

Individual Trade-Offs Between Possible Benefits and Risks of Cancer Treatments: Results from a Stated Preference Study with Patients with Multiple Myeloma

DOUWE POSTMUS,^{a,b} SARAH RICHARD,^c NATHALIE BERE,^b GERT VAN VALKENHOEF,^a JAYNE GALINSKY,^c ERIC LOW,^c ISABELLE MOULON,^b MARIA MAVRIS,^b TOMAS SALMONSSON,^{b,d} BEATRIZ FLORES,^e HANS HILLEGE,^a FRANCESCO PIGNATTI^b

^aDepartment of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands; ^bEuropean Medicines Agency, London, United Kingdom; ^cMyeloma UK, Edinburgh, United Kingdom; ^dLäkemedelsverket Medical Products Agency, Uppsala, Sweden; ^eMedicines & Healthcare products Regulatory Agency, London, United Kingdom

Disclosures of potential conflicts of interest may be found at the end of this article.

Key Words. Patient preferences • Regulatory science • Benefit-risk assessment • Multicriteria decision analysis

ABSTRACT

Background. The objectives of this study were to elicit the preferences of patients with multiple myeloma regarding the possible benefits and risks of cancer treatments and to illustrate how such data may be used to estimate patients' acceptance of new treatments.

Patients and Methods. Patients with multiple myeloma from the cancer charity Myeloma UK were invited to participate in an online survey based on multicriteria decision analysis and swing weighting to elicit individual stated preferences for the following attributes: (a) 1-year progression-free survival (PFS, ranging from 50% to 90%), (b) mild or moderate toxicity for 2 months or longer (ranging from 85% to 45%), and (c) severe or life-threatening toxicity (ranging from 80% to 20%).

Results. A total of 560 participants completed the survey. The average weight given to PFS was 0.54, followed by 0.32 for

severe or life-threatening toxicity and 0.14 for mild or moderate chronic toxicity. Participants who ranked severe or life-threatening toxicity above mild or moderate chronic toxicity (56%) were more frequently younger, working, and looking after dependent family members and had more frequently experienced severe or life-threatening side effects. The amount of weight given to PFS did not depend on any of the collected covariates. The feasibility of using the collected preference data to estimate the patients' acceptance of specific multiple myeloma treatments was demonstrated in a subsequent decision analysis example.

Conclusion. Stated preference studies provide a systematic approach to gain knowledge about the distribution of preferences in the population and about what this implies for patients' acceptance of specific treatments. *The Oncologist* 2018;23:44–51

Implications for Practice: This study demonstrated how quantitative preference statements from a large group of participants can be collected through an online survey and how such information may be used to explore the acceptability of specific treatments based on the attributes studied. Results from such studies have the potential to become an important new tool for gathering patient views and studying heterogeneity in preferences in a systematic way, along with other methods, such as focus groups and expert opinions.

INTRODUCTION

Benefit-risk assessment is the cornerstone of therapeutic and regulatory decisions. This requires value judgments regarding the relative importance of the expected favorable and unfavorable effects associated with treatment. While the regulator's role is to make decisions that are in the best interest of the patient, it is being increasingly recognized that patients, physicians, and regulators may have different views on the importance of one treatment outcome compared with another (e.g., when the impact of a certain treatment outcome on a

patient's daily life is hard to understand without experience of living with the condition). Information about how individual patient preferences are distributed in the target population could help rendering the decision-making in such preference-sensitive settings more patient-centered [1].

The European Medicines Agency (EMA) has developed a framework for interaction with patients and consumers and their organizations [2]. One of the objectives of the framework is to facilitate participation of patients in benefit-risk

Correspondence: Francesco Pignatti, M.D., Human Medicines Evaluation Division, European Medicines Agency, 30 Churchill Place, London E14 5EU, United Kingdom. Telephone: 44-020-3660-6000; e-mail: francesco.pignatti@ema.europa.eu Received May 31, 2017; accepted for publication September 6, 2017; published Online First on October 27, 2017. <http://dx.doi.org/10.1634/theoncologist.2017-0257>

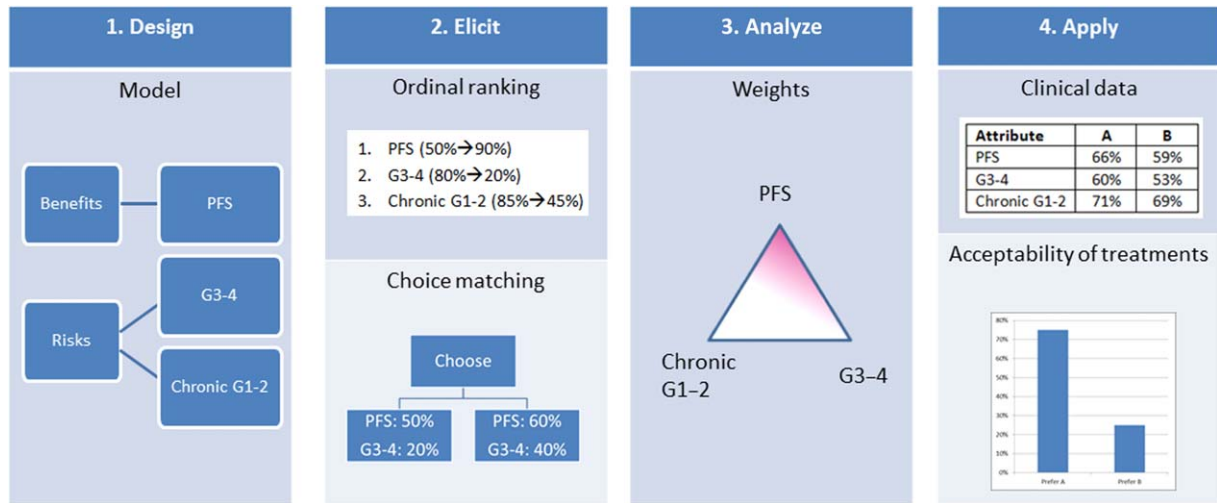


Figure 1. Schematic representation of the steps taken in this study. In step 1 (**Design**), a model is set up with the help of a focus group, with attributes of interest and ranges on which preferences will be collected. In step 2 (**Elicit**), a questionnaire is used first to elicit the ordinal ranking of the attributes considering the full range of alternatives; then choice matching is used to elicit trade-offs by comparing different levels for pairwise attributes, and the process is repeated for all attributes. In step 3 (**Analyze**), the data collected in the survey are transformed into a set of weights reflecting the relative importance of the considered attributes. In Step 4 (**Apply**), the elicited preferences are combined with data from a clinical trial to assess acceptability of two treatments based on the individual elicited weights.

Abbreviations: G1–2, mild or moderate toxicity; G3–4, severe or life-threatening toxicity; PFS, progression-free survival.

evaluation and to explore how methods for eliciting patient preferences can complement other methods for gathering patient views, such as focus groups and asking a few patients to provide input during the assessment. To this end, the EMA previously conducted a pilot study in which individual preferences about hypothetical cancer treatments were elicited from a diverse group of patients, health care professionals, and regulators [3]. While the results from that study suggested that collecting preference data through an online survey based on multicriteria decision analysis (MCDA) and swing weighting was feasible and could provide some indication as to how preferences might vary across stakeholders, it also became apparent that there was considerable heterogeneity in how individual participants valued the attributes of treatments included in the survey. However, because subject characteristics were not collected, it was not possible to further characterize the within-group variability observed in the study. To explore this aspect, we conducted the present study in the area of relapsed or refractory multiple myeloma with a larger sample of patients from the cancer charity Myeloma UK. First, we describe the variation in individual preferences in the surveyed population and the extent to which this variation can be explained by demographic and clinical characteristics. Next, we use a real example to illustrate how the elicited preference statements can be combined with data on progression-free survival (PFS) and toxicity from a randomized clinical trial to estimate the acceptability of two treatments for patients with multiple myeloma (ixazomib in combination with lenalidomide and dexamethasone against placebo in combination with lenalidomide and dexamethasone).

MATERIALS AND METHODS

To elicit individual patient preferences for different attributes of treatments of multiple myeloma, we conducted an online

survey based on MCDA. A schematic representation of the approach used for this study is provided in Figure 1. First, the attributes and attribute levels to include in the survey were selected with the help of a focus group, and a questionnaire was developed to elicit statements regarding the relative importance of those attributes from prospective participants. Subsequently, the survey was administered to patients from the cancer charity Myeloma UK. Next, the ordinal preference statements collected in the survey were transformed into subject-level estimates of the part worth associated with each attribute level by applying the procedure described in the supplemental online material. This was followed by an empirical analysis phase in which the variation in individual preferences was described and further analyzed and a decision analysis phase in which the distribution of the preferences in the surveyed population was combined with data from a clinical trial to estimate the acceptability of two treatment regimens.

Attributes and Attribute Levels

The attributes and attribute levels that were included in the survey are summarized in Table 1. PFS, operationalized as the probability of being progression-free for 1 year or longer, was selected as the measure of treatment benefit. Following our previous pilot study [3], the adverse events were aggregated into two broad categories based on their duration and amount of interference with a patient’s usual or daily activities.

In the questionnaire, PFS was explained as “the time during which treatment can induce and maintain a remission or at least delay the growth of cancer and worsening of symptoms and can delay the need for additional treatment for some time.” Mild or moderate side effects were explained as “side effects that do not immediately require medical intervention but whose symptoms are likely to interfere with normal daily activities. Examples include moderate levels of fatigue (not relieved by rest), diarrhoea (up to 6 stools per day), and moderate pain.” Severe or life-threatening side effects were explained

Table 1. Attributes and attribute levels considered in the survey

Attribute	Definition	Levels
Progression-free survival	Probability of being progression-free for 1 year or longer	50%, 60%, 70%, 80%, 90%
Mild or moderate chronic toxicity	Probability of experiencing mild to moderate side effects for two months or longer	45%, 55%, 65%, 75%, 85%
Severe or life-threatening toxicity	Probability of experiencing severe to life-threatening side effects	20%, 35%, 50%, 65%, 80%

as “side effects that require medical intervention and whose symptoms strongly interfere with normal daily activities. Hospitalisation, dose reduction, or treatment discontinuation may be required. Examples include severe pain (limiting daily activities), skin problems requiring intravenous antibiotics, inflammation affecting eating and swallowing, diarrhoea (7 or more stools per day), and incontinence.”

Preference Model

In MCDA, a participant’s preferences for different combinations of the attribute levels are represented by means of a mathematical function that assigns to each combination of attribute levels a preference score reflecting the relative desirability of this combination of attribute levels compared with all other combinations of attribute levels. These scores are normalized so that the least preferred combination of attribute levels (PFS, 50%; mild or moderate chronic toxicity, 85%; severe or life-threatening toxicity, 80%) is assigned a value of 0 and the most preferred combination of attribute levels (PFS, 90%; mild or moderate chronic toxicity, 45%; severe or life-threatening toxicity, 20%) a value of 1. How the preference scores of the other combinations of attribute levels are calculated depends on the shape of the value function. For this study, we assumed this function to be additive, so that the preference score (overall value) assigned to a combination of attribute levels is the sum of the part worth (partial values) associated with each of the individual attribute levels in this combination. Three combinations of attribute levels of particular interest are the corner point combinations: (a) PFS, 90%; mild or moderate chronic toxicity, 85%; severe or life-threatening toxicity, 80%; (b) PFS, 50%; mild or moderate chronic toxicity, 45%; severe or life-threatening toxicity, 80%; and (c) PFS, 50%; mild or moderate chronic toxicity, 85%; severe or life-threatening toxicity, 20%. The preference scores associated with these corner point combinations reflect how important the “swing” (improvement) from worst to best on one attribute is compared with the swing from worst to best on the other attributes and can therefore be seen as a measure of the relative importance of the attributes. The preference score associated with the corner point combination for which an attribute is at its most preferred level is called the weight of that attribute. By definition, the weights are equal to the part-worth values associated with the attributes’ most preferred levels, which are non-negative and sum to 1. A more detailed discussion of the preference model used in this study can be found in the supplemental online material.

Questionnaire Instrument

The development of the questionnaire went through several phases. First, the overall context of the survey was discussed in a focus group with patients with multiple myeloma, clinicians,

advocates, and regulators. Feedback was obtained on the appropriateness of the chosen attributes, the wording of the electronic consent, elicitation questions, the explanatory text, the complexity of the exercise, and other issues that could affect a patient’s ability or willingness to participate in the study. Subsequently, a first version of the online questionnaire was developed and pretested with a different group of patients with multiple myeloma, from whom we received detailed comments regarding the level of cognitive burden, clarity of the instructions and explanations provided, and ease of use of the elicitation software. Based on this feedback, a revised version of the questionnaire was developed.

A full script of the final version of the questionnaire can be found in the supplemental online material. In short, to assess the attribute weights, participants were first asked to rank the swing from worst to best for each of the three attributes from most important to least important. For example, one participant, say Mr. Smith, may feel that increasing the probability of being progression-free for 1 year or longer from 50% to 90% is more important than decreasing the probability of experiencing severe or life-threatening toxicity from 80% to 20%, and that this in turn is more important than decreasing the probability of experiencing mild or moderate chronic toxicity from 85% to 45%. For the parameterization of Mr. Smith’s preference model, we then know that (a) the weight attached to PFS must be higher than the weight attached to severe or life-threatening toxicity and the weight attached to mild or moderate chronic toxicity and (b) the weight attached to severe or life-threatening toxicity must higher than the weight attached to mild or moderate chronic toxicity. Subsequently, starting from a situation in which the most important attribute (PFS for Mr. Smith, but this could be different for other participants) is at its worst value and the second most important attribute (severe or life-threatening toxicity for Mr. Smith) is at its best value, participants were asked to determine to what extent the performance on the former needs to be improved to offset a swing from best to worst on the latter. This was assessed in an indirect manner by asking the participants to respond to two pairwise comparison questions that narrowed the interval in which this indifference value must lie to a fraction of the attribute’s full-scale range, as described in more detail in Postmus et al. [3]. For example, from these two pairwise comparison questions, we may learn that for Mr. Smith to accept an increase in the probability of experiencing severe or life-threatening toxicity from 20% to 80%, the probability of being progression-free for 1 year or longer must increase from 50% to somewhere between 70% and 80%. For the parameterization of Mr. Smith’s preference model, it must then hold that the weight attached to severe or life-threatening toxicity is between the part worth

associated with a 1-year PFS of 70% and the part worth associated with a 1-year PFS of 80%. Next, this question format was repeated to assess how much participants were willing to trade between the second most important and the least important attribute. Note that the exact trade-offs considered in these four pairwise comparison questions varied from one participant to another depending on how they rank ordered the three attributes and on how they answered preceding pairwise comparison questions. To assess the shape of the partial value functions, participants were asked to rank the increases in value associated with going from one attribute level to another from greatest to smallest, allowing them to assign the same rank to those improvements that they considered to be equally valuable. These questions were the same for all study participants. In addition to the above preference elicitation questions, the survey also contained questions to collect demographic and clinical information and open-ended questions to collect participant feedback about the study.

Participant Recruitment

Patients with multiple myeloma who had given prior permission to receive mailings were invited by the cancer charity Myeloma UK to participate in the study. A prelaunch announcement of the study was also posted on the Myeloma UK website together with an option to register for the study mailing. On June 3, 2016, e-mail invitations with a personal link to the online questionnaire were sent to a total of 2,204 individuals. This was followed by a reminder email on June 10, 2016, and the survey was closed on June 17, 2016.

Empirical Analysis

The subject-level part-worth values associated with each of the attribute levels were summarized using boxplots, and the joint distribution of the attribute weights was presented graphically on a ternary plot. Subsequently, the variation in the weights was further characterized by describing the proportion of participants falling in different segments of the ternary plot. Finally, to explore whether weights differed significantly across different subgroups of participants defined by demographic and clinical characteristics, chi-square tests were performed with an alpha of 5% as the selected level of significance. All these analyses were exploratory.

Decision Analysis Example

A recently approved drug for multiple myeloma was chosen to illustrate how the elicited preference statements can be combined with data on PFS and toxicity from a randomized clinical trial to estimate the acceptability of two treatments for patients with multiple myeloma. Ixazomib is an oral proteasome inhibitor that has recently been approved in the European Economic Area for use in combination with lenalidomide and dexamethasone for the treatment of adult patients with multiple myeloma who have received at least one prior therapy. To estimate the proportion of patients ranking the ixazomib-containing combination above the standard regimen, we conducted a stochastic multicriteria acceptability analysis [4] with the preference model parameters sampled from the empirical distribution of these parameters in the surveyed population and treatment effect estimates obtained from a randomized controlled trial of ixazomib or placebo in combination with a standard regimen of

lenalidomide and dexamethasone in patients with relapsed, refractory, or relapsed and refractory multiple myeloma [5]. The proportion of patients with a PFS of 1 year or longer was estimated from the final statistical analysis dataset for this study endpoint (data cutoff: October 2014). The proportion of patients experiencing mild or moderate chronic toxicity and the proportion of patients experiencing severe or life-threatening toxicity were estimated from the latest safety analysis dataset (data cutoff: July 2015). For mild or moderate chronic toxicity, all National Cancer Institute Common Toxicity Criteria for Adverse Events mild (grade 1) and moderate (grade 2) treatment-related adverse events (i.e., adverse events at least possibly related, as determined by investigator, to any of the three agents in the combination) with a resolution date more than 60 days after toxicity start date or unresolved at the time of the data cutoff but with toxicity start date at least 60 days before data cutoff were included in the estimation of this event rate. For severe or life-threatening toxicity, all treatment-related adverse events of grade 3 or 4 were considered.

RESULTS

Participants

Of the 2,204 individuals who received the email invitation, 795 (36.1%) clicked through to the online questionnaire. Of those visiting the survey website, 650 (81.8%) provided electronic consent after reading the study information sheet and 563 (70.8%) completed the full questionnaire, resulting in a 25.5% response rate and an 86.6% completion rate. Two participants withdrew their consent after completion of the questionnaire, and one participant was excluded due to declaring not to have multiple myeloma, resulting in a study population of 560 participants.

Demographic and clinical characteristics of the survey participants are provided in Table 2. Participants were most frequently between 61 and 70 years of age, and 56% of the participants were male. Most participants did not live alone, were not working, and did not have dependent family members. The time of diagnosis was most frequently more than 4 years ago, and 70% of the participants had received at least two lines of treatment. Forty-four percent of the participants were waiting to start (for instance, newly diagnosed or between therapies) or receiving treatment.

Distribution of Individual Preferences

The subject-level part-worth values of each attribute level are summarized in Figure 2, and the joint distribution of the attribute weights is displayed in Figure 3. The part worth associated with the attributes' least preferred attribute values is normalized to 0 and thus the same for all participants. The part-worth values of the other attribute levels within an attribute are scaled according to how important these attribute levels are compared with the least preferred attribute level. The part worth associated with an attribute's most preferred attribute value is equal to the weight of that attribute and is scaled according to how important the swing from the least preferred to the most preferred attribute level is compared with the swings from worst to best on the other attributes.

Table 2. Demographic and clinical characteristics of the survey participants

Characteristics	Percentage of total group (n = 560)
Men	56%
Age	
60 years or younger	31%
61 to 70 years	44%
71 years and older	25%
White	97%
Working (full-time or part-time)	26%
Living alone	15%
Looking after dependent family members	14%
Time of diagnosis	
Less than 2 years ago	23%
Between 2 and 4 years ago	24%
More than 4 years ago	53%
Current status	
Awaiting or receiving treatment	44%
In remission or not requiring treatment	56%
Previous treatments	
Fewer than two lines of treatment	30%
Two or more lines of treatment	70%
Side effects	
None or mild/moderate, lasting for up to 2 months	31%
Mild/moderate, lasting more than 2 months	57%
Severe/life-threatening	12%
Regular contact with a patient organization/support group	45%
Other chronic health problems	32%

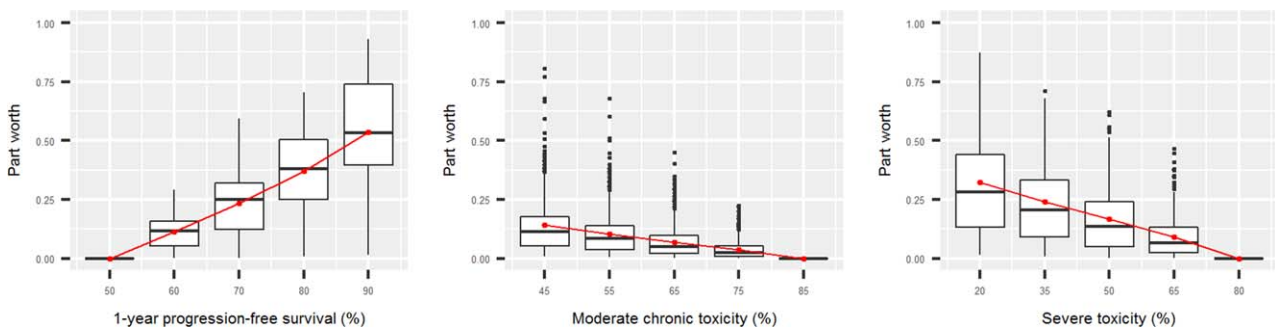
**Figure 2.** Part worth associated with the different levels for each attribute. The red points represent the average (mean) part worth at each attribute level.

Figure 2 shows that the median (quartile 1 to quartile 3 range) weights given to PFS, mild or moderate chronic toxicity, and severe or life-threatening toxicity were 0.53 (0.39–0.74), 0.11 (0.05–0.18), and 0.28 (0.134–0.44), respectively. For 58% of the participants, the weight given to PFS exceeded the cumulative weight given to the two toxicity attributes, meaning that the majority of participants considered increasing the probability of being progression-free for 1 year or longer from 50% to 90% to be more important than simultaneously decreasing the probability of experiencing severe or life-threatening toxicity from 80% to 20% and the probability of

experiencing mild or moderate chronic toxicity from 85% to 45%. In Figure 3, this corresponds to all points that lie above the horizontal gridline at 0.5 on the axis displaying the weight given to PFS. However, as is shown by the distribution of the weights across the different colored areas of the ternary plot, there was considerable heterogeneity among participants with respect to the relative importance given to the two toxicities, with 56% attaching a higher weight to severe or life-threatening toxicity and 44% attaching a higher weight to mild or moderate chronic toxicity. When comparing the demographic and clinical characteristics of these groups (Table 3), we found

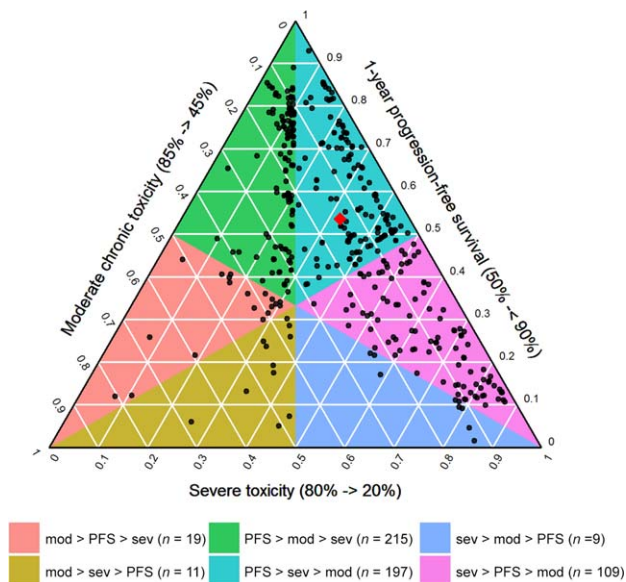


Figure 3. Ternary plot showing the joint distribution of the attribute weights. The left axis displays the weight given to mild or moderate chronic toxicity, the right axis displays the weight given to PFS, and the bottom axis displays the weight given to severe or life-threatening toxicity. The black points represent the attribute weights of the individual study participants, and the red diamond represents the average weight given to the three attributes (0.54 for PFS, 0.32 for severe or life-threatening toxicity, and 0.14 for mild or moderate chronic toxicity). The colored polygons represent areas with a different ordinal ranking of the attribute weights.

Abbreviations: mod, moderate chronic toxicity; PFS, progression-free survival; sev, severe toxicity.

that participants who gave a higher weight to severe or life-threatening toxicity were more frequently age 70 or younger, working, and looking after dependent family members and had more frequently experienced severe or life-threatening side effects. The demographic and clinical characteristics did not differ significantly among those giving a higher weight to PFS and those giving a higher cumulative weight to the two toxicities.

Decision Analysis Example

The probability of being progression-free for 1 year or longer was 66% for the ixazomib-containing regimen and 59% for the placebo-containing regimen [5], the probability of experiencing mild or moderate chronic toxicity was 71% for the ixazomib-containing regimen and 69% for the placebo-containing regimen, and the probability of experiencing severe or life-threatening toxicity was 60% for the ixazomib-containing regimen and 53% for the placebo-containing regimen (data received on file from Takeda Pharma A/S, the marketing authorization holder for Ninlaro [ixazomib]). Based on these effect sizes and the distribution of the individual preferences in the surveyed population, following decision analysis, the proportion of patients ranking the ixazomib-containing regimen above the placebo-containing regimen was 76%.

Participant Feedback

In total, 261 participants (47%) provided free-text comments about the survey. Out of those participants, 31% mentioned that they found some of the questions difficult to understand, while 8% stated that they found the questionnaire easy to comprehend. There were also some participants who mentioned

that the instructions were clear but that the questions themselves were not always easy to answer and that as such the questionnaire was thought-provoking and sometimes even a bit upsetting. Others pointed to the mathematical nature of the questions and that this resulted in the exercise feeling rather academic. Finally, there were several participants who mentioned that some of the questions were a bit repetitive and that they therefore felt that the questionnaire could have been shorter.

DISCUSSION

To understand the preferences of patients with multiple myeloma for the benefits and risks of treatments, we conducted an online survey based on MCDA with 560 participants from the cancer charity Myeloma UK. We found that the average weight given to PFS was higher than the average cumulative weight given to the two toxicity attributes. The amount of weight given to PFS was not associated with any of the collected covariates. However, there was considerable heterogeneity with respect to the relative importance given to those two toxicities, with participants who gave a higher weight to severe or life-threatening toxicity more frequently being younger, working, and looking after dependent family members and having more frequently experienced severe or life-threatening side effects.

Mühlbacher et al. [6] previously conducted a discrete choice experiment to elicit patients' preferences for eight different attributes associated with the treatment of multiple myeloma. In their study, the possibility of having further treatment options and life expectancy was considered to be the most important attribute, while adverse events—ranging from short-term, transient to long-term, permanent—were considered to be the least important attribute. The authors also found that preferences only varied slightly in different subgroups of patients, which is similar to what we found when comparing the weight given to PFS with the cumulative weight given to the two toxicities. However, because they only included one generic toxicity attribute, they were not able to assess the relative importance of different types of toxicity as we did in our study. It is here that we found larger differences in preferences across different subgroups of patients.

Previous patient preference studies in oncology have generally assessed trade-offs over a wider range of attributes, including aspects such as route of administration [7], cost of treatment [8], and length of therapy-free intervals [6]. While all of these attributes are likely to be important considerations in the setting of shared decision-making, whereby patients and health care professionals decide together the best course of action, one should strike a balance between number of attributes and total length of the survey. Furthermore, regulatory decision-making is more narrowly focused on establishing the benefit-risk balance of the products under evaluation to the exclusion of economic considerations. These are the reasons why we only included attributes related to key measures of clinical efficacy and safety in our survey.

A second important consideration is the selection and operationalization of these attributes. In multiple myeloma, quality of response, PFS, and overall survival are all well-established efficacy endpoints. However, because the former two endpoints are generally considered to be related to the latter, they

Table 3. Comparison of the demographic and clinical characteristics between those giving a higher weight to mild or moderate chronic toxicity and those giving a higher weight to severe or life-threatening toxicity

Characteristics	Mild or moderate toxicity more important (n = 245)	Severe or life-threatening toxicity more important (n = 315)	p value
Men	52%	59%	.11
Age			<.001
60 years or younger	26%	35%	
61 to 70 years	41%	47%	
71 years and older	33%	18%	
White	97%	97%	.61
Working (full-time or part-time)	19%	31%	.003
Living alone	15%	14%	.72
Looking after dependent family members	10%	17%	.013
Time of diagnosis			.76
Less than 2 years ago	23%	23%	
Between 2 and 4 years ago	26%	23%	
More than 4 years ago	51%	54%	
Current status			.13
Awaiting or receiving treatment	47%	41%	
In remission or not requiring treatment	53%	59%	
Previous treatments			.074
Less than two lines of treatment	26%	33%	
Two or more lines of treatment	74%	67%	
Side effects			.033
None or mild/moderate, lasting for up to 2 months	30%	32%	
Mild/moderate, lasting more than 2 months	62%	53%	
Severe/life-threatening	8%	15%	
Regular contact with a patient organization/support group	45%	45%	1
Other chronic health problems	35%	31%	.37

should not be simultaneously included as attributes in a multi-criteria model [9]. This is both to ensure that the weights given to the different attributes still have a clear interpretation and to prevent participants from being confronted with seemingly absurd trade-off questions, such as what duration of overall survival they would be willing to forgo to increase the duration of PFS by a specific amount. Patients participating in the focus group at the start of the study considered it more relevant to think in terms of time spent with the disease being under control or in remission than to think in terms of overall survival. They also thought that formulating the elicitation questions in terms of duration of the disease being under control might be less upsetting to some participants. For these reasons, we decided to focus on PFS in our study. The toxicities were aggregated into two generic categories to ensure that our results would be independent of the safety profiles of specific multiple myeloma treatments and therefore more generally applicable when assessing acceptability of treatments. For a more detailed discussion of the pros and cons of aggregating the toxicities into two generic categories rather than including multiple individual side effects, we refer to Postmus et al [3].

Although the survey was built after extensive discussion and pretesting to optimize contextual information provided

and the wording of the elicitation questions, it emerged from the free-text comments at the end of the questionnaire that there was still a group of participants who struggled to understand these questions. It is therefore recommended that further refinements to the survey should aim to make it more user friendly. For example, eliciting weights about probability may be replaced by actual effects (e.g., specific values for the duration of PFS instead of the probability that the duration of PFS exceeds a certain time point), or the mode of administration could be changed from online self-administered to interviewer-guided, which is already a standard way of working in decision support approaches such as decision conferencing [10]. However, such more elaborate elicitation approaches are challenging in a survey-based setting where the objective is to collect preference statements from a large number of geographically dispersed individuals. A limitation of this study is that we did not collect any information on level of education or psychosocial and cognitive factors, meaning that we were not able to further explore how the construction and elicitation of the preferences might have been affected by factors such as health literacy, numeracy, and the willingness of patients to actively engage in the medical decision-making.

The decision analysis example based on ixazomib showed how quantitative preference statements from a larger group of patients could be used in practice by different decision makers when evaluating new treatments. While the example illustrates the use of the method, it should not be considered to replace a comprehensive benefit-risk analysis of the ixazomib combination regimen, as a number of important aspects would require further consideration. First, the Kaplan-Meier survival curves of ixazomib and placebo for PFS started to separate relatively late, with larger differences observed at later time points [5]. Because the benefit-risk assessment was conducted with 1-year PFS as the measure of treatment efficacy, this may have resulted in a possible underrepresentation of the overall treatment effect of ixazomib. Second, as intermittent adverse events were reported as ongoing in our example, the estimates of the proportion of patients experiencing mild or moderate chronic toxicity were likely to include not only adverse events of expected long duration, such as neuropathy, but also intermittent adverse events, such as gastrointestinal events. Finally, the observation time was longer for the ixazomib group, resulting in larger differences in the adverse event rates between the two study groups than what would have been obtained if equal observation times and subsequent treatments were to be considered. Most of these limitations could be addressed using further analyses and assumptions, but this was beyond the purpose of this paper. To allow for further analyses, the empirical distribution of the subject-level preferences as well as an accompanying R script file is available online [11].

CONCLUSION

This study demonstrated how quantitative preference statements from a large group of participants with multiple myeloma can be collected through an online survey and how

such information may be used to explore the heterogeneity of preferences in the population, to identify subgroups with similar preferences, and to estimate the acceptability of specific treatments. Although the usefulness of stated preference studies in drug regulation is still not well established, such studies, along with other methods such as focus groups and expert opinions, have the potential to become an important tool for gathering patient views in a systematic way to inform regulatory and treatment decisions.

ACKNOWLEDGMENTS

We thank the study participants for their time and effort invested in the survey, the patients and experts participating in the focus group and survey test panel for their critical feedback on the survey design, and the researchers from Takeda for providing the clinical trial results for the example. The views presented here are those of the authors and should not be understood or quoted as those of the European Medicines Agency.

AUTHOR CONTRIBUTIONS

Conception/design: Douwe Postmus, Sarah Richard, Nathalie Bere, Gert van Valkenhoef, Maria Mavris, Beatriz Flores, Hans Hillege, Francesco Pignatti
Provision of study material or patients: Sarah Richard
Collection and/or assembly of data: Gert van Valkenhoef
Data analysis and interpretation: Douwe Postmus, Gert van Valkenhoef, Francesco Pignatti
Manuscript writing: Douwe Postmus, Francesco Pignatti
Final approval of manuscript: Douwe Postmus, Sarah Richard, Nathalie Bere, Gert van Valkenhoef, Jayne Galinsky, Eric Low, Isabelle Moulon, Maria Mavris, Tomas Salmonsson, Beatriz Flores, Hans Hillege, Francesco Pignatti

DISCLOSURES

The authors indicated no financial relationships.

REFERENCES

1. Medical Device Innovation Consortium. Medical device innovation consortium (MDIC) patient centered benefit-risk project report: A framework for incorporating information on patient preferences regarding benefit and risk into regulatory assessments of new medical technology. Available at http://mdic.org/wp-content/uploads/2015/05/MDIC_PCBR_Framework_Web1.pdf. Accessed May 24, 2017.
2. European Medicines Agency. Revised framework for interaction between the European Medicines Agency and patients and consumers and their organisations. October 16, 2014. Available at http://www.ema.europa.eu/docs/en_GB/document_library/Other/2009/12/WC500018013.pdf. Accessed May 24, 2017.
3. Postmus D, Mavris M, Hillege HL 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;99:548–554.
4. Tervonen T, van Valkenhoef G, Buskens E et al. A stochastic multicriteria model for evidence-based decision making in drug benefit-risk analysis. *Stat Med* 2011;30:1419–1428.
5. Moreau P, Masszi T, Grzasko N et al. Oral ixazomib, lenalidomide, and dexamethasone for multiple myeloma. *N Engl J Med* 2016;374:1621–1634.
6. Mühlbacher AC, Lincke HJ, Nübling M. Evaluating patients' preferences for multiple myeloma therapy, a Discrete-Choice-Experiment. *Psychosoc Med* 2008;5:Doc10.
7. Landfeldt E, Eriksson J, Ireland S et al. Patient, physician, and general population preferences for treatment characteristics in relapsed or refractory chronic lymphocytic leukemia: A conjoint analysis. *Leuk Res* 2016;40:17–23.
8. Malhotra C, Farooqui MA, Kanesvaran R et al. Comparison of preferences for end-of-life care among patients with advanced cancer and their caregivers: A discrete choice experiment. *Palliat Med* 2015;29:842–850.
9. Tervonen T, Naci H, van Valkenhoef G et al. Applying multiple criteria decision analysis to comparative benefit-risk assessment: Choosing among statins in primary prevention. *Med Decis Making* 2015;35:859–871.
10. Phillips LD. Decision conferencing. In: Edwards W, Miles RF Jr, von Winterfeldt D, eds. *Advances in Decision Analysis: From Foundations to Applications*. Cambridge: Cambridge University Press, 2007:375–399.
11. Postmus D. Stated preference study with multiple myeloma patients. Mendeley Data, v1. May 24, 2017. Available at <https://doi.org/10.17632/n348fw6k85.1>. Accessed July 28, 2017.



See <http://www.TheOncologist.com> for supplemental material available online.