Preference Testing in Medical Devices: Current Framework and Regulatory Gaps

Amy Lewis, Despoina Douka, Angeliki Koukoura, Vasiliki Valla, Amie Smirthwaite, Susanne Holm Faarbaek, Efthathios Vassiliadis

1 Evnia, Copenhagen, Denmark; 2 RQM+, Nottingham, England; 3 BK Medical, Herlev, Denmark

Abstract: Preference testing is a valuable source of information that can be provided by both healthcare professionals (HCPs) and patients (users). It can be used to improve the design and development of medical devices by feeding into device usability and, ultimately, risk management. Furthermore, it can aid with selecting the most appropriate clinical endpoints to be used in the clinical evaluation of a device and increase patient engagement by incorporating patient-relevant outcomes. Preference testing is widely conducted in the food industry but is not widespread in the medical field due to limited guidelines and a lack of regulatory framework. As such, manufacturers may be unaware of the benefits of preference testing and fail to take full advantage of it, or conversely, may use inappropriate methodology and/or analyses and consequently fail to collect meaningful data. In this position paper, we aim to highlight the benefits and uses of preference testing, along with potential methods that could be used for preference testing of medical devices. A key step towards the wider implementation of preference testing in medical devices is the publication of international standards and guidelines for the collection, assessment, and implementation of preference data into the life cycle of a medical device.

Keywords: preference testing, medical device, PPI, patient preference

Preference Testing

Preference testing refers to a research method in which a consumer/user indicates the most liked (preferred) product, usually from a choice of two. Preference testing is recommended when information on choice is required or the ordinal relationship between two products needs to be established. In the context of medical device design, this could be used to address preferences for different design features (e.g., size, shape, weight), usability factors (interaction between a device and its user(s)), or clinical endpoints (e.g., efficacy, side effect, etc.). This can include both HCP and patient preferences, for example, preference testing could be used to establish the clinical outcomes/benefits that a patient would like to experience (e.g., improved quality of life). Common methods include paired-preference testing, best-worst scaling, and discrete choice experiments (see Table 1). An overview of the advantages and disadvantages of preference testing is provided in Table 2.

Acceptance is related to preference testing and describes the overall liking or acceptability of a product and can be measured in terms of behavior, such as the purchasing of a product. Acceptance testing is recommended when information on the magnitude of liking/disliking of a product is required, and to establish the hedonic status (overall liking or acceptability) of two products. Common methods include 9-point hedonic, labelled affective magnitude and unstructured line scales (see for a description). Acceptance testing in this instance should not be confused with User Acceptance Testing (UAT) that may occur as part of validation and verification activities required for a medical device to conform with the relevant regulations. UAT is used in various research settings, for example, in clinical trials/investigations to test and verify electronic Case Report Forms (eCRFS) and the databases that store them to ensure that they are fit for purpose (i.e., that they store and collect data as intended).

Although preference and acceptability are clearly related, they are not interchangeable terms as a product can be acceptable yet not preferred over an alternative product, thus both types of information may be tested independently. For example,
### A. Qualitative Preference Testing Methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
<th>Example (Health Industry Related)</th>
</tr>
</thead>
</table>
| Delphi Method        | Primarily used by groups of experts to gain consensus through controlled feedback when there is limited or insufficient knowledge/evidence on the topic of discussion. | - Creation of guidelines and/or treatment protocols (eg creation of design guidance tools – Safety Risk Assessment for healthcare settings; The Center for Health Design, 2017).<sup>53</sup>  
- Consensus on recommendations for the use of fluorodeoxyglucose positron emission tomography (FDG-PET) in the detection of dementing neurodegenerative disorders.<sup>19</sup> |
| Dyadic Interview     | Two-person interviews that combine the advantages of individual and focus groups interviews by assisting the researcher to formulate favorable circumstances for social interaction and depth. | Photovoice dyadic interviews in people with dementia to evaluate the well-being of the patients (eg see studies from Wang and Burris, 1997).<sup>54</sup>                                                                                                      |
| Concept Mapping      | Concept mapping is a mixture of methods that involve brainstorming and unstructured sorting via multiple expert sources and is analyzed by multidimensional scaling and/or hierarchical cluster analysis statistical methods. | Concept mapping can be used to guide a) action planning, b) program development, and c) evaluation procedures and/or measurements for the health sector:<sup>55</sup>  
- Development of conceptual frameworks in depression, anxiety, Quality of Life/Care (QoL/QoC).  
- Development of guidelines in public health management (eg public health management of lower prevalence chronic conditions such as epilepsy).  
- Stakeholder’s perspective identification in Traumatic Brain Injury (TBI), primary health care.  
- Health planning/evaluation in mental health/illness, telemedicine in child protective services. |

### B. Quantitative Preference Testing Methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
<th>Example (Health Industry Related)</th>
</tr>
</thead>
</table>
| Discrete Choice Experiment (DCE) | DCEs assess preferences and choices via examining various constituents of treatments, services, and trade-offs among involved stakeholders.<sup>24,26</sup> | DCEs have been broadly accepted and used in pharmaco-economics and outcomes research that involve evaluation of patient preferences for treatments, drug comparison, biologics, and medical devices:<sup>19</sup>  
- Study on the benefits and risks of rofecoxib relative to naproxen, 2010.  
- Johnson & Johnson study on rivaroxaban for the prophylaxis of deep vein thrombosis (DVT) and pulmonary embolism (PE) in patients undergoing hip or knee replacement surgery, 2009. |
| Best-worst Scaling (BWS) | Types of Surveys to evaluate lists of attributes and/or profiles to signify the most or least appealing between them.<sup>30,56</sup> | BWS can serve as a valuable tool to identify patients’/HCPs preferences for health related QoL.<sup>34,57</sup>  
- Pilot BWS study as an empirical example conducted in 2005 to evaluate the QoL of elderly people (aged >65). |
| Q-Methodology          | A rank-ordering method (agree-disagree) that associates the assets of both qualitative and quantitative methods and aims to appraise subject’s preferences based on a “forced normal distribution”.<sup>58</sup> | Q-Methodology may act as a beneficial apparatus to form the basis for examining attitudes (attributes that form cognitive relationships and subsequently behavior) within the fields of health education and promotion.<sup>58</sup> |
| Control Preference Scale (CPS) | CPS as a method utilizes the Unfolding statistical Theory (determination of preference distribution) and is applied to assess patient preferences regarding the desired control a patient aims to have over their medical treatment.<sup>59</sup> | CPS is broadly used in health care settings to evaluate the preferences of patients that deal with life-threatening diseases (eg breast cancer, prostate cancer, etc.).<sup>59</sup>  
- Henrikson et al, exploratory interview study among male patients suffering from localized prostate cancer, 2011. |

(Continued)
a manufacturer may wish to use acceptance testing to establish if their product is liked by the intended user and to modify the product accordingly, whereas preference testing could be used to gather information on how their product compares to a competitor product (i.e., if their product is the preferred product). In this article, we will focus on preference testing, specifically, patient preference testing (PPI, defined below), due to recent discussions surrounding PPI within the medical field.

Table 1 (Continued).

<table>
<thead>
<tr>
<th>Quantitative Preference Testing Methods</th>
<th>Example (Health Industry Related)</th>
</tr>
</thead>
</table>
| Standard Gamble (SG)                   | Estimation/measurement of cardinal preferences that are related to chronic or temporary health outcomes.\(^{60}\) SG provides the ability to measure multiple health states to generate testable comparisons among different populations and clinical settings.\(^ {60}\)  
  - Yen et al, study on a Cost-Effectiveness Analysis of Treatment Strategies for Acute Uncomplicated Pyelonephritis in Women, 2003.\(^ {61}\) |
| Threshold Technique (TT)               | A patient preference method that detects the most amount of change a patient will accept in one attribute to achieve a change in another attribute.\(^ {40, 62}\) TT is broadly used in empirical health related studies to provide evidence on patient/HCP preferences and the accuracy of these preferences in terms of their results and estimates.\(^ {62}\)  
  - Devereaux et al, prospective observational study on differences between perspectives of physicians and patients on anticoagulation in patients with atrial fibrillation, 2001 (TT used to evaluate patient/HCP preferences towards the reduction of stroke risk).\(^ {63}\) |
| Swing Weighting (SW)                   | Statistical preference method that evaluates attribute decision changes through the swinging effect of attributes and their assigned weights.\(^ {40, 64}\) SW is used to obtain patient preferences based on treatment benefit-risk trade-offs (attributes included may be the clinical benefit, adverse events, convenience, etc.).\(^ {44}\)  
  - Tervonen et al, empirical study on swing weighting and discrete choice experiments for elicitation of patient benefit-risk preferences. |
| Paired Preference Test                 | Paired preference tests examine statistically significant preferences on specified sensory attributes among two options/products for a given population.\(^ {43}\) Paired comparisons assist with identifying patient preferences in multiple clinical settings:  
  - Sutherland et al, study on measuring satisfaction with health care: a comparison of single with paired rating strategies, 1989.\(^ {65}\) |

Table 2 Advantages and Disadvantages of Preference Testing\(^ {1}\)

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preference tests are simple to set up. Respondents can easily understand the nature of the task. Risk factors associated with preferences can be more precisely determined before test. Preference measures are likely to be seen as more relevant to consumers. Preference is a criterion-free measure.</td>
<td>Actual liking or disliking or both of the product in preference is unknown. Decision about “no preference” must be made. Paired comparison data are not particularly useful for product development because they give no real guidance on what is liked or disliked about a product (only which product is preferred).</td>
</tr>
</tbody>
</table>

Note: Adapted from E2943-15, A. Standard guide for two-sample acceptance and preference testing with consumers; 2021; Available from: https://www.astm.org/e2943-15r21.html. Accessed June 16, 2022.\(^ {1}\)
Preference Testing in Practice
Preference testing is used within a variety of fields, such as insurance, public policy, and product development/design. For example, preference testing is widely used within the food industry (e.g., in confectionery and meat) and is often referred to as Consumer Preference Testing (CPT). It involves measuring consumers’ preferences of sensory properties of food (appearance, odor, texture, and flavor) and can be used during product development, improvement, quality control, storage considerations, and process development.

Preference testing has recently become a topic of discussion in the medical field, particularly within pharmaceuticals. Preference testing can be used to obtain Patient Preference Information (PPI), described by the USA Food and Drug Administration (FDA) as

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.

Thus, PPI describes how much one outcome/intervention is valued by a patient in comparison to alternative options, and measures the benefit-risk trade-offs that patients are willing to accept.

In the USA, the FDA has provided guidance on the use of PPI in the design and conduct of medical device clinical investigations, acknowledging the benefit that patient preferences can provide. For example, it is proposed that patient input can help with easier/faster recruitment in clinical investigations, with a reduction in drop-out rates and increased follow-up participation. Moreover, the data that is collected can be of greater relevance to the outcomes that matter to patients. The Center for Devices and Radiological Health (CDRH) has played a pivotal role in highlighting the importance of preference testing in medical devices by using PPI in their regulatory decision-making, resulting in the approval of the first device to treat obesity. The device (Maestro Rechargeable System) failed to meet its clinical trial endpoint of a 10% weight loss; however, a survey of patient preferences revealed patients were willing to accept the associated risks of using the device for the expected amount of weight loss the device could provide. Consequently, it was determined (based on the clinical trial, an expert panel, and patient preferences) that the benefit of the device outweighed the risks. Indeed, the use of PPI is now part of the 21st Century Cures Act (enacted in 2016). The Act requires that the FDA provides guidance on the collection of patient experience data, defined as “information about the impact of a medical condition or related therapy on a patient and the patient’s preferences for treatment”. The FDA must also publish a statement and its review of any patient experience data used in support of an application.

The Regulatory Gap in Preference Testing of Medical Devices
Currently, it is uncommon for patient and/or HCP preferences to be accounted for within a medical device’s lifecycle, and even if they are, the lack of standards and regulations on how preference information should be collected, analyzed, and used within decision processes are absent. Within Europe, regulation for the use of preference information within the medical field is non-existent, even in the recently enforced Medical Device Regulation (2017/745); however, there is ongoing discussion on how preference information could be used to enable regulatory decisions, mainly with respect to reimbursement and access to medical products. The PREFER project, a 5-year public-private collaborative project funded by the Innovative Medicines Initiative, aims to establish recommendations for best-practices for including PPI throughout the medical treatment lifecycle (including medicinal products, devices and services). The PREFER project highlights four key areas that need to be addressed by all involved stakeholders (e.g., industry, regulatory authorities, HCP, etc.) if PPI is to be used to improve medical treatments for patients:

1. Consensus/shared understanding of the most appropriate methodology for patient preference studies.
2. Consensus/shared understanding of how patient preference studies can be used to inform decision processes within industry, regulatory authorities, HTA, and reimbursement agencies.
3. Consensus/shared understanding on the best time-point to collect PPI in the product’s lifecycle.
4. The lack of information on how to conduct patient preference studies that involve different patient populations, acute and chronic diseases, and both prevalent and uncommon diseases.
Similarly, Whichello et al\textsuperscript{15} propose that recommendations are needed for all stakeholders (eg industry, regulatory authorities, HCP, etc.) on how to collect PPI, who should best conduct patient preference studies and how results should be interpreted.\textsuperscript{15} Furthermore, reimbursement/HTA and regulatory authorities must decide the weight that is given to PPI in comparison to the information required by regulations.\textsuperscript{15}

In August 2021, the PREFER project proposed a framework for the implementation of patient preferences in decision-making, which is currently under review by the European Medicines Agency (EMA) and HTA bodies (EUnetHTA).\textsuperscript{17} The aim of the framework is to address stakeholders’ concern over a lack of clear and practical framework for collecting information on patient preferences. The framework covers advice on study purpose and objectives, design and conduct, and method selection (including Discrete Choice Experiment, Swing weighting, Best-worse scaling and Threshold Technique, see Table 1), as well as how the collected PPI can be used to inform decision-making.\textsuperscript{17,18}

In the United States (U.S.), the process of incorporating PPI into regulatory decision-making is a step ahead. An FDA guidance document is available on PPI and recommendations for patient preference testing (see Ref\textsuperscript{10} and Table 3). This guidance encourages the collection and submission of PPI for devices using the premarket approval, humanitarian device exemption (HDE), or de novo classification pathway, and emphasizes that PPI should be considered when evaluating the benefit-risk profile of a device.\textsuperscript{10} The FDA guidance also provides a list of cases in which PPI may be particularly beneficial, for instance, for devices that have a direct patient interface or devices that are intended to directly affect health-related quality of life, amongst others.\textsuperscript{10}

Additional guidance is available from the Patient Centered Benefit-Risk framework developed by the Medical Device Innovation Consortium (MDIC).\textsuperscript{19} This guidance focuses on the use of PPI to ensure a patient-centered approach for determining risk-benefit assessment (ie to identify risks and benefits that are most important to patients), whilst also emphasizing that PPI should be an additional source of information and should not replace clinical data on safety and performance\textsuperscript{19} (for a detailed review see\textsuperscript{20}). Both these guidance documents are geared towards patient preferences and not patient and HCP preferences, as the FDA believes that usability testing is more suitable for obtaining HCP preferences; however, the FDA will also consider the preferences of care-partners and HCPs if they are relevant to the benefit-risk assessment.\textsuperscript{10}

Despite the availability of guidance documents in the US, PPI data collection is not mandatory and is very much still an evolving area of activity.\textsuperscript{10} Moreover, preference testing is exempt from FDA medical device regulation in accordance with 21 CFR 812.2 (b), so long as the purpose is not for determining safety and effectiveness of the device and there is no risk to the subjects:

A device undergoing consumer preference testing, testing of a modification, or testing of a combination of two or more devices in commercial distribution, if the testing is not for the purpose of determining safety or effectiveness and does not put subjects at risk. (21 CFR Part 812.2 Investigational Device Exemptions)

For preference testing of food products, relevant international standards are available (eg ISO 13299:2016: Sensory analysis – Methodology – General guidance for establishing a sensory profile), yet standards for preference testing of medical devices is limited, with standards relating only to preference testing methodology, namely, two-sample acceptance and preference testing (ASTM E2943) and Standard Test Method for Paired Preference Test (ASTM E2263). These standards, although useful for the collection of preference information, do not provide information on how the collected data should be analyzed and used within the decision-making processes throughout a medical device’s lifecycle. This means that manufacturers may use collected preference information in any way that they see fit (provided they can provide justification for the methods used), regardless of whether the analysis they chose is the most valid and reliable method. Consequently, any preference information that is used may not accurately reflect the true preferences of the patient and/or HCP and may result in unnecessary design changes.

Preference Testing: Design, Usability, and Risks
A key implication of the lack of regulatory framework and guidance on preference testing is that the link between preference testing, usability and risk is overlooked. Preference testing, usability and risk are all inter-linked, with
Table 3 FDA Recommended Qualities for Patient Preference Testing

<table>
<thead>
<tr>
<th>Recommended Qualities</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Patient centeredness</strong></td>
<td>The patient, and not the healthcare professional, should be the focus of the study. Risk-benefit preferences should be obtained from well-informed patients.</td>
</tr>
<tr>
<td><strong>2. Representativeness of the Sample and Generalizability of Results</strong></td>
<td>Preferences should be obtained from a representative sample of adequate size to enable generalisation of the results to the intended population. If preference differences between subgroups are of interest, the sample size of each subgroup should also be of adequate size.</td>
</tr>
<tr>
<td><strong>3. Capturing Heterogeneity of Patients’ Preferences</strong></td>
<td>Individual patient preferences will vary, no matter the commonalities that may exist between individuals. This variability in individual patient preference needs to be accounted for. Note: PPI can help identify subgroup preferences of risk-benefit trade-offs (eg, patients with higher pain or increased severity of a condition/disease may be willing to accept more risk than those with a lower pain/less severity).</td>
</tr>
<tr>
<td><strong>4. Established Good Research Practices by Recognized Professional Organizations</strong></td>
<td>Guidelines for good research practices from recognized professional organizations should be used, such as the guideline from the International Society for Pharmacoeconomics and Outcomes Research (ISPOR).</td>
</tr>
<tr>
<td><strong>5. Effective Communication of Benefit, Harm, Risk, and Uncertainty</strong></td>
<td>The context of the benefit-risk trade-offs, the level of effectiveness and the severity of treatment-related harms need to be defined and patients need to understand probabilities using appropriate numeric, verbal, and graphic representations of uncertainty.</td>
</tr>
<tr>
<td><strong>6. Minimal Cognitive Bias</strong></td>
<td>Potential cognitive bias, such as framing, anchoring, simplifying heuristics, and ordering effect, should be minimized.</td>
</tr>
<tr>
<td><strong>7. Logical Soundness</strong></td>
<td>Internal-validity tests of logic and consistency should be included.</td>
</tr>
<tr>
<td><strong>8. Relevance</strong></td>
<td>Preferences in relation to harm, risk, benefit, and uncertainty should be obtained or well justified if not. Specific endpoints should be clearly defined in relation to their clinical outcomes.</td>
</tr>
<tr>
<td><strong>9. Robustness of Analysis of Results</strong></td>
<td>Appropriate analysis of the results is needed for correct interpretation of the data. Sources of uncertainty need to be understood and can be reported via confidence intervals and standard errors.</td>
</tr>
<tr>
<td><strong>10. Study Conduct</strong></td>
<td>The study should be conducted by trained research staff. If the study is self-administered, the patients should be given a tutorial and quiz on how to complete the study.</td>
</tr>
<tr>
<td><strong>11. Comprehension by Study Participants</strong></td>
<td>Patients should fully comprehend the harm, risk, benefit, uncertainty, and any other medical information presented to them (ie, all questions must be fully understood for data to be valid).</td>
</tr>
</tbody>
</table>

and feasibility, (3) design and development – verification and validation, (4) final validation – product launch preparation, and (5) product launch and postlaunch assessment (see Figure 2).

Usability is directly correlated with the mitigation of risks associated with correct use and use errors, meaning that it is an essential process in the design and development of a medical device. As preference testing is conducted prior to usability, preferences can affect usability through design changes (ie by adding/removing/adjusting features) and therefore will have an impact on risks related to device use. Any change related to device design should be followed by usability testing and implemented only if it does not pose unacceptable risks for the user. Figure 3 presents the process to follow after obtaining the results from a preference test in relation to usability. In all scenarios, usability testing includes a risk management process, by which it can be extracted whether the user’s preference affects the usability of the device or not. If a preference has no impact on usability, the proposed preference can be implemented in the development plan of the device without any further change (Scenario 1 Figure 3). If the examined preference has an unacceptable impact on usability and risk, then changes to

![Figure 1](https://doi.org/10.2147/MDER.S368420)
the preference should be made (to reduce risk), or the preference should not be implemented and the next preference in the list should be examined (Scenario 2 in Figure 3). If risk management shows that preference testing results in minor but acceptable risks for the user, the preference is accepted and is implemented in the final design of the device (Scenario 3 in Figure 3). In all cases, a risk/hazard analysis must be conducted by the manufacturer to identify possible foreseen and unforeseen risks associated with the design of the device (Steps 1–4 in Figure 3).

**Potential Differences Between Patient and HCP Preferences**

Although the FDA believes that usability testing is more suitable for obtaining HCP preferences than preference testing, preference information from HCPs can reveal differences in opinion on what the most important clinical outcomes of treatment may be. For example, being pain-free may seem like an obvious endpoint from both the patient and HCP, however, if patients do not consider being pain-free as a realistic endpoint, then it will not be of high preference to them. These potential differences in opinion between patients and HCPs are demonstrated in a recent systematic review that compared patient and HCP preferences on attributes of healthcare interventions (such as outcomes, infrastructure, safety, and so on). Findings revealed that HCPs placed more focus on attributes of structure (ie material resources, organisational structures and human resources) and care outcomes, especially mortality, compared to patients. Likewise, a publication regarding HCP preferences for treating patients with non-alcoholic steatohepatitis found that, although both patients and HCPs agreed on treatments that improved liver health, there were differences related to their secondary priorities. Specifically, patients focused on the impact of symptoms related to the disease, while HCPs found the impact of the progression of the disease on long-term liver damage to be as important as improving liver health. This long-term treatment was much lower in patients’ priorities.

---

**Figure 2** Stage-gate design concept for medical devices, including the conduction of preference testing. Data from sources.22,25

---

Source: Lewis et al.

https://doi.org/10.2147/MDER.S368420

DovePress
Further differences between patients and HCP preferences are highlighted in a recent article that established the key factors that define an orthopedic surgeons’ preferences on knee and hip implants.\(^{26}\) Factors related to technology were rated as the most important for choosing an implant, which included improved outcomes for patients, longevity of implant, and design and ease of implant and instrumentation, among others. Additionally, the absence of a similar implant from another vendor, and the knowledge, availability, thoroughness, and follow-up of the sales representative were also rated highly. Finance and cost were considered of lower importance.\(^{26}\) These findings indicate that there are preferences that can be obtained from HCPs that would be unlikely to be assessed in patients, such as the availability of similar devices, highlighting that both patient and HCP preferences provide valuable sources of information that may be missed if only one source were accounted for.

Obtaining both patient and HCP preferences is thus important to obtain a full overview of all relevant preferences that can aid with usability solutions that positively impact the risk/benefit assessment of a device. Both sources of data are imperative to create a complete device to understand and optimize design decisions. For example, during the pre-market phase, preference testing may identify sub-groups of the intended population that are willing to accept a higher risk-benefit trade-off (ie patients with higher pain or increased severity of a condition/disease may be willing to accept more risk than those with a lower pain/less severity). At the post-market stage, clinical follow-ups under the European Medical Device Regulation 2017/745 (MDR), although designed to establish performance and safety of the device, may highlight endpoints that can be used for preference testing, such as issues with usability from both patient and HCP perspectives.
Patient Preferences and Quality of Life (QoL)
PPI can result in greater transparency and accountability during product development and improve and assess the most important outcomes for patients. For example, in pharmaceuticals patient preference studies are used as part of a drug’s clinical development program to aid with selecting the most appropriate clinical endpoints (ie the endpoints that are of most value to patients) and to ensure patient-relevant outcomes are incorporated, as well as to identify the patients that may benefit most from innovative technology.

As a result, PPI can have a direct impact on patient quality of life (QoL), defined by the World Health Organization (WHO) as

an individual’s perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns.

PPI can improve QoL outcomes for patients by taking into account their goals, expectations, standards and concerns in relation to treatment outcomes. For example, a clinical trial conducted as part of the PREFER project found that patients rated their QoL as greater when treated for chronic non-malignant pain using oxycodone PR/naloxone PR, compared to previous treatment using the WHO-step I and/or WHO-step II analgesics, indicating the importance of patient preferences on QoL in treatment options. Likewise, QoL has been shown to have an impact on patient preferences for the treatment of multiple sclerosis. Patients’ treatment preferences were influenced by efficacy, side effects, and mode of administration, with some side-effects considered as too detrimental to their present QoL to be acceptable. For example, one patient reported that

It’s here and now that I’m alive, not in ten years, right? And that’s why I have to say that it’s now that my children need me and it is now that my husband must have me,

thus highlighting that, for some patients, current QoL is more important than long-term health outcomes, and that preferences regarding the balance between treatment efficacy and current QoL differ between patients.

Taking account of PPI is thus crucial if the most appropriate treatment option is to be selected for the patient. This point has recently been highlighted by a survey examining current shared decision-making between patients and their clinicians regarding treatment options for cancer. The survey revealed that although 62% of patients thought that personal QoL preferences were important for treatment decision-making, only 37% reported that they discussed QoL preferences with their clinician before starting treatment. Thus, highlighting a demand that is not yet being met.

The importance of QoL preferences is also highlighted by the impact that QoL outcomes have on reimbursement and treatment use. For example, in England and Wales, The National Institute for Health and Care Excellence (NICE) uses an incremental cost-effectiveness ratio to determine whether a treatment is used/continued to be used, which is based upon the cost per quality-adjusted life-year (QALY). Similarly, in Canada, the Canadian Agency for Drugs and Technologies in Health (CADTH) recommends reimbursement for cancer drugs based upon a cost-effectiveness ratio using QALY. The QALY measures are based upon preference measures of QoL, thus clearly demonstrating the importance of QoL preferences on reimbursement, and ultimately influences whether a treatment is recommended/approved.

The Value of Preference Testing in the Lifecycle of Medical Devices
Although PPI is yet to be widely implemented within medical treatments, its value within decision-making throughout the lifecycle of a medical treatment is largely acknowledged. For instance, a survey conducted in 2006 by the International Network of Agencies for Health Technology Assessment (INAHTA) found that involving consumers in decision-making processes “broadens the perspective of those assessments and of the advice provided to decision makers”.

Furthermore, the use of PPI for the selection of clinical endpoints should increase participants’ willingness to be part of a clinical trial and consequently aid with clinical development, as well as improving the likelihood of the target population’s use of the medical treatment/device once marketed.

Regarding the stage at which the incorporation of HCP/patient preferences could be most beneficial, a literature review assessing the benefits and barriers to implementation of user involvement in medical devices found that
involvement of users at all stages of product development was beneficial; however, it was most beneficial if implemented during the early stages (eg design concept), design and development, prototype testing, and clinical trials. Moreover, a recent study conducted an in-depth investigation into the critical decision-making stages from key stakeholders (industry, regulatory authorities and reimbursement/HTA) to determine at which points PPI could be beneficial within the lifecycle of medicines. For industry stakeholders, PPI was considered valuable for decision-making processes throughout the entire lifecycle, and especially in the design and development stages. For the regulatory authorities, PPI was acknowledged as being important, but was not readily accepted within their decision-making stages due to a lack of recognized and structured protocols for including PPI. For example, one EU respondent said, “All these things are very important but you have to create a way of measuring the impact of taking into account patient preference”. For HTA, although mechanisms for PPI collection to help identify the most important reimbursement/HTA topics were in place and PPI was sometimes included in assessment dossiers, it was not required.

It should be noted that there may be slight differences in opinion for medical devices at the regulatory decision-making level, due to differences in regulatory process for medicines and medical devices. Although the general principles surrounding the regulation of medicines and medical devices are the same (ie benefit-risk assessment and post-market monitoring), there are differences with regard to the specifics of regulation. For example, within medicines, pre-market assessment is conducted by the competent authorities themselves, whereas for medical devices, competent authorities work through notified bodies, with the notified bodies being responsible for ensuring compliance with the relevant regulations. Thus, incorporating PPI into the regulatory decision-making process in medical devices will require the involvement of both notified bodies and competent authorities.

Potential Preference Testing Methods for Medical Devices and Appropriate Method Selection

Preference testing methods can be broadly categorized into either qualitative or quantitative methods, which can involve individual subjects, groups, or a combination of both (ie concept mapping method, see Table 1). Although traditionally used within social sciences, qualitative preference testing methods can also be applied in complex medical settings (ie health services research) to assist in identifying HCPs and/or patients’ preferences, which may facilitate a stable and structured basis for medical decision-making processes. Qualitative methods provide descriptive data that is collected via participant and/or phenomenon observation and includes information that reflects the participants’ subjective experiences and decisions. Quantitative methods (also described as elicitation methods) differ from qualitative methods in terms of providing quantifiable and measurable data. These data can be quantified, statistically analyzed and used to inform decision-making processes. As such, both qualitative and quantitative methods can be used to collect different types of preference information from HCPs, patients and/or their legal representatives, both of which can inform regulatory and medical-related approval decisions.

An overview of qualitative and quantitative preference testing methods that are broadly used within health care settings is presented in Table 1. Regardless of which preference testing method is chosen, an appropriate group of auditors (eg expert HCPs) should be selected, as preferences will vary based on the HCP/patient group targeted. Additional considerations should include sample size adequacy (representative sample to provide statistical confidence), study design attributes (easy to implement and comprehend) and meaningful statistics (replicable results and data analysis that will reflect “non-preference” options). It is also important to consider that preferences obtained using a preference test may not necessarily be an accurate reflection of preferences obtained in everyday, real-world settings. For example, when using a 9-point hedonic scale as a measure of preferences, a participant may give two products the same rating, yet still have a preference for one product over the other but have no way of indicating this (ie preference information is lost). As well as losing information on preferences, false preferences can also be obtained. For instance, research has indicated that paired preference testing can result in false preferences due to a response bias (eg a preconception that a preference should be given). For example, a study conducted to assess the preferences for two cigarettes (type A and type B) found that 80% of participants reported a preference for one of the types, despite the fact that the two types were actually identical. Potential methods to
combat these issues are to create preference tests that do not resemble a preference task and to include identical “placebo” items to act as a control data that show the results of the testing conditions per se, and not actual preferences.

**Real-World Examples of Preference Testing of a Medical Device**

Although preference testing is not yet widely implemented in the medical field, there are examples of its use in practice. A study on patients’ preferences was used as part of the approval decision for a novel obesity device, namely the VBLOC Maestro® Rechargeable System. The patient preference test obtained information on the maximum amount of mortality risk that a patient would accept for a percentage of total body weight loss. For example, a device with a 0.01% risk of mortality would be accepted by a patient if 10% total body weight loss for five years was the expected outcome. This information on the risk-benefit acceptance was not known prior to this preference study, and as such, the FDA did not have the clinical data to be able to quantify the trade-offs that patients were willing to accept. The preference study provided these data, which was subsequently used as part of the approval for this novel device.

Likewise, a patient preference study by Medtronic on hypertension revealed that patients would accept a 20% risk of adverse events if their office-based systolic blood pressure was lowered. The study also found that a minimal acceptance benefit of less than a 2.5 mmHg reduction in office-based systolic blood pressure was needed for patients to accept the treatment risks. The findings differed from those observed in published literature, indicating that patients may accept a lower reduction in blood pressure (reduced benefit) and may tolerate higher risks than previously observed. These findings have important implications for devices intended to lower blood pressure, as the lowering of blood pressure by small amounts is considered beneficial to patients, despite potential treatment risks.

Preference testing was also implemented in a study to assess patient preferences between two different types of portable infusion-pump devices. The preference test compared an elastomeric infusor (Baxter) with an electronically controlled mechanical pump. Despite the numerous technical benefits of the electronically controlled pumps, patients preferred the elastomeric pump (testing p-value <0.01), due to its comfort.

Finally, preference testing was used as part of a study to evaluate recommendations for the use of fluorodeoxyglucose positron emission tomography (FDG-PET) as an aid for the diagnosis of dementing neurodegenerative disorders. Literature searches were conducted to assess the sensitivity, specificity and accuracy (amongst others) of FDG-PET in the detection of dementing neurodegenerative disorders. An expert panel (the European Association of Nuclear Medicine Neuroimaging Committee and the Neuroimaging and Dementia Study Groups of the European Academy of Neurology) then voted for or against the use of FDG-PET based on the obtained evidence and their expertise. They concluded that FDG-PET is useful for all dementias and mild cognitive impairment (with the exception of amyotrophic lateral sclerosis and Huntington’s Disease) but that it was not suitable for pre-clinical conditions. Thus, using a preference testing method (Delphi method), a consensus was achieved regarding recommendations for the clinical use of FDG-PET. These examples thus highlight the value that PPI can add to the lifecycle of a medical device.

**Challenges and Lessons Learned**

Despite the benefits of preference testing, there remain numerous challenges that restrict its further development and use within the medical field. A major concern is the difficulty regarding the choice of a suitable method for conducting a preference study, due to the wide range of different methods and types of information that are required at different stages of the product lifecycle. This may result in inappropriate methods being used and thus limit the value of the data obtained. The second major concern is the current level of patient and HCP knowledge/awareness of the value of preference data in the decision-making process. However, within the US, guidance documents provided by the FDA are highlighting the importance of PPI and paving the way for its incorporation within medical device decision-making. Likewise, within the EU, the PREFER project is increasing awareness of the value of PPI and will be key to establishing best practices for using PPI within the medical device lifecycle.

In conclusion, preference information is a valuable source of data that can provide input into the usability, and ultimately, the benefit-risk profile of a medical device. Currently, preference testing is not covered by EU regulation and is covered by an extremely limited framework within the US. Moreover, guidelines and standards for implementing preference testing with medical devices are lacking/insufficient. The result of which may lead manufacturers to employ
inappropriate/unreliable preference testing methods, or simply avoid conducting preference testing altogether. This paper serves to raise awareness of the value of patient and HCP preference data and its potential use throughout a device’s lifecycle. The creation of sufficient guidelines to enable stakeholders (eg industry, regulatory, HCP, etc.) to use patient and/or HCP preference information effectively, and/or incorporation of preference testing into medical device regulations, would greatly aid with the widespread implementation of preference information in medical devices.

Disclosure

Dr Amie Smirthwaite reports on working for a medical devices consultancy. The authors report no other conflicts of interest in this work.

References

11. FDA. Patient-engagement in the design and conduct of medical device clinical investigations; 2019.
17. PREFER, CHMP & EUnetHTA parallel scientific advice: qualification of a framework and “points to consider” for method selection along with five methods for performing patient preference studies to inform regulatory and HTA body medical product decision-making; 2021.
18. PREFER, Missed the PREFER framework webinar? Recording now available; 2021.


38. Pope C, Mays N. Qualitative research: reaching the parts other methods cannot reach: an introduction to qualitative methods in health and health services research. *BMJ*. 1995;311(6996):42. doi:10.1136/bmj.311.6966.42


