Preference-Based Assessments

Suitability of Preference Methods Across the Medical Product Lifecycle: A Multicriteria Decision Analysis

Jorien Veldwijk, PhD, Esther de Bekker-Grob, PhD, Juhaeri Juhaeri, PhD, Eline van Overbeeke, PhD, Stephanie Tcherny-Lessenor, MD, MSc, MPH, Cathy Anne Pinto, PhD, Rachael L. DiSantostefano, PhD, Catharina G.M. Groothuis-Oudshoorn, PhD

ABSTRACT

Objectives: This study aimed to understand the importance of criteria describing methods (eg, duration, costs, validity, and outcomes) according to decision makers for each decision point in the medical product lifecycle (MPLC) and to determine the suitability of a discrete choice experiment, swing weighting, probabilistic threshold technique, and best-worst scale cases 1 and 2 at each decision point in the MPLC.

Methods: Applying multicriteria decision analysis, an online survey was sent to MPLC decision makers (ie, industry, regulatory, and health technology assessment representatives). They ranked and weighted 19 methods criteria from an existing performance matrix about their respective decisions across the MPLC. All criteria were given a relative weight based on the ranking and rating in the survey after which an overall suitability score was calculated for each preference elicitation method per decision point. Sensitivity analyses were conducted to reflect uncertainty in the performance matrix.

Results: Fifty-nine industry, 29 regulatory, and 5 health technology assessment representatives completed the surveys. Overall, “estimating trade-offs between treatment characteristics” and “estimating weights for treatment characteristics” were highly important criteria throughout all MPLC decision points, whereas other criteria were most important only for specific MPLC stages. Swing weighting and probabilistic threshold technique received significantly higher suitability scores across decision points than other methods. Sensitivity analyses showed substantial impact of uncertainty in the performance matrix.

Conclusion: Although discrete choice experiment is the most applied preference elicitation method, other methods should also be considered to address the needs of decision makers. Development of evidence-based guidance documents for designing, conducting, and analyzing such methods could enhance their use.

Keywords: decision makers, medical product lifecycle, multicriteria decision analysis, preference elicitation, preference methods, stakeholders.

VALUE HEALTH. 2023; 26(4):579–588

Introduction

Increasingly decision makers look for ways to measure patients’ preferences and include such information in decision making along the medical product lifecycle (MPLC). Including preference information might be apparent for some decisions such as for identifying unmet medical needs and selecting endpoints for randomized controlled trials or for the purpose of quantitative benefit-risk assessment. Nevertheless, the exact role of patient preferences in other industry decision points, especially regulatory and health technology assessment (HTA)/reimbursement-related decisions, is less clear. Both regulatory and HTA agencies advocate for increased transparency in their decision-making process and a focus on patient-centered decision making. The Food and Drug Administration has issued guidance for the conduct of patient preference studies (PPS), and the European Medicines Agency recently provided a positive qualification opinion and asked for public consultation on a preference elicitation framework. For HTA and reimbursement, the inclusion of preferences in decision making seems more distant. Current cost-utility analysis frameworks do not allow for easy inclusion of patient preference information and require more structural changes. Nevertheless, initiatives are undertaken; for instance, the National Institute for Health and Care Excellence published their perspective on the use of PPS in HTA and provided scientific advice on the conduct of PPS. Nevertheless, the weighting or incorporation of preferences against the standard information (eg, clinical data, cost-effectiveness data) in decision making along the MPLC remains
debated. According to previous research, the MPLC, in total, consists of approximately 15 decision points for different decision makers: pharmaceutical industry, regulators, and HTA agency/body.  

Decision makers themselves indicated that most of these decisions could include patient preference information to some extent. At the same time, decision makers likely require different types of information with varying depth and focus along the MPLC, making it complicated to select one or few suitable preference elicitation methods that fit the needs of all decision makers across all decision points of the MPLC.

A recent literature review identified 22 preference elicitation methods grouped into ranking, rating, indifference methods, and discrete choice methods. Within each category, most commonly used methods in healthcare to elicit preferences of patients were discrete choice experiment (DCE), probabilistic threshold technique (PTT), swing weighting (SW), best-worst scaling case 1 (BWS1), and best-worst scaling case 2 (BWS2). In a first effort to identify methods most suitable for satisfying stakeholders’ needs, Whichello and colleagues combined a Q-method and analytical hierarchy process to appraise all 22 preference elicitation methods. The relative weight of criteria describing methods (eg, duration, costs, validity, and outcomes) was evaluated for 4 hypothetical MPLC scenarios: 2 variations of early clinical development stages, 1 late phase III scenario, and 1 postmarketing scenario. 

Weighting of criteria for methods appraisal was mostly based on response of representatives from industry and academia (1 HTA and 1 regulator responded). To provide insight into suitability of methods across the full MPLC and thereby facilitate methods selection and systematic implementation of preference elicitation along the MPLC, it remains crucial to understand the importance of methods criteria for each decision maker (ie, industry, regulator, HTA agency/body) at their respective critical decision points in the MPLC. Therefore, this study aimed to evaluate the importance of methods criteria to fully appraise the performance of 5 commonly used preference elicitation methods against these methods criteria according to decision makers at different moments along the MPLC.

Methods

We used multicriteria decision analysis (MCDA) in this study, which is a methodology for appraising alternatives on individual, often conflicting criteria, and combining them into 1 overall appraisal. Common steps in MCDA are (1) defining the decision problem (including decision makers), (2) selecting criteria, (3) measuring performance, (4) weighting of criteria, (5) aggregating results, (6) sensitivity analyses, and (7) interpretation of results. 

Step 2 to 6 will be detailed below given that step 1 was outlined in the introduction (ie, to provide insight into suitability of methods across the full MPLC and thereby facilitate methods selection and systematic implementation of preference elicitation along the MPLC) and step 7 will be covered in the Results and Discussion sections of this article.

Selecting Criteria and Measuring Performance

Whichello et al initially identified 35 method criteria as being most important for selecting a qualitative or quantitative patient preference method based on literature reviews and previous studies. These were subsequently restricted to 19 criteria (12 operational and 7 outcomes related criteria) by means of a Q-method experiment among stakeholders (N = 54 being academic, representative from industry or regulatory/HTA agency, physician, patient (representative), or consultant). Whichello et al subsequently developed a performance matrix (Table 2) specifying the performance of each method for each criterion was created based on semistructured interviews with preference method experts (N = 17) and a literature review. Further details on the method and development of the performance matrix can be found elsewhere.

Weighting of Criteria

Three surveys were developed to assess the relative importance, that is, the weights, of the methods criteria for each of the critical decision points in the MPLC in which patient preference information could be considered in addition to the current evidence used for decision making. Furthermore, the surveys were tailored to decision processes of the 3 decision-maker groups in the MPLC in such a way that surveys for industry representative included 6 industry-related decision points (ie, select and prioritize targets and leads, prioritize studies, prioritize assets, optimize and prioritize assets, regulatory submission and launch, manage MPLC, and prioritize opportunities), the survey for regulators contained 1 regulatory decision point (scientific opinion), and the survey for HTA agency/body representative included 1 HTA decision point (appraisal).

Respondents (recruitment strategies are described below) were invited to participate by sharing the link for the survey and an explanatory letter. The survey started with an explanation of what a PPS constitutes, and after that, 4 background questions were included to get insight in the respondents and their experience with such preference studies. In the next part of the survey, the respondents were asked to rank the method criteria included in the performance matrix (Table 2) from most to least important for each of the decision points that related to their specific decisional framework (for instance, HTA representatives were only asked about the importance of the methods criteria related to appraisal). To avoid ordering bias, the order in which the criteria were presented to respondents was randomized. For the criteria that a respondent ranked in their top 10, the respondents were asked to rate (on a scale of 100) the criterion compared with their top-ranked criterion, the score of which was set to 100. They were specifically not asked to weight all 19 criteria owing to the high cognitive burden of such a task resulting in fatigue and further potential bias induced by such a request.

The surveys were constructed by the research team and reviewed by different decision makers (ie, 5 industry representatives, 2 Food and Drug Administration representatives, and 1 Belgium HTA representative [also representing the EUnetHTA]). Thereafter, the surveys were pretested by means of 3 think-aloud interviews (using conveyance sampling) to refine language, relevance, and usability of the survey. After the pretest, changes were made to the surveys related to the explanation of the decision points, what constitutes a preference study and the content/meaning of the criteria. The surveys were developed in Lighthouse Studio 9.7.0. Industry representatives within the Patient Preferences in Benefit-Risk Assessments During the Drug Life Cycle (PREFER) consortium and the Benefit-Risk Assessment, Communication, and Evaluation special interest group were asked to invite industry representatives to complete the survey. When disseminating the survey, it was requested to forward the invitation to colleagues at different departments (eg, regulatory-policy, drug safety, epidemiology, clinical development, health outcomes research, value and access groups). Regulatory representatives were contacted via the European Medicines Agency. HTA agency/body representatives were contacted via the head of the PREFER HTA advisory board. In total, 20 to 40 respondents per group of decision makers were anticipated to result in sufficient data to arrive at meaningful conclusions.
Based on the ranking position and the points from the rating exercise, all the criteria were given a relative average weight $w_i$ for each decision point.\textsuperscript{23,24} For each respondent, the criteria with an individual ranking outside the top 10 were given a weight of 0. The weights of the other criteria were calculated by scaling the ratings with the total sum of the points such that the sum of all weights equaled to 100. Next, the individual weights were averaged over all respondents giving the average weight $w_i$ for each criterion. Subsequently, for each preference elicitation method, an overall value was calculated per critical decision point along the MPLC.

**Aggregation**

Based on whether the methods met certain criteria (the scoring, see performance matrix in Table 2\textsuperscript{22,27-31}) and the points allocated to that criterion (the weighting) for that particular decision point.\textsuperscript{13,24} The overall value of the separate methods for each critical decision point was calculated as:

$$\text{total value} = \frac{1}{k} \sum_{i=1}^{k} w_i x_i$$

where $x_i$ indicates the scoring of a method on criterium $i$ (0 or 1), $w_i$ the average weight of criterium $i$ for a critical decision point.
Trade-offs between treatment characteristics are not common practice in SW; it can theoretically be done but only with (too) many assumptions.

Low costs for DCE as the qualitative work across methods is equally much and specialized software and expertise for DCE is no longer a necessity (free packages such as DCE-SW).8 Attributes, such as DCE is not advised with 8 attributes owing to complexity of choice tasks.9

Bootstrap confidence intervals for the overall values per method.33,34 The overall value can in principle range between 0 and 100. Bootstrap sampling was used to estimate nonparametric confidence intervals for the overall values per method.35

**Sensitivity Analyses**

Separate sensitivity analyses were conducted to account for the uncertainty in the performance matrix because of (1) a lack of consensus among experts or (2) conflicting evidence from literature and experts (in these cases, final decisions were based on literature).22 Conducted sensitivity analyses are listed below including a rational for each of the analyses. Analyses are grouped based on their origin (ie, either based on uncertainty in performance matrix or based on additional insights).

1. Analysis based on uncertainties in the original methods performance matrix
   A. Assigning a value of 0 to all criteria for which a value of 1 was uncertain
   B. Assigning a value of 0 to all criteria for which a value of 1 was uncertain and assigning a value of 1 to all criteria for which a value of 0 was assigned with uncertainty

2. Analysis based on insights from PREFER case studies and expert consultation within the consortium

C. Assigning low cognitive load (of method on patient) for all methods (given that recent research reported DCE not to be perceived difficult by respondents27-31)

D. Reassigning methods criteria according to the revised performance matrix in Table 3

E. Reassigning methods criteria according to the revised performance matrix in Table 3 and indicating a 1 for DCE for “estimates external validity” (given that research has shown and is currently conducted on external validity in DCE studies36-40)

F. Reassigning methods criteria according to the revised performance matrix in Table 3 and indicating a 1 for BWS2 for “estimating trade-offs between treatment characteristics” (given that latent class analyses and mixed logit models can be used for the analyses, the estimates resulting from such analysis for BWS2 could be used to calculate secondary outcomes measures such as trade-offs)

<table>
<thead>
<tr>
<th>Criteria</th>
<th>DCE</th>
<th>BWS1</th>
<th>BWS2</th>
<th>SW</th>
<th>PTT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low cost of the patient preference study</td>
<td>1^†</td>
<td>1^†</td>
<td>1</td>
<td>1^†</td>
<td>1</td>
</tr>
<tr>
<td>Quick sessions with participants (≤ 30 min)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Low frequency of sessions (&lt; 2)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Study duration (≤ 6 months)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8 or more treatment characteristics can be explored</td>
<td>0</td>
<td>1^†</td>
<td>0^§</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Small sample size (≤ 100)</td>
<td>0</td>
<td>1^†</td>
<td>0^§</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>A low cognitive load on patients</td>
<td>1^†</td>
<td>1^†</td>
<td>1^†</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Low complexity of instructions</td>
<td>0</td>
<td>0^†</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Public acknowledgment by your organization as an acceptable method</td>
<td>1</td>
<td>1^†</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Easy to add new treatment characteristics</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>The patient preference study does not include interaction among participants</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Group dynamic with participants</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>The patient preference study results allow for the calculation of risk attitudes</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Exploring reasons behind a preference in qualitative detail</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Estimating weights for treatment characteristics</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Estimating trade-offs between treatment characteristics</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Quantifying heterogeneity in preferences</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Internal validation methods can be incorporated</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Establishes external validity</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note. 1 indicates the method complies to the criteria and 0 means it does not. Further indications for changes compared with the original performance matrix and explanations of reasoning behind all changes have been indicated using symbols * and † to **. *Sample size of DCE and BWS2 cannot be > 100 to perform the common practice statistical models (conditional logit, MIXL, or LCA). †Low complexity of instructions for all methods in accordance with results of recent publications.22 §Group dynamic with participants for DCE unrealistic as also in lab conducted experiment you get individual outcomes. ‡Calculation of risk attitudes not possible in BWS and SW given that attributes are not actively traded against each other such as in DCE and PTT where people can focus on avoiding (all) risks. ††Trade-offs between treatment characteristics are not common practice in SW; it can theoretically be done but only with (too) many assumptions.

k total number of criteria, and i index of summation.33,34

Results

In total, 59 industry representatives, 29 regulatory representatives, and 5 HTA agency/body representatives completed the survey. Most participants were from the United States and
Table 3. Overview of respondents’ characteristics stratified by the decision maker.

<table>
<thead>
<tr>
<th>Respondents’ characteristics</th>
<th>Industry (n = 59)</th>
<th>Regulators (n = 29)</th>
<th>HTA/payer (n = 5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country of residence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Belgium</td>
<td>0</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>France</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Germany</td>
<td>14</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Italy</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>4</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Switzerland</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sweden</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>United States</td>
<td>24</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Other*</td>
<td>5</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>Department of employment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regulatory affairs</td>
<td>7</td>
<td>12.5%</td>
<td></td>
</tr>
<tr>
<td>Real-world evidence group</td>
<td>11</td>
<td>19.6%</td>
<td></td>
</tr>
<tr>
<td>Patient (or) drug safety</td>
<td>17</td>
<td>30.4%</td>
<td></td>
</tr>
<tr>
<td>Epidemiology or pharmacoepidemiology</td>
<td>24</td>
<td>42.9%</td>
<td></td>
</tr>
<tr>
<td>Clinical development</td>
<td>8</td>
<td>14.3%</td>
<td></td>
</tr>
<tr>
<td>Medical affairs</td>
<td>5</td>
<td>8.9%</td>
<td></td>
</tr>
<tr>
<td>Health economics and outcomes research</td>
<td>9</td>
<td>16.1%</td>
<td></td>
</tr>
<tr>
<td>Market access</td>
<td>4</td>
<td>7.1%</td>
<td></td>
</tr>
<tr>
<td>Consultancy agency</td>
<td>5</td>
<td>8.9%</td>
<td></td>
</tr>
<tr>
<td>Other: statistics</td>
<td>2</td>
<td>3.6%</td>
<td></td>
</tr>
<tr>
<td>Experience with PPS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Read PPS</td>
<td>46</td>
<td>78.0%</td>
<td>5</td>
</tr>
<tr>
<td>Organized/managed PPS</td>
<td>25</td>
<td>42.4%</td>
<td>3</td>
</tr>
<tr>
<td>Evaluated PPS</td>
<td>26</td>
<td>44.1%</td>
<td>3</td>
</tr>
<tr>
<td>Used PPS in work</td>
<td>30</td>
<td>50.8%</td>
<td>2</td>
</tr>
<tr>
<td>Attended webinars/conferences on PPS</td>
<td>44</td>
<td>74.6%</td>
<td>3</td>
</tr>
<tr>
<td>Other experience</td>
<td>10</td>
<td>16.9%</td>
<td>1</td>
</tr>
<tr>
<td>Do not know what a PPS is†</td>
<td>2</td>
<td>3.4%</td>
<td>0</td>
</tr>
</tbody>
</table>

HTA indicates health technology assessment; PPS, patient preference study.
*Most reported countries under “other” were Austria, Latvia, Slovakia, Ireland, Greece, Portugal, Poland, Finland, Denmark, and Canada.
†Respondents who indicated not to know what a PPS is were excluded from the survey.

Germany (see Table 3 for a full overview). Industry representatives had an average of 9.9 (SD = 7.3) years of experience in their current position, with a range from 1 to 30 years. Among regulators and HTA representatives, respectively, the average years of experience in their current position were 8.7 years (SD = 5.5; range 2–23 years) and 11 years (SD = 6.1; range 3–16 years). From industry, most self-identified as working in “epidemiology or pharmacoepidemiology” or “patient (or) drug safety” (see Table 3 for full overview). Respondents differed in their familiarity with PPS; although the majority had read PPS, only approximately half of the respondents had used PPS in their work (see Table 3 for full overview).

Weighting of Methods Criteria

Methods criteria that were appointed the largest values were reported per decision point of each of the decision makers. When a criterion had a total value of 8 or more (meaning the criterion is 50% more important than the value that would have been calculated if all criteria were equally important), the criterion was marked among the highest weighted criteria. Values and standard deviations of all criteria are listed in Table 4, with the top-weighed criteria being specifically indicated (*). Please see Table 1 for a full definition of all criteria.

Overall, “estimating trade-offs between treatment characteristics” and “estimating weights for treatment characteristics” were important criteria throughout all decision points of the MPLC. “Exploring reasons behind preferences in qualitative detail” seemed most important in the early industry decisions and in HTA/appraisal. “External validity,” “internal validation methods can be incorporated,” and “quantifying heterogeneity in preferences” showed to be more important from clinical development phase 3 and onward to the later stages in the MPLC.
Both for BWS1 and BWS2, the total values across decision points were relatively stable implying them to be equally suitable for all decision points. There was more variability across total values of the other methods included (Table 5). Based on the valuation of the methods criteria, DCEs seemed to be most suitable during clinical development and regulatory launch. SW and PTT seemed to be most suitable throughout all industry decision points but total values were lower for regulatory and HTA decision making. When comparing the suitability of the methods across the decision points, SW and PTT were valued significantly better for all decision points than the other methods.

Dividing the values of methods based on operational versus outcome criteria, all methods tended to score lowest for HTA decision making when looking at operational criteria only. The DCE method in total scored lowest for operational criteria across decision points and in comparison with other methods. When only
considering outcome criteria, BWS1 and BWS2 scored relatively lower than the DCE, SW, and PTT methods across all decision points.

### Sensitivity Analyses

Sensitivity analyses show that the overall value of methods changed substantially from the base case (Appendix in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.11.019) (Fig. 1A) depending on the scoring in the performance matrix (Appendix in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.11.019) (Fig. 1B-G). Although the total value of BWS1 and BWS2 remained quite consistent, the total value of the DCE substantially increased in some instances whereas the total value of mainly SW was reduced in several analyses.

### Discussion

This study evaluated the importance of methods criteria according to decision makers at different moments along the MPLC to appraise the performance of 5 commonly used preference elicitation methods. Weights were calculated for a total of 19 elicitation methods. The top-ranked criteria for all decision makers across all decision points included “whether a method could estimate trade-offs between treatment characteristics” and “estimate weights for treatment characteristics.” “Exploring reasons behind preferences in qualitative detail” seemed most important in the early industry decisions and in HTA/appraisal. External validity, internal validity, and the quantification of preference heterogeneity showed to be more important from clinical development phase 3 and for regulatory and HTA decision makers. Scoring the methods based on these weights across decision points of the MPLC has shown that SW and PTT had significantly higher scores across all MPLC decision points than DCE, BWS1, and BWS2. DCE scored higher for all industry decision points (except for select and prioritize targets and leads) and regulatory decision making. All methods had better scores for industry-related decision points than regulatory and HTA decisions.

Not all methods criteria were equally important for each decision point according to the decision makers. This was in line with expectations based on a previous interview study regarding what type of information is being used at each decision point and concerns and expectations for PPS. Regulatory decision makers put relatively more weight on external validity of a method, and HTA decision makers put relatively more weight on the ability to explore reasons behind preferences and in qualitative detail than industry decision makers. This likely explains why all preference elicitation methods score relatively lower for HTA and regulatory decision points than industry decision points.

Methods were appraised using a previously established performance matrix, but also based on adapted matrices, which showed substantial differences in the overall scoring of methods. Owing to the ongoing advancements in the field of preference elicitation methods (eg, improvements in their [experimental] design, analysis), performance matrices of preference methods should continue to be updated with empirical evidence. Furthermore, there may be value in using a more detailed performance matrix that allows a less strict value function. Although the performance matrix used in the current study is based on a binary value function allowing methods to comply or not with a certain criterion, an alternative (eg, partial) value function might be more appropriate for several criteria. For instance, according to the current matrix, all methods can be used to identify preference heterogeneity. Although subgroup analysis can be conducted on the data retrieved for all methods, only some methods allow for further investigation of heterogeneity even within subgroups by means of more complex modeling strategies (ie, via mixed logit models or latent class analysis). In addition, assessment of external validity is lacking for most methods except for DCE where recent research is showing favorable results. Although the existing empirical evidence does not fully establish external validity for DCE, it is trending in a favorable direction. The sensitivity analyses that were conducted as part of this study clearly show the impact of small changes in the performance matrix on the overall appraisal of methods.

Although this study was conducted in an international multidisciplinary team and recruited decision makers across the MPLC to determine the weights of methods criteria for all critical decisions points, this study is subject to some limitations. First, a very limited number of HTA representatives (n = 5) responded to the survey making outcomes of the MCDA related to HTA decisions less reliable. Related to this point, owing to the applied...
recruitment strategy, overall response rate cannot be reported. Second, a large set of criteria was included in this study; therefore, respondents were asked to rank only their 10 criteria and in some cases axiomatic properties might have been violated. For example, some methods criteria were showing dependence, which is not accounted for in the MCDA.

This study showed that methods differed in their suitability across specific decision points of the MPLC. In the healthcare setting, DCEs are most applied for eliciting preferences, likely in part owing to the fact that insights into the design and conduct of this method have been published. Nevertheless, other methods including the PTT, SW, BWS1, BWS2, and the remaining 8 methods that were marked promising by Whichello et al should be considered when setting up future preference studies given that also those methods comply with the top-weighted methods criteria according to decision makers. Additional research leading up to evidence-based guidance documents for designing, conducting, and analyzing such methods could enhance their use and implementation. Nevertheless, methods appraisal based on performance matrices should never be the sole determinant for method selection in a case study. Other important considerations such as the research question, requested endpoints, and operational aspects of the study should also be taken into account.

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2022.11.019.

Figure 1. Total value of methods across base case and other scenarios. (A) Base case according to the original performance matrix. (B) Assigning a value of 0 to all criteria for which a value of 1 was uncertain. (C) Assigning a value of 0 to all criteria for which a value of 1 was uncertain and assigning a value of 1 to all criteria for which a value of 0 was assigned with uncertainty. (D) Assigning low cognitive load to all methods. (E) Reassigning methods criteria according to the revised performance matrix in Table 3. (F) Reassigning methods criteria according to the revised performance matrix and indicating a value of 1 for DCE for “establishes external validity.” (G) Reassigning methods criteria according to the revised performance matrix and indicating a value of 1 for BWS2 for “estimating trade-offs between treatment characteristics.”

BWS indicates best-worst scale; DCE, discrete choice experiment; HTA, health technology assessment; PTT, probabilistic threshold technique; SW, swing weighting.
(Tcherny-Lessenot); Merck & Co, Kenilworth, NJ, USA (Pinto); Janssen Research & Development, LLC, Titusville, NJ, USA (DiSantostefano); Health Technology and Services Research, Faculty of Behavioural and Management Science, University of Twente, Enschede, The Netherlands (Groothuis-Oudshoorn).

Correspondence: Jorien Veldwijk, PhD, Erasmus School of Health Policy & Management, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands. Email: veldwijk@eshpm.eur.nl

Author Contributions: Concept and design: Veldwijk, de Bekker-Grob, Juhaeri, Tcherny-Lessenot, Pinto, van Overbeeke, Groothuis-Oudshoorn

Acquisition of data: Veldwijk, Groothuis-Oudshoorn

Analysis and interpretation of data: Veldwijk, de Bekker-Grob, Juhaeri, Tcherny-Lessenot, Pinto, van Overbeeke, Groothuis-Oudshoorn, DiSantostefano

Drafting of the manuscript: Veldwijk, Groothuis-Oudshoorn

Critical revision of the paper for important intellectual content: Veldwijk, de Bekker-Grob, Juhaeri, Tcherny-Lessenot, Pinto, van Overbeeke, Groothuis-Oudshoorn, DiSantostefano

Statistical analysis: Veldwijk, Groothuis-Oudshoorn

Provision of study materials or patients: Veldwijk

Obtaining funding: Veldwijk, de Bekker-Grob, Juhaeri, Tcherny-Lessenot, DiSantostefano

Administrative, technical, or logistic support: Veldwijk

Supervision: Groothuis-Oudshoorn

Conflict of Interest Disclosures: Dr Veldwijk reported receiving grants from IML, during the conduct of the study. Drs Juhaeri and Tcherny-Lessenot are employees of Sanofi. They receive a salary from Sanofi and own Sanofi shares. Dr Juhaeri also has an investment portfolio that from time to time includes shares of other biopharmaceutical companies. Dr Pinto is an employee of MSD, a subsidiary of Merck & Co, Inc, Kenilworth, NJ. She receives a salary and owns shares of corporate stock. Dr DiSantostefano reported receiving other from Johnson & Johnson, during the conduct of the study, and other from Johnson & Johnson, outside the submitted work. Dr van Overbeeke is an employee of NV/SA and owns Pfizer Inc shares and options based on her employment. No other disclosures were reported. This article and its contents reflect the view of the authors and not the view of PREFER, IMI, the European Union, or the European Federation of Pharmaceutical Industries and Associations.

Funding/Support: This study formed part of the PREFER project. The PREFER project has 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 program and the European Federation of Pharmaceutical Industries and Associations.

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Acknowledgment: The authors thank all consortium members of PREFER who contributed to the review and dissemination of the survey or early drafts of the protocol for this study. A special thanks to Jennifer Viberg Johansson for her assistance in the design and pretesting of the initial surveys; Jürgen Kübler for his guidance in the initiation stages of this study; Bennett Levitan and Irina Cleemput for their help with defining appropriate guidance for regulators and HTA representatives in our surveys; Brett Hauber, Annalisa Rubino, and Nathalie Bere for their assistance in the recruitment stages of this study; and Francesco Pignatti for his assistance with recruitment and his valuable comments when reviewing the manuscript.

REFERENCES


