tasks	Explain how to complete choice tasks	Explain general information on diabetes, use of glucose monitors in diabetes self-management, and attributes used in choice task	Explain disease symptoms, choice task, and attributes used in choice task
Comparison of different educational tools	No	No	No	Yes	Yes
Alternative educational material	N/A	N/A	N/A	Plain text with static graphics	Plain text with static graphics
Found <u>between group</u> differences in preference estimates	Yes; participants who spent more time on the educational tools tolerated more additional bleeds and reductions in QoL	N/A	N/A	Yes; unsystematic differences in relative importance and parameter level estimates	Yes; substantial differences in relative importance of attributes

PREFER case study	PAVING-haemophilia	No split sample		Split sample	
		Lung cancer	NMD, Newcastle	Diabetes, Utrecht	Rheumatoid arthritis, Uppsala
Comparison of dropouts, flatliners, dominant decision making[c]	Yes; no outcomes presented	N/A	N/A	No dropouts during standard text; n=62 during enhanced educational materials	Plain text had more flat-liners (n=29) and respondents answering the survey in <5 mins (n=35) than the enhanced educational tool (n=3 and n=8, respectively)
Assessed comprehension	Yes; n=1 risk grid comprehension question was included; n=1 participant excluded from analysis for failing check	N/A	N/A	No	No
Assessed health literacy and numeracy	Yes, for demographics	Yes, for demographics	Yes, for demographics	Yes, for subgroup analysis	Yes, for demographics
Found differences in reliability[d]	Yes; participants who spent more time on the educational tools were more consistent in their preferences	N/A	N/A	No	No

PREFER case study	No split sample			Split sample	
	PAVING-haemophilia	Lung cancer	NMD, Newcastle	Diabetes, Utrecht	Rheumatoid arthritis, Uppsala
Found significant differences in time spent on the choice tasks	N/A	N/A	N/A	No	Yes, enhanced educational Edu-Arm spent slightly more time

[a] Participants were considered as self-reported if recruited through other channels than clinical care or patient organisations. [b] Edu-Arm = participants that received enhanced multimedia educational material. [c] Flat-lining: choosing only one alternative frequently (80–100%), e.g. if the options are A or B, the participant always chooses A. [d] Reliability refers here to the scale homogeneity or error variance, which in these studies can be said to be a measure of choice consistency.

Case study example for Section 7.3: The PAVING study

The PAVING study investigated what trade-offs between treatment characteristics adult Belgian haemophilia A and B patients were willing to make when asked to choose between a standard of care and gene therapy.(219)

Step 1: Identification of educational needs

The educational needs were identified in meetings with a stakeholder advisory board (including patients, caregivers, medical specialists, HTA bodies and payers, industry market access experts, and patient education [EUPATI] experts) and in semi-structured interviews with haemophilia A and B patients. The general view was that, although most patients had a basic awareness of gene therapy, they did not know about the mode of action or the practicalities, benefits, risks, and uncertainties of the therapy. The aim of the educational module was thus to ensure that patients participating in the study had a consistent understanding of their disease and current treatment, along with the mechanism of action, administration of gene therapy, and its risks and benefits. The preference elicitation method (probabilistic threshold technique) was perceived by both researchers and patient partners to be simple to understand by the target population and did not require enhanced educational materials.

Step 2: Selection and adaptation of formats

The content of the educational materials was developed during the qualitative phase of the preference study and included information on the disease, current therapies, administration, efficacy, and safety of the treatments (**Figure A7-1**). The suggested educational materials were reviewed in interviews with haemophilia patients and subsequently revised. Patients in the semi-structured interviews were provided with illustrations to visualise difficult concepts.

Step 3: Select and adapt educational features

The education materials included a video, animations, interactive navigation, and a voiceover to ensure that those with low health literacy would be better able to understand the information (**Figure A7-2**).
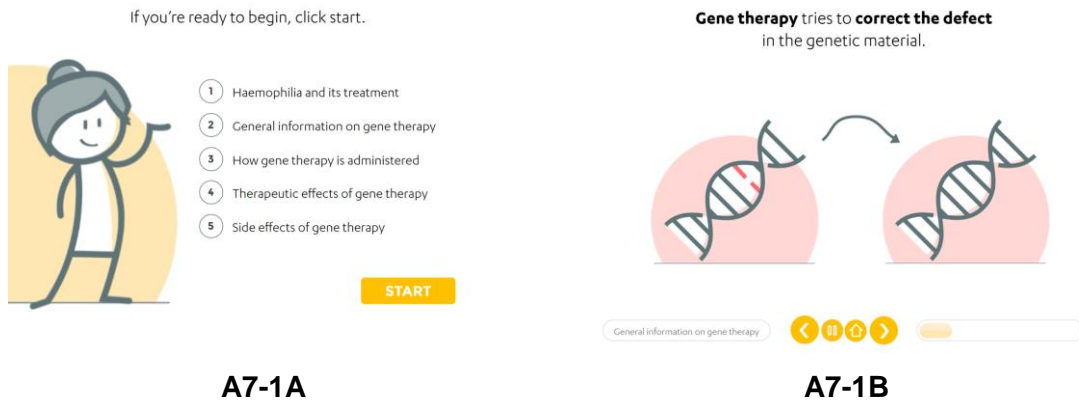
A7-1A

A7-1B

Figure A7-1. Example of multimedia graphics used to educate patients with haemophilia A and B in a preference elicitation study.

A7-2A

A7-2B

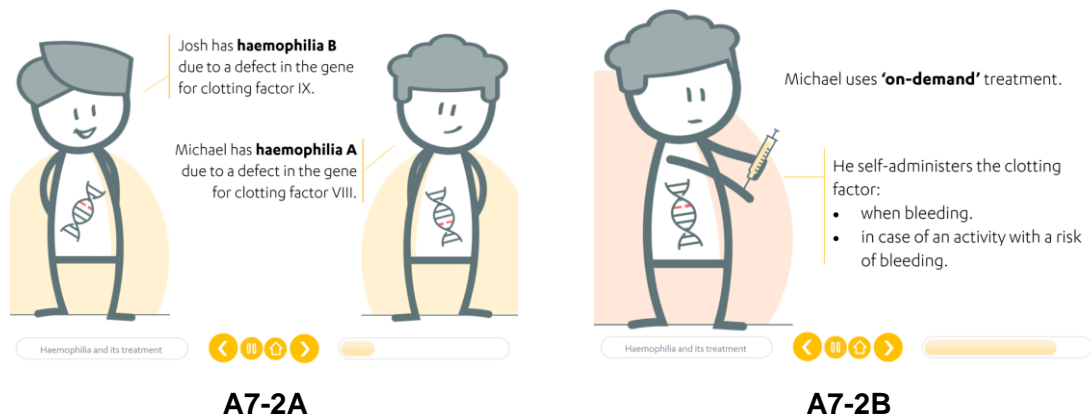**Figure A7-2.** Example of multimedia graphics used to educate patients with haemophilia A and B in a preference elicitation study.

Because conducting patient preference studies is time-consuming and expensive, re-using previously collected preference data in other decision contexts is an attractive proposition; however, in contrast with other scientific fields, research on the transferability of patient preferences is still in its infancy. To enhance the sustainability of research efforts and increase the speed of use of patient preference information in decision-making, more research is needed to synthesise existing patient preference information, as well as to understand its transferability. As part of the PREFER project, a checklist was proposed to assess the degree of transferability across countries or diseases / patient groups / indications. The checklist contains questions relating to methodology, and population and healthcare context characteristics that may affect the transferability of patient preference results to other decision contexts. It considers two typical transferability situations:

- the transferability of study results obtained within one country to another country

- the transferability of study results obtained for one specific patient group, disease, or indication of a medical product to a related patient group or disease or another indication of the same medical product.

For transferability across countries, the assumption is that there is a relevant and methodologically sound patient preference study that was executed in one country and the question is if the results of this study can be transferred to another country with a similar target patient population. This implies that a study with methodological flaws will never be transferable to other countries (knock-out criterion).

For transferability across diseases, patient groups, or indications, the assumption is that there is a relevant and methodologically sound patient preference study that was executed for one disease, patient group or indication and the question is if the results of this study can be transferred to another disease or patient group (within a spectrum of related diseases), or another indication for a given medical product. Again, a study with methodological flaws will never be transferable to other diseases, patient groups or indications. Examples of transferability questions between contexts include:

- Are preferences for attributes of pharmaceutical treatments for chronic moderate-to-severe musculoskeletal pain in osteoarthritis patients transferable to chronic low back pain patients?

- Are preferences of melanoma patients using immunotherapy transferable to patients with lung cancer using similar immunotherapy?

- Are preferences of patients with NMD transferable between subgroups of patients (e.g. myotonic dystrophy type 1 and mitochondrial myopathies)?

The checklist contains three categories of issues that should be considered to assess transferability:

- methodological characteristics
- population characteristics
- healthcare context characteristics.

More details are shown in **Box A8-1**.

Box A8-1. Patient preference study transferability checklist.

Methodological characteristics

- attributes
- levels and their range

Population characteristics

- sociodemographic and educational characteristics of the population
- epidemiologic characteristics
- attitudinal characteristics (only for transferability across countries)
- cultural and religious beliefs (only for transferability across countries)
- cognitive characteristics (only for transferability across diseases / patient groups / indications)

Healthcare context characteristics

- commercial and financial (reimbursement) availability of medical products
- geographical accessibility of medical products
- level of experience with use of medical products or adverse health effects
- level of trust of patients in treatment with a medical product
- the standard of care for treatment of a certain disease
- healthcare and social security system (only for transferability across countries)

A9 Annex references

1. Postmus D, Richard S, Bere N, van Valkenhoef G, Galinsky J, Low E, *et al.* Individual Trade-Offs Between Possible Benefits and Risks of Cancer Treatments: Results from a Stated Preference Study with Patients with Multiple Myeloma. Oncologist. 2018;23(1):44–51.

2. Postmus D, Mavris M, Hillege HL, Salmonson T, Ryll B, Plate A, *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.

3. European Medicines Agency. The patient's voice in the evaluation of medicines. How patients can contribute to assessment of benefit and risk. 18 October 2013. https://www.ema.europa.eu/en/documents/report/report-workshop-patients-voice-evaluation-medicines_en.pdf.

4. U.S. Food & Drug Administration. FDA News Release. FDA approves system for the delivery of ear tubes under local anesthesia to treat ear infection. 25 Nov 2019. https://www.fda.gov/news-events/press-announcements/fda-approves-system-delivery-ear-tubes-under-local-anesthesia-treat-ear-infection. Accessed 22 April 2022.

5. U.S. Food & Drug Administration. Summary of Safety and Effectiveness data (SSED) Tula System. 25 November 2019. https://www.accessdata.fda.gov/cdrh_docs/pdf19/P190016B.pdf.

6. Patalano F, Gutzwiller FS, Shah B, Kumari C, Cook NS. Gathering Structured Patient Insight to Drive the PRO Strategy in COPD: Patient-Centric Drug Development from Theory to Practice. Adv Ther. 2020;37(1):17–26.

7. National Institute for Health and Care Excellence. NICE provides first scientific advice on patient preference study design. 14 February 2019. https://www.nice.org.uk/news/article/nice-provides-first-scientific-advice-on-patient-preference-study-design. Accessed 22 April 2022.

8. Ho MP, Gonzalez JM, Lerner HP, Neuland CY, Whang JM, McMurry-Heath M, *et al.* Incorporating patient-preference evidence into regulatory decision making. Surg Endosc. 2015;29(10):2984–93.

9. U.S. Food & Drug Administration. FDA News Release. FDA Approves New Kind Of Device To Treat Obesity. 14 Jan 2015. https://wayback.archive-it.org/7993/20170111191732/http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm430223.htm. Accessed 22 April 2022.

10. Janssen Research & Development L. Advisory Committee Briefing Document: Esketamine Nasal Spray for Patients with Treatment-resistant Depression. 16 January 2019. https://www.fda.gov/media/121377/download. Janssen Research & Development, LLC.

11. U.S. Food & Drug Administration. Joint Meeting of the Psychopharmacologic Drugs Advisory Committee (PDAC) and the Drug Safety and Risk Management (DSaRM) Advisory Committee 12 February 2019. https://www.fda.gov/advisory-

committees/february-12-2019-joint-meeting-psychopharmacologic-drugs-advisory-committee-pdac-and-drug-safety-and. Accessed 22 April 2022.

12. Medical Device Innovation Consortium (MDIC). MDICx Series. NxStage. 2016. https://mdic.org/event/a-case-study-in-how-patient-preference-information-contributes-to-regulatory-decisions-for-medical-devices/.

13. U.S. Food & Drug Administration. Department of health and human services. Approval letter for NXStage System One. 24 August 24 2017. https://www.accessdata.fda.gov/cdrh_docs/pdf17/k171331.pdf.

14. U.S. Food & Drug Administration. CDRH Patient Engagement. 24 January 2022. https://www.fda.gov/about-fda/cdrh-patient-science-and-engagement-program/cdrh-patient-engagement. Accessed 22 April 2022.

15. Pivot X, Spano JP, Espie M, Cottu P, Jouannaud C, Pottier V, *et al.* Patients' preference of trastuzumab administration (subcutaneous versus intravenous) in HER2-positive metastatic breast cancer: Results of the randomised MetaspHer study. Eur J Cancer. 2017;82:230–6.

16. 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 (PrefMab). Ann Oncol. 2017;28(4):836–42.

17. Genentech. Rituxan Hycela prescribing information. 2017. https://www.gene.com/download/pdf/rituxan_hycela_prescribing.pdf.

18. Benz HL, Caldwell B, Ruiz JP, Saha A, Ho M, Christopher S, *et al.* Patient-Centered Identification of Meaningful Regulatory Endpoints for Medical Devices to Treat Parkinson's Disease. MDM Policy Pract. 2021;6(1):1–13.

19. U.S. Food & Drug Administration. Comparing Qualitative and Quantitative Approaches to Eliciting Patient Preferences: A Case Study on Innovative Upper Limb Prostheses. 2016. https://www.fda.gov/science-research/advancing-regulatory-science/comparing-qualitative-and-quantitative-approaches-eliciting-patient-preferences-case-study. Accessed 15 February 2022.

20. U.S. Food & Drug Administration Advancing Regulatory Science. Patient and Provider Views on Clinical Endpoints: A Qualitative Preference Study of Minimally Invasive Glaucoma Surgical (MIGS) Devices. 29 March 2018. https://www.fda.gov/science-research/advancing-regulatory-science/patient-and-provider-views-clinical-endpoints-qualitative-preference-study-minimally-invasive.

21. Hauber B, Mange B, Zhou M, Chaudhuri S, Benz HL, Caldwell B, *et al.* Parkinson's Patients' Tolerance for Risk and Willingness to Wait for Potential Benefits of Novel Neurostimulation Devices: A Patient-Centered Threshold Technique Study. MDM Policy Pract. 2021;6(1):1–13.

22. Janssen EM, Benz HL, Tsai JH, Bridges JF. Identifying and prioritizing concerns associated with prosthetic devices for use in a benefit-risk assessment: a mixed-methods approach. Expert Rev Med Devices. 2018;15(5):385–98.

23. Tarver ME, Neuland C. Integrating Patient Perspectives into Medical Device Regulatory Decision Making to Advance Innovation in Kidney Disease. Clinical Journal of the American Society of Nephrology. 2021;16(4):636.

24. U.S. Food & Drug Administration. FY 2018 Report from the Director. 2018. https://www.fda.gov/vaccines-blood-biologics/fy-2018-report-director.

25. Ho M, Gonzalez JM, Mansfield C, Wilson L. Three preference studies sponsored by FDA CBER to inform benefit-risk assessment for biologics. Presented at the Virtual ISPOR 2021 Conference; 20 May 2021.

26. 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. 2015. https://mdic.org/wp-content/uploads/2015/05/MDIC_PCBR_Framework_Proof5_Web.pdf.

27. Bridges JF, Hauber AB, Marshall D, Lloyd A, Prosser LA, Regier DA, 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;14(4):403–13.

28. U.S. Food & Drug Administration. Department of Health and Human Services. Center for Devices and Radiological Health & Center for Biologics Evaluation and Research. Patient Preference Information - Voluntary Submission, Review in Premarket Approval Applications, Humanitarian Device Exemption Applications, and De Novo Requests, and Inclusion in Decision Summaries and Device Labeling. Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders. 24 August 2016. https://www.fda.gov/media/92593/download. Accessed 22 April 2022.

29. Mühlbacher A, Johnson FR. Choice Experiments to Quantify Preferences for Health and Healthcare: State of the Practice. Appl Health Econ Health Policy. 2016;14(3):253–66.

30. van Overbeeke E, Whichello C, Janssens R, Veldwijk J, Cleemput I, Simoens S, et al. Factors and situations influencing the value of patient preference studies along the medical product lifecycle: a literature review. Drug Discov Today. 2019;24(1):57–68.

31. European Medicines Agency. ICH guideline M4E(R2) - Common technical document for the registration of pharmaceuticals for human use – Efficacy. 15 July 2016. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-m4e-r2-common-technical-document-registration-pharmaceuticals-human-use-efficacy-step-5_en.pdf.

32. van Overbeeke E, Janssens R, Whichello C, Schölin Bywall K, Sharpe J, Nikolenko N, et al. Design, Conduct, and Use of Patient Preference Studies in the Medical Product Life Cycle: A Multi-Method Study. Frontiers in Pharmacology. 2019;10(1395).

33. U.S. Food & Drug Administration. Patient-Focused Drug Development: Collecting Comprehensive and Representative Input Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders. June 2020. https://www.fda.gov/media/139088/download.

34. European Medicines Agency. ICH E9 Statistical principles for clinical trials. Note for guidance on statistical principles for clinical trials. September 1998.

https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5_en.pdf.

35. U.S. Food & Drug Administration. Patient-Focused Drug Development: Collecting Comprehensive and Representative Input Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders. June 2020. https://www.fda.gov/media/139088/download.

36. Nunley S, Glynn P, Rust S, Vidaurre J, Albert DVF, Patel AD. 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;92:53–6.

37. Oremus M, Tarride JE, Pullenayegum E, Clayton N, Mugford G, Godwin M*, et al.* Caregivers' willingness-to-pay for Alzheimer's disease medications in Canada. Dementia (London). 2015;14(1):63–79.

38. Hauber AB, Arellano J, Qian Y, González JM, Posner JD, Mohamed AF*, et al.* Patient preferences for treatments to delay bone metastases. Prostate. 2014;74(15):1488–97.

39. International Council for Harmonisation of technical requirements for pharmaceuticals for human use. ICH harmonised guideline. Addendum on estimands and sensitivity analysis in clinical trials. To the guideline on statistical principles for clinical trials. E9 (R1). 2019. https://www.gmp-compliance.org/files/guidemgr/E9-R1_Step4_Guideline_2019_1203.pdf.

40. Pope C, Ziebland S, Mays N. Qualitative research in health care. Analysing qualitative data. BMJ. 2000;320(7227):114–6.

41. Lacey A, Luff D. Qualitative Research Analysis. The NIHR RDS for the East Midlands / Yorkshire & the Humber, 2007 https://www.rds-yh.nihr.ac.uk/wp-content/uploads/2013/05/9_Qualitative_Data_Analysis_Revision_2009.pdf.

42. Hollin IL, Craig BM, Coast J, Beusterien K, Vass C, DiSantostefano R*, et al.* Reporting Formative Qualitative Research to Support the Development of Quantitative Preference Study Protocols and Corresponding Survey Instruments: Guidelines for Authors and Reviewers. Patient. 2020;13(1):121–36.

43. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC medical research methodology. 2013;13(1):117.

44. Bradley EH, Curry LA, Devers KJ. Qualitative data analysis for health services research: developing taxonomy, themes, and theory. Health Serv Res. 2007;42(4):1758–72.

45. Howitt D. Introduction to Qualitative Research Methods in Psychology, 3rd Edition: Pearson; 2016.

46. Whichello C, van Overbeeke E, Janssens R, Schölin Bywall K, Russo S, Veldwijk J*, et al.* Factors and Situations Affecting the Value of Patient Preference Studies: Semi-Structured Interviews in Europe and the US. Frontiers in Pharmacology. 2019;10(1009).

47. Coast J, Al-Janabi H, Sutton EJ, Horrocks SA, Vosper AJ, Swancutt DR, *et al.* Using qualitative methods for attribute development for discrete choice experiments: issues and recommendations. Health Econ. 2012;21(6):730–41.

48. 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;22(9):1063-9.

49. Russo S, Jongerius C, Faccio F, Pizzoli SFM, Pinto CA, Veldwijk J, *et al.* Understanding Patients' Preferences: A Systematic Review of Psychological Instruments Used in Patients' Preference and Decision Studies. Value Health. 2019;22(4):491–501.

50. Johnson FR, Yang JC, Reed SD. The Internal Validity of Discrete Choice Experiment Data: A Testing Tool for Quantitative Assessments. Value Health. 2019;22(2):157–60.

51. Janssen EM, Marshall DA, Hauber AB, Bridges JFP. Improving the quality of discrete-choice experiments in health: how can we assess validity and reliability? Expert Rev Pharmacoecon Outcomes Res. 2017;17(6):531–42.

52. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, *et al.* Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. Value Health. 2005;8(2):94–104.

53. Mansfield C, Poulos C, Boeri M, Hauber B. Performance of a comprehension question in discrete-choice experiment surveys (DCE). Poster presented at the ISPOR 2019 European Conference; November 5, 2019. Copenhagen, Denmark. [abstract] Value Health. 2019 Dec; 22(S3). https://rtihs.org/publications/performance-comprehension-question-discrete-choice-experiment-surveys-dce.

54. Ryan M, Watson V, Entwistle V. Rationalising the 'irrational': a think aloud study of discrete choice experiment responses. Health Econ. 2009;18(3):321–36.

55. U.S. Food & Drug Administration. Plain Language Principles. 11 January 2022. https://www.fda.gov/about-fda/plain-writing-its-law/plain-language-principles. Accessed 22 April 2022.

56. Jacobson KL, Parker RM. Health Literacy Principles: Guidance for Making Information Understandable, Useful, and Navigable. 22 December 2014. https://nam.edu/wp-content/uploads/2015/06/HealthLiteracyGuidance.pdf.

57. Sroka-Saidi K, Boggetti B, Schindler TM. Transferring regulation into practice: The challenges of the new layperson summary of clinical trial results. Medical Writing. 2015;24(1):24–7.

58. National Institute for Health Research. Plain English summaries. 5 April 2021. https://www.nihr.ac.uk/documents/plain-english-summaries/27363. Accessed 22 April 2022.

59. European Medicines Agency. Day 80 assessment report overview and D120 LOQ template_en Rev 10.19. 2019. https://www.ema.europa.eu/en/human-regulatory/marketing-authorisation/assessment-templates-guidance.

60. Bouvy JC, Cowie L, Lovett R, Morrison D, Livingstone H, Crabb N. Use of Patient Preference Studies in HTA Decision Making: A NICE Perspective. Patient. 2020;13(2):145–9.

61. The Pharmaceutical Benefits Board. General guidelines for economic evaluations from the Pharmaceutical Benefits Board (LFNAR 2003:2). 1 May 2003. https://www.tlv.se/download/18.2e53241415e842ce95514e9/1510316396792/Guidelines-for-economic-evaluations-LFNAR-2003-2.pdf.

62. eunethta: European Network for Health Technology Assessment. Methods for health economic evaluations A guideline based on current practices in Europe. May 2015. https://www.eunethta.eu/wp-content/uploads/2018/01/Methods-for-health-economic-evaluations-A-guideline-based-on-current-practices-in-Europe_Guideline_Final-May-2015.pdf.

63. Versteegh MM, Brouwer WBF. Patient and general public preferences for health states: A call to reconsider current guidelines. Social science & medicine (1982). 2016;165:66–74.

64. Mohamed AF, Hauber AB, Johnson FR, Coon CD. Patient preferences and linear scoring rules for patient-reported outcomes. Patient. 2010;3(4):217–27.

65. Johnson FR, Hauber AB, Osoba D, Hsu MA, Coombs J, Copley-Merriman C. Are chemotherapy patients' HRQoL importance weights consistent with linear scoring rules? A stated-choice approach. Qual Life Res. 2006;15(2):285–98.

66. Harrison M, Bansback N, Aguiar M, Koehn C, Shojania K, Finckh A, *et al.* Preferences for treatments to prevent rheumatoid arthritis in Canada and the influence of shared decision-making. Clin Rheumatol. 2020;39(10):2931–41.

67. Bansback N, Harrison M, Marra C. Does Introducing Imprecision around Probabilities for Benefit and Harm Influence the Way People Value Treatments? Medical Decision Making. 2016;36(4):490–502.

68. Mohamed A, Johnson F, Hauber A, Lescrauwaet B, Masterson A. Physicians' stated trade-off preferences for chronic hepatitis B treatment outcomes in Germany, France, Spain, Turkey, and Italy. European journal of gastroenterology & hepatology. 2012;24:419–26.

69. Hauber AB, Mohamed AF, Watson ME, Johnson FR, Hernandez JE. Benefits, Risk, and Uncertainty: Preferences of Antiretroviral-Naïve African Americans for HIV Treatments. AIDS Patient Care and STDs. 2008;23(1):29–34.

70. Fairchild AO, Katz EG, Reed SD, Johnson FR, Dibernardo AB, Hough DW, *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. 2020;37:67–78.

71. Johnson FR, Van Houtven G, Ozdemir S, Hass S, White J, Francis G, *et al.* Multiple sclerosis patients' benefit-risk preferences: serious adverse event risks versus treatment efficacy. J Neurol. 2009;256(4):554–62.

72. Johnson & Johnson Pharmaceutical Research & Development. Advisory Committee Briefing Document. Rivaroxaban for the Prevention of Stroke and Non-Central Nervous System (CNS). Systemic Embolism in Patients With Atrial Fibrillation. 5

August 2011. https://wayback.archive-it.org/7993/20170405212019/https:/www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/CardiovascularandRenalDrugsAdvisoryCommittee/UCM270797.pdf.

73. National Institute for Health and Care Excellence. Single Technology Appraisal. Betibeglogene autotemcel for treating transfusion-dependent beta-thalassaemia [ID968]. Committee Papers. 2021. https://www.nice.org.uk/guidance/gid-ta10334/documents/committee-papers.

74. Singer DE, Chang Y, Fang MC, Borowsky LH, Pomernacki NK, Udaltsova N, *et al.* The net clinical benefit of warfarin anticoagulation in atrial fibrillation. Ann Intern Med. 2009;151(5):297-305.

75. Hsu JC, Hsieh C-Y, Yang Y-HK, Lu CY. Net Clinical Benefit of Oral Anticoagulants: A Multiple Criteria Decision Analysis. PloS one. 2015;10(4):e0124806.

76. Barnett AS, Cyr DD, Goodman SG, Levitan BS, Yuan Z, Hankey GJ, *et al.* Net clinical benefit of rivaroxaban compared with warfarin in atrial fibrillation: Results from ROCKET AF. International Journal of Cardiology. 2018;257:78–83.

77. Pocock SJ, Ariti CA, Collier TJ, Wang D. 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):176–82.

78. Evans SR, Rubin D, Follmann D, Pennello G, Huskins WC, Powers JH, *et al.* Desirability of Outcome Ranking (DOOR) and Response Adjusted for Duration of Antibiotic Risk (RADAR). Clin Infect Dis. 2015;61(5):800–6.

79. Broekhuizen H, MJ IJ, Hauber AB, Groothuis-Oudshoorn CGM. Weighing Clinical Evidence Using Patient Preferences: An Application of Probabilistic Multi-Criteria Decision Analysis. Pharmacoeconomics. 2017;35(3):259–69.

80. Broekhuizen H, Groothuis-Oudshoorn CGM, Vliegenthart R, Groen HJM, MJ IJ. Assessing Lung Cancer Screening Programs under Uncertainty in a Heterogeneous Population. Value Health. 2018;21(11):1269–77.

81. Thokala P, Devlin N, Marsh K, Baltussen R, Boysen M, Kalo Z, *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 Health. 2016;19(1):1–13.

82. Baltussen R, Marsh K, Thokala P, Diaby V, Cleemput I, Garau M, *et al.* Multi-criteria decision analysis to support HTA agencies – benefits, limitations and the way forward. Value in Health. 2019:1283–8.

83. Broekhuizen H, Groothuis-Oudshoorn CG, van Til JA, Hummel JM, MJ IJ. A review and classification of approaches for dealing with uncertainty in multi-criteria decision analysis for healthcare decisions. Pharmacoeconomics. 2015;33(5):445–55.

84. U.S. Food & Drug Administration. Patient Preference Information (PPI) in Medical Device Decision-Making. 27 October 27 2020. https://www.fda.gov/about-fda/cdrh-patient-science-and-engagement-program/patient-preference-information-ppi-medical-device-decision-making. Accessed 23 April 2022.

85. U.S. Food & Drug Administration. Factors to Consider Regarding Benefit Risk in Medical Device Product Availability, Compliance, and Enforcement Decisions. Guidance for Industry and Food and Drug Administration Staff. 27 December 2016. https://www.fda.gov/media/98657/download.

86. European Commission. Enterprise and Industry Directorate-General. A Guideline on Summary of Product Characteristics (SmPC). September 2009. https://ec.europa.eu/health/system/files/2016-11/smpc_guideline_rev2_en_0.pdf.

87. eunethta: European Network for Health Technology Assessment. HTA Core Model. https://www.eunethta.eu/hta-core-model/. Accessed 23 April 2022.

88. Huls SPI, Whichello CL, van Exel J, Uyl-de Groot CA, de Bekker-Grob EW. What Is Next for Patient Preferences in Health Technology Assessment? A Systematic Review of the Challenges. Value Health. 2019;22(11):1318–28.

89. Janssens R, Russo S, van Overbeeke E, Whichello C, Harding S, Kübler J, *et al.* Patient Preferences in the Medical Product Life Cycle: What do Stakeholders Think? Semi-Structured Qualitative Interviews in Europe and the USA. Patient. 2019;12(5):513–26.

90. European Medicines Agency Committee for Human Medicinal Products. Guideline for good clinical practice E6(R2) Step 5. 1 December 2016. EMA/CHMP/ICH/135/1995. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-6-r2-guideline-good-clinical-practice-step-5_en.pdf.

91. Birch R, Simons G, Wähämaa H, McGrath CM, Johansson EC, Skingle D, *et al.* Development and formative evaluation of patient research partner involvement in a multi-disciplinary European translational research project. Research involvement and engagement. 2020;6:6.

92. Pollock J, Raza K, Pratt AG, Hanson H, Siebert S, Filer A, *et al.* Patient and researcher perspectives on facilitating patient and public involvement in rheumatology research. Musculoskeletal Care. 2017;15(4):395–9.

93. De Souza S, Williams R, Johansson E, Zabalan C, Esterine T, Bakkers M, *et al.* PARE0007 Patient and public involvement in clinical trial design. Annals of the Rheumatic Diseases. 2020;79(Suppl 1):1289–90.

94. Boudes M, Robinson P, Bertelsen N, Brooke N, Hamoir A, Sargeant I, *et al.* Understanding Stakeholders Expectations for Patient Engagement in Medicines Lifecycle: A Qualitative Survey. Poster presentation. 2017. http://patientfocusedmedicine.org/wp-content/uploads/2017/10/PFMD_UnderstandingStakeholdersExpectation.pdf.

95. European Medicines Agency. Benefit-risk methodology project Work package 1 report: description of the current practice of benefit-risk assessment for centralised procedure products in the EU regulatory network. Report by the EMA Benefit-Risk Methodology Project Team. 25 May 2011. https://www.ema.europa.eu/en/documents/report/benefit-risk-methodology-project-work-package-1-report-description-current-practice-benefit-risk_en.pdf.

96. European Medicines Agency. Regulatory and methodological standards to improve benefit-risk evaluation of medicines. Workshop report. 10 April 2014.

https://www.ema.europa.eu/en/documents/report/regulatory-methodological-standards-improve-benefit-risk-evaluation-medicines-workshop-report_en.pdf.

97. Tervonen T, Angelis A, Hockley K, Pignatti F, Phillips LD. Quantifying Preferences in Drug Benefit-Risk Decisions. Clin Pharmacol Ther. 2019;106(5):955–9.

98. European Medicines Agency. Regulatory Science: Research needs. 13 December 2021. https://www.ema.europa.eu/en/documents/other/regulatory-science-research-needs_en.pdf.

99. Beyer A, Fasolo B, Costa J, Bana e Costa C. Application of MACBETH to the elicitation of preferences for treatment outcomes. Working paper. January 2014. https://www.researchgate.net/publication/321013177_Application_of_MACBETH_to_the_elicitation_of_preferences_for_treatment_outcomes.

100. IMI PROTECT Benefit-Risk Group. Patient and Public Involvement Report version 1.0. Recommendations for Patient and Public Involvement in the assessment of benefit and risk of medicines. https://protectbenefitrisk.eu/documents/ppi150413v1.pdf.

101. Hockley KS. Patient and public involvement in the benefit-risk assessment of medicines and regulatory decision-making. 2013. https://core.ac.uk/download/pdf/76990632.pdf.

102. Chachoua L, Dabbous M, François C, Dussart C, Aballéa S, Toumi M. Use of Patient Preference Information in Benefit–Risk Assessment, Health Technology Assessment, and Pricing and Reimbursement Decisions: A Systematic Literature Review of Attempts and Initiatives. Frontiers in Medicine. 2020;7.

103. Kievit W, Tummers M, van Hoorn R, Booth A, Mozygemba K, Refolo P, *et al.* Taking patient heterogeneity and preferences into account in health technology assessments Int J Technol Assess Health Care. 2017;33(5):562–9.

104. European Medicines Agency. ICH reflection paper on proposed ICH guideline work to advance patient focused drug development. 10 December 2020. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-reflection-paper-proposed-ich-guideline-work-advance-patient-focused-drug-development_en.pdf.

105. Bauer P, König F. The risks of methodology aversion in drug regulation. Nature Reviews Drug Discovery. 2014;13(5):317–8.

106. Soekhai V, de Bekker-Grob EW, Ellis AR, Vass CM. Discrete Choice Experiments in Health Economics: Past, Present and Future. Pharmacoeconomics. 2019;37(2):201–26.

107. Clark MD, Determann D, Petrou S, Moro D, de Bekker-Grob EW. Discrete choice experiments in health economics: a review of the literature. Pharmacoeconomics. 2014;32(9):883–902.

108. de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. Health Econ. 2012;21(2):145–72.

109. Hauber BA, Fairchild AO, Reed Johnson F. Quantifying benefit-risk preferences for medical interventions: an overview of a growing empirical literature. Appl Health Econ Health Policy. 2013;11(4):319–29.

110. Al-Faruque F. FDA Weighs Patients' Risk Tolerance in Approving Obesity Device. 20 January 2015 https://pink.pharmaintelligence.informa.com/PS077790/FDA-Weighs-Patients-Risk-Tolerance-in-Approving-Obesity-Device. Accessed 23 April 2022.

111. Mühlbacher AC. Patient-centric HTA: different strokes for different folks. Expert Review of Pharmacoeconomics & Outcomes Research. 2015;15(4):591–7.

112. Johnson FR, Lancsar E, Marshall D, Kilambi V, Mühlbacher A, Regier DA, *et al.* Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. Value Health. 2013;16(1):3–13.

113. McFadden D. Quantitative methods for analyzing travel behaviour on individuals: some recent developments. Hensher D, Stopher P Behavioral Travel Modelling (1978) London: Croom Helm.

114. Van Houtven G, Johnson FR, Kilambi V, Hauber AB. Eliciting benefit-risk preferences and probability-weighted utility using choice-format conjoint analysis. Medical decision making : an international journal of the Society for Medical Decision Making. 2011;31(3):469–80.

115. Hauber AB, Tunceli K, Yang JC, Gantz I, Brodovicz KG, Alexander CM, *et al.* A survey of patient preferences for oral antihyperglycemic therapy in patients with type 2 diabetes mellitus. Diabetes Ther. 2015;6(1):75–84.

116. Hauber AB, Arden NK, Mohamed AF, Johnson FR, Peloso PM, Watson DJ, *et al.* A discrete-choice experiment of United Kingdom patients' willingness to risk adverse events for improved function and pain control in osteoarthritis. Osteoarthritis Cartilage. 2013;21(2):289–97.

117. Ryan M, Wordsworth S. Sensitivity of Willingness to Pay Estimates to the Level of Attributes in Discrete Choice Experiments. Scottish Journal of Political Economy. 2000;47(5):504–24.

118. Determann D, Gyrd-Hansen D, de Wit GA, de Bekker-Grob EW, Steyerberg EW, Lambooij MS, *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):681–92.

119. Hauber AB, González JM, Groothuis-Oudshoorn CG, Prior T, Marshall DA, Cunningham C, *et al.* Statistical Methods for the Analysis of Discrete Choice Experiments: A Report of the ISPOR Conjoint Analysis Good Research Practices Task Force. Value Health. 2016;19(4):300–15.

120. Bridges JF, Mohamed AF, Finnern HW, Woehl A, Hauber AB. Patients' preferences for treatment outcomes for advanced non-small cell lung cancer: a conjoint analysis. Lung Cancer. 2012;77(1):224–31.

121. Janssen EM, Hauber AB, Bridges JFP. Conducting a Discrete-Choice Experiment Study Following Recommendations for Good Research Practices: An Application for Eliciting Patient Preferences for Diabetes Treatments. Value Health. 2018;21(1):59–68.

122. Lancsar E, Fiebig DG, Hole AR. Discrete Choice Experiments: A Guide to Model Specification, Estimation and Software. Pharmacoeconomics. 2017;35(7):697–716.

123. de Bekker-Grob EW, Donkers B, Jonker MF, Stolk EA. Sample Size Requirements for Discrete-Choice Experiments in Healthcare: a Practical Guide. Patient. 2015;8(5):373–84.

124. Marshall D, Bridges JF, Hauber B, Cameron R, Donnalley L, Fyie K, *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. 2010;3(4):249–56.

125. Yang J-C, Johnson FR, Kilambi V, Mohamed AF. Sample size and utility-difference precision in discrete-choice experiments: A meta-simulation approach. Journal of choice modelling. 2015;16(C):50–7.

126. Soekhai V, Whichello C, Levitan B, Veldwijk J, Pinto CA, Donkers B, *et al.* Methods for exploring and eliciting patient preferences in the medical product lifecycle: a literature review. Drug Discovery Today. 2019;24(7):1324–31.

127. Tervonen T, Gelhorn H, Sri Bhashyam S, Poon JL, Gries KS, Rentz A, *et al.* MCDA swing weighting and discrete choice experiments for elicitation of patient benefit-risk preferences: a critical assessment. Pharmacoepidemiol Drug Saf. 2017;26(12):1483–91.

128. de Bekker-Grob EW, Donkers B, Bliemer MCJ, Veldwijk J, Swait JD. Can healthcare choice be predicted using stated preference data? Social science & medicine (1982). 2020;246:112736.

129. Angelis A, Kanavos P. Multiple Criteria Decision Analysis (MCDA) for evaluating new medicines in Health Technology Assessment and beyond: The Advance Value Framework. Social science & medicine (1982). 2017;188:137–56.

130. Dodgson J, Spackman M, Pearman A, Phillips L. Multi-criteria analysis: a manual. http://eprints.lse.ac.uk/12761/1/Multi-criteria_Analysis.pdf. London: Department for Communities and Local Government, 2009.

131. Medical Device Innovation Consortium. Patient Centered Benefit-Risk Project (PCBR). Appendix A: Catalog of Methods for Assessing Patient Preferences for Benefits and Harms of Medical Technologies. 22 April 2015. https://mdic.org/wp-content/uploads/2018/05/MDIC_PCBR_Framework_Web.pdf.

132. Mussen F, Salek S, Walker S. A quantitative approach to benefit-risk assessment of medicines - part 1: the development of a new model using multi-criteria decision analysis. Pharmacoepidemiol Drug Saf. 2007;16 Suppl 1:S2–S15.

133. Olson DL. Decision Aids for Selection Problems.: Springer New York; 1996.

134. Dolan JG. Multi-criteria clinical decision support: A primer on the use of multiple criteria decision making methods to promote evidence-based, patient-centered healthcare. Patient. 2010;3(4):229–48.

135. Marsh K, M IJ, Thokala P, Baltussen R, Boysen M, Kaló Z, *et al.* Multiple Criteria Decision Analysis for Health Care Decision Making--Emerging Good Practices: Report 2 of the ISPOR MCDA Emerging Good Practices Task Force. Value Health. 2016;19(2):125–37.

136. Keeney RL, Raiffa H. Decisions with Multiple Objectives: Preferences and Value Trade-Offs. Cambridge: Cambridge University Press; 1993.

137. Marsh K, Lanitis T, Neasham D, Orfanos P, Caro J. Assessing the value of healthcare interventions using multi-criteria decision analysis: a review of the literature. Pharmacoeconomics. 2014;32(4):345–65.

138. Gregory R, Failing L, Harstone M, Long G, McDaniels T, Ohlson D. Structured Decision Making: A Practical Guide to Environmental Management Choices: Wiley-Blackwell; 2012.

139. Edwards W. How to Use Multiattribute Utility Measurement for Social Decisionmaking. IEEE Transactions on Systems, Man, and Cybernetics. 1977;7:326–40.

140. Edwards W, Barron FH. SMARTS and SMARTER: Improved Simple Methods for Multiattribute Utility Measurement. Organizational Behavior and Human Decision Processes. 1994;60(3):306–25.

141. Diaby V, Campbell K, Goeree R. Multi-criteria decision analysis (MCDA) in health care: A bibliometric analysis. Operations Research for Health Care. 2013;2(1):20–4.

142. Mussen F, Salek S, Walker S. Appendix 3: Reflection Paper on Benefit–Risk Assessment Methods in the Context of the Evaluation of Marketing Authorisation Applications of Medicinal Products for Human Use – Committee for Medicinal Products for Human Use (March 2008). 2009. p. 233–49.

143. European Medicines Agency. Benefit-Risk Methodology Project. London, 12 March 2009. Doc. Ref. EMEA/108979/2009. https://www.ema.europa.eu/en/documents/report/benefit-risk-methodology-project_en.pdf.

144. European Medicines Agency. Benefit-risk methodology project. Work package 4 report: Benefit-risk tools and processes. 9 May 2012 EMA/297405/2012 – Revision 1 Human Medicines Development and Evaluation. https://www.ema.europa.eu/en/documents/report/benefit-risk-methodology-project-work-package-4-report-benefit-risk-tools-processes_en.pdf.

145. Marsh K, Zaiser E, Orfanos P, Salverda S, Wilcox T, Sun S*, et al.* Evaluation of COPD Treatments: A Multicriteria Decision Analysis of Aclidinium and Tiotropium in the United States. Value Health. 2017;20(1):132–40.

146. Vermersch P, Martinelli V, Pfleger C, Rieckmann P, Alonso-Magdalena L, Galazka A*, et al.* Benefit-risk Assessment of Cladribine Using Multi-criteria Decision Analysis (MCDA) for Patients With Relapsing-remitting Multiple Sclerosis. Clin Ther. 2019;41(2):249–60.e18.

147. Felli JC, Noel RA, Cavazzoni PA. A multiattribute model for evaluating the benefit-risk profiles of treatment alternatives. Medical decision making : an international journal of the Society for Medical Decision Making. 2009;29(1):104–15.

148. van Valkenhoef G, Tervonen T, Zhao J, de Brock B, Hillege HL, Postmus D. Multicriteria benefit-risk assessment using network meta-analysis. J Clin Epidemiol. 2012;65(4):394–403.

149. Rutten-van Mölken M, Leijten F, Hoedemakers M, Tsiachristas A, Verbeek N, Karimi M, *et al.* Strengthening the evidence-base of integrated care for people with multi-morbidity in Europe using Multi-Criteria Decision Analysis (MCDA). BMC Health Services Research. 2018;18(1):576.

150. Leijten F, Tsiachristas A, Boland M, Hoedemakers M, Verbeek N, Islam K, *et al.* Development of an analytical framework to perform a comprehensive evaluation of integrated care programmes for multi-morbidity using Multi-Criteria Decision Analysis. Deliverable report 2 MCDA framework. 31 January 2017. https://www.selfie2020.eu/wp-content/uploads/2018/05/SELFIE-WP4-Deliverable-Report-2.pdf.

151. Nutt DJ, King LA, Phillips LD. Drug harms in the UK: a multicriteria decision analysis. The Lancet. 2010;376(9752):1558–65.

152. Airoldi M, Morton A, Smith J, Bevan G. Healthcare prioritisation at the local level: a socio-technical approach. Priority Setting for Population Health. Working paper series. 2011. http://eprints.lse.ac.uk/59892/1/__lse.ac.uk_storage_LIBRARY_Secondary_libfile_shared_repository_Content_Bevan%2C%20G_Bevan_Healthcare_prioritisation_local_2011_Bevan_Healthcare_prioritisation_local_author.pdf.

153. van Amsterdam J, Nutt D, Phillips L, van den Brink W. European rating of drug harms. J Psychopharmacol. 2015;29(6):655–60.

154. Nutt DJ, Phillips LD, Balfour D, Curran HV, Dockrell M, Foulds J, *et al.* Estimating the harms of nicotine-containing products using the MCDA approach. Eur Addict Res. 2014;20(5):218–25.

155. Marsh K, Hawken N, Brookes E, Kuehn C, Liden B. Patient-centered benefit-risk analysis of transcatheter aortic valve replacement. F1000Res. 2019;8:394.

156. SriBhashyam S, Marsh K, Quartel A, Weng HH, Gershman A, Longo N, *et al.* A benefit-risk analysis of pegvaliase for the treatment of phenylketonuria: A study of patients' preferences. Mol Genet Metab Rep. 2019;21:100507.

157. Kaltoft MK, Nielsen JB, Dowie J. Dual Purpose, Dual Audience: MCDA-Based Tools Can Simultaneously Support Personal Health Decisions and Educate Persons and Clinicians. Stud Health Technol Inform. 2018;255:257–61.

158. Wagner M, Samaha D, Cuervo J, Patel H, Martinez M, O'Neil WM, *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):1215–31.

159. Tervonen T, Naci H, van Valkenhoef G, Ades AE, Angelis A, Hillege HL, *et al.* Applying Multiple Criteria Decision Analysis to Comparative Benefit-Risk Assessment: Choosing among Statins in Primary Prevention. Medical decision making : an international journal of the Society for Medical Decision Making. 2015;35(7):859–71.

160. Weber M, Borcherding K. Behavioral influences on weight judgments in multiattribute decision making. European Journal of Operational Research. 1993;67(1):1–12.

161. Phillips LD. Best Practice for MCDA in Healthcare. In: Marsh K, Goetghebeur M, Thokala P, Baltussen R, editors. Multi-Criteria Decision Analysis to Support Healthcare Decisions. Cham: Springer International Publishing; 2017. p. 311–29.

162. Marsh K, Thokala P, Mühlbacher A, Lanitis T. Incorporating Preferences and Priorities into MCDA: Selecting an Appropriate Scoring and Weighting Technique. In: Marsh K, Goetghebeur M, Thokala P, Baltussen R, editors. Multi-Criteria Decision Analysis to Support Healthcare Decisions. Cham: Springer International Publishing; 2017. p. 47–66.

163. IMI PREFER. Diabetes. 2020. https://www.imi-prefer.eu/case-studies/diabetes/.

164. Keeney RL. Creativity in Decision Making with Value-Focused Thinking. Sloan Management Review. 1994;35(4):33–41.

165. 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.

166. Lahdelma R, Salminen P. SMAA-2: Stochastic Multicriteria Acceptability Analysis for Group Decision Making. Operations Research. 2001;49(3):444–54.

167. Tervonen T, Figueira J. A Survey on Stochastic Multicriteria Acceptability Analysis Methods. Journal of Multi-Criteria Decision Analysis. 2008;15:1–14.

168. Tervonen T, van Valkenhoef G, Buskens E, Hillege HL, Postmus D. A stochastic multicriteria model for evidence-based decision making in drug benefit-risk analysis. Stat Med. 2011;30(12):1419–28.

169. Drug Information Systems. Transparent and efficient health care decision making. https://www.drugis.org/index.

170. van Valkenhoef G, Tervonen T, Zwinkels T, de Brock B, Hillege H. ADDIS: A decision support system for evidence-based medicine. Decision Support Systems. 2013;55(2):459–75.

171. Catalyze. Hiview3. http://www.catalyzeconsulting.com/software/hiview3/. Accessed 23 April 2022.

172. IMI PREFER. Lung cancer. 2020. https://www.imi-prefer.eu/case-studies/lung-cancer/. Accessed 23 April 2022.

173. Goetghebeur MM, Wagner M, Khoury H, Levitt RJ, Erickson LJ, Rindress D. Bridging health technology assessment (HTA) and efficient health care decision making with multicriteria decision analysis (MCDA): applying the EVIDEM framework to medicines appraisal. Medical decision making : an international journal of the Society for Medical Decision Making. 2012;32(2):376–88.

174. Durbach I, Stewart T. Using expected values to simplify decision making under uncertainty. Omega. 2009;37:312–30.

175. Stewart T. Simplified approach for multicriteria decision making under uncertainty. Journal of Multi-Criteria Decision Analysis. 1995;4:246–58.

176. Keeney RL. Common Mistakes in Making Value Trade-Offs. Operations Research. 2002;50(6):935–45.

177. Finn A, Louviere JJ. Determining the Appropriate Response to Evidence of Public Concern: The Case of Food Safety. Journal of Public Policy & Marketing. 1992;11(2):12–25.

178. Louviere JJ, Flynn TN, Marley AAJ. Best-Worst Scaling: Theory, Methods and Applications, Cambridge University Press (Cambridge). 2015.

179. Whitty JA, Oliveira Gonçalves 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;11(3):301–17.

180. Burton N, Burton M, Rigby D, Sutherland CAM, Rhodes G. Best-worst scaling improves measurement of first impressions. Cogn Res Princ Implic. 2019;4(1):36.

181. Cheung KL, Wijnen BF, Hollin IL, Janssen EM, Bridges JF, Evers SM, *et al.* Using Best-Worst Scaling to Investigate Preferences in Health Care. Pharmacoeconomics. 2016;34(12):1195–209.

182. Yu T, Holbrook JT, Thorne JE, Flynn TN, Van Natta ML, Puhan MA. Outcome Preferences in Patients With Noninfectious Uveitis: Results of a Best-Worst Scaling Study. Invest Ophthalmol Vis Sci. 2015;56(11):6864–72.

183. Yuan Z, Levitan B, Burton P, Poulos C, Brett Hauber A, 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;30(9):1733–41.

184. Gallego G, Bridges JF, Flynn T, Blauvelt BM, Niessen LW. Using best-worst scaling in horizon scanning for hepatocellular carcinoma technologies. Int J Technol Assess Health Care. 2012;28(3):339–46.

185. Marti J. A best-worst scaling survey of adolescents' level of concern for health and non-health consequences of smoking. Social science & medicine (1982). 2012;75(1):87–97.

186. Cozmuta R, Merkel PA, Wahl E, Fraenkel L. Variability of the impact of adverse events on physicians' decision making. BMC Med Inform Decis Mak. 2014;14:86.

187. Husni ME, Fernandez A, Hauber B, Singh R, Posner J, Sutphin J, *et al.* Comparison of US patient, rheumatologist, and dermatologist perceptions of psoriatic disease symptoms: results from the DISCONNECT study. Arthritis Res Ther. 2018;20(1):102.

188. Hauber B, Caloyeras J, Posner J, Brommage D, Tzivelekis S, Pollock A. Hemodialysis patients' preferences for the management of anemia. BMC Nephrology. 2017;18(1):253.

189. Silverman S, Calderon A, Kaw K, Childers TB, Stafford BA, Brynildsen W, *et al.* Patient weighting of osteoporosis medication attributes across racial and ethnic groups: a study of osteoporosis medication preferences using conjoint analysis. Osteoporosis International. 2013;24(7):2067–77.

190. Aizaki H, Nakatani T, Sato K. Stated Preference Methods Using R. 1st ed: Chapman and Hall/CRC; 2014.

191. Cheung KL, Mayer S, Simon J, de Vries H, Evers S, Kremer IEH*, et al.* Comparison of statistical analysis methods for object case best-worst scaling. J Med Econ. 2019;22(6):509–15.

192. McFadden D, Train K. Mixed MNL models for discrete response. Journal of Applied Econometrics. 2000;15(5):447–70.

193. Imaeda A, Bender D, Fraenkel L. What is most important to patients when deciding about colorectal screening? J Gen Intern Med. 2010;25(7):688–93.

194. Torbica A, Fattore G, Ayala F. Eliciting preferences to inform patient-centred policies: the case of psoriasis. Pharmacoeconomics. 2014;32(2):209–23.

195. 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.

196. Ross M, Bridges JF, Ng X, Wagner LD, Frosch E, Reeves G*, et al.* A best-worst scaling experiment to prioritize caregiver concerns about ADHD medication for children. Psychiatr Serv. 2015;66(2):208–11.

197. Harrell Jr FE. Regression Modeling Strategies With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis. Second Edition. Sringer Series in Statistics. 2015.

198. Hess A, A D. Handbook of Choice Modelling: Edward Elgar Publishing; 2014.

199. Flynn TN, Huynh E, Peters TJ, Al-Janabi H, Clemens S, Moody A*, et al.* Scoring the Icecap-a capability instrument. Estimation of a UK general population tariff. Health Econ. 2015;24(3):258–69.

200. Mansfield C, Ndife B, Chen J, Gallaher K, Ghate S. Patient preferences for treatment of metastatic melanoma. Future Oncol. 2019;15(11):1255–68.

201. Szeinbach SL, Barnes JH, McGhan WF, Murawski MM, Corey R. Using Conjoint Analysis to Evaluate Health State Preferences. Drug Information Journal. 1999;33(3):849–58.

202. Ratcliffe J, Stevens K, Flynn T, Brazier J, Sawyer M. An assessment of the construct validity of the CHU9D in the Australian adolescent general population. Qual Life Res. 2012;21(4):717–25.

203. Coast J, Flynn TN, Natarajan L, Sproston K, Lewis J, Louviere JJ*, et al.* Valuing the ICECAP capability index for older people. Social science & medicine (1982). 2008;67(5):874–82.

204. Dalziel K, Catchpool M, García-Lorenzo B, Gorostiza I, Norman R, Rivero-Arias O. 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;38(5):499–513.

205. Potoglou D, Burge P, Flynn T, Netten A, Malley J, Forder J*, et al.* Best-worst scaling vs. discrete choice experiments: an empirical comparison using social care data. Social science & medicine (1982). 2011;72(10):1717–27.

206. Severin F, Schmidtke J, Mühlbacher A, Rogowski WH. Eliciting preferences for priority setting in genetic testing: a pilot study comparing best-worst scaling and discrete-choice experiments. Eur J Hum Genet. 2013;21(11):1202–8.

207. Knox SA, Viney RC, Street DJ, Haas MR, Fiebig DG, Weisberg E*, et al.* What's good and bad about contraceptive products?: a best-worst attribute experiment comparing the values of women consumers and GPs. Pharmacoeconomics. 2012;30(12):1187–202.

208. Flynn TN. Valuing citizen and patient preferences in health: recent developments in three types of best-worst scaling. Expert Rev Pharmacoecon Outcomes Res. 2010;10(3):259–67.

209. Whitty JA, Walker R, Golenko X, Ratcliffe J. A think aloud study comparing the validity and acceptability of discrete choice and best worst scaling methods. PloS one. 2014;9(4):e90635.

210. Flynn TN, Louviere JJ, Peters TJ, Coast J. Estimating preferences for a dermatology consultation using Best-Worst Scaling: Comparison of various methods of analysis. BMC medical research methodology. 2008;8(1):76.

211. Ryan M, Krucien N, Hermens F. The eyes have it: Using eye tracking to inform information processing strategies in multi-attributes choices. Health Econ. 2018;27(4):709–21.

212. 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):20–7.

213. Ratcliffe J, Huynh E, Chen G, Stevens K, Swait J, Brazier J*, et al.* Valuing the Child Health Utility 9D: Using profile case best worst scaling methods to develop a new adolescent specific scoring algorithm. Social science & medicine (1982). 2016;157:48–59.

214. Hollin IL, Dimmock AE, Bridges JF, Danoff SK, Bascom R. Collecting patient preference information using a Clinical Data Research Network: demonstrating feasibility with idiopathic pulmonary fibrosis. Patient preference and adherence. 2019;13:795–804.

215. Shields GE, Brown L, Wells A, Capobianco L, Vass C. Utilising Patient and Public Involvement in Stated Preference Research in Health: Learning from the Existing Literature and a Case Study. The Patient - Patient-Centered Outcomes Research. 2021;14(4):399–412.

216. Mühlbacher AC, Kaczynski A. Making Good Decisions in Healthcare with Multi-Criteria Decision Analysis: The Use, Current Research and Future Development of MCDA. Appl Health Econ Health Policy. 2016;14(1):29–40.

217. Hauber B, Coulter J. 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;18(1):31–46.

218. Devereaux PJ, Anderson DR, Gardner MJ, Putnam W, Flowerdew GJ, Brownell BF*, et al.* Differences between perspectives of physicians and patients on anticoagulation in patients with atrial fibrillation: observational study. Bmj. 2001;323(7323):1218–22.

219. van Overbeeke E, Michelsen S, Hauber B, Peerlinck K, Hermans C, Lambert C*, et al.* Patient perspectives regarding gene therapy in haemophilia: Interviews from the PAVING study. Haemophilia. 2021;27(1):129–36.

220. Landrum Peay H, Fischer R, Tzeng JP, Hesterlee SE, Morris C, Strong Martin A*, et al.* Gene therapy as a potential therapeutic option for Duchenne muscular dystrophy: A qualitative preference study of patients and parents. PloS one. 2019;14(5):e0213649.

221. Tomlinson D, Bartels U, Gammon J, Hinds PS, Volpe J, Bouffet E*, et al.* Chemotherapy versus supportive care alone in pediatric palliative care for cancer: comparing the preferences of parents and health care professionals. Cmaj. 2011;183(17):E1252–8.

222. Robinson A, Spencer A, Moffatt P. A framework for estimating health state utility values within a discrete choice experiment: modeling risky choices. Medical decision making : an international journal of the Society for Medical Decision Making. 2015;35(3):341–50.

223. Rodríguez-Míguez E, Pinto-Prades JL, Mosquera-Nogueira J. Eliciting Health State Utilities Using Paired-Gamble Methods: The Role of the Starting Point. Value in Health. 2019;22(4):446–52.

224. Percy ME, Llewellyn-Thomas H. Assessing Preferences about the DNR Order:Does It Depend on How You Ask? Medical Decision Making. 1995;15(3):209–16.

225. Brundage MD, Davidson JR, Mackillop WJ. Trading treatment toxicity for survival in locally advanced non-small cell lung cancer. J Clin Oncol. 1997;15(1):330–40.

226. Kennedy ED, To T, Steinhart HA, Detsky A, Llewellyn-Thomas HA, McLeod RS. Do patients consider postoperative maintenance therapy for Crohn's disease worthwhile? Inflammatory Bowel Diseases. 2007;14(2):224–35.

227. Mørkbak MR, Olsen SB. A within-sample investigation of test–retest reliability in choice experiment surveys with real economic incentives. Australian Journal of Agricultural and Resource Economics. 2015;59(3):375–92.

228. Dales RE, O'Connor A, Hebert P, Sullivan K, McKim D, Llewellyn-Thomas H. Intubation and mechanical ventilation for COPD: development of an instrument to elicit patient preferences. Chest. 1999;116(3):792–800.

229. Sung L, Feldman BM, Schwamborn G, Paczesny D, Cochrane A, Greenberg ML*, et al.* Inpatient versus outpatient management of low-risk pediatric febrile neutropenia: measuring parents' and healthcare professionals' preferences. J Clin Oncol. 2004;22(19):3922–9.

230. Steures P, Berkhout JC, Hompes PG, van der Steeg JW, Bossuyt PM, van der Veen F*, et al.* Patients' preferences in deciding between intrauterine insemination and expectant management. Hum Reprod. 2005;20(3):752–5.

231. Gupta AA, Donen RM, Sung L, Boydell KM, Lo KC, Stephens D*, et al.* Testicular Biopsy for Fertility Preservation in Prepubertal Boys with Cancer: Identifying Preferences for Procedure and Reactions to Disclosure Practices. J Urol. 2016;196(1):219–24.

232. Kattan MW. Test-Treatment Threshold. 2009. In: Encyclopedia of Medical Decision Making [Internet]. Thousand Oaks, California: SAGE Publications, Inc.

233. Kahneman D, Tversky A. Prospect Theory: An Analysis of Decision under Risk. Econometrica. 1979;47(2):263–91.

234. Whichello C, Levitan B, Juhaeri J, Patadia V, DiSantostefano R, Pinto CA, *et al.* Appraising patient preference methods for decision-making in the medical product lifecycle: an empirical comparison. BMC Medical Informatics and Decision Making. 2020;20(1):114.

This document should be cited as follows:

The PREFER consortium (2022). PREFER Recommendations: Why, when and how to assess and use patient preferences in medical product decision-making. Zenodo. DOI: 10.5281/zenodo.6470922