

# Weighing Clinical Evidence Using Patient Preferences: An Application of Probabilistic Multi-Criteria Decision Analysis

Henk Broekhuizen<sup>1</sup> · Maarten J. IJzerman<sup>1</sup> · A. Brett Hauber<sup>2</sup> ·  
Catharina G. M. Groothuis-Oudshoorn<sup>1</sup>

© The Author(s) 2016. This article is published with open access at Springerlink.com

**Abstract** The need for patient engagement has been recognized by regulatory agencies, but there is no consensus about how to operationalize this. One approach is the formal elicitation and use of patient preferences for weighing clinical outcomes. The aim of this study was to demonstrate how patient preferences can be used to weigh clinical outcomes when both preferences and clinical outcomes are uncertain by applying a probabilistic value-based multi-criteria decision analysis (MCDA) method. Probability distributions were used to model random variation and parameter uncertainty in preferences, and parameter uncertainty in clinical outcomes. The posterior value distributions and rank probabilities for each treatment were obtained using Monte-Carlo simulations. The probability of achieving the first rank is the probability that a treatment represents the highest value to patients. We illustrated our methodology for a simplified case on six HIV treatments. Preferences were modeled with normal

distributions and clinical outcomes were modeled with beta distributions. The treatment value distributions showed the rank order of treatments according to patients and illustrate the remaining decision uncertainty. This study demonstrated how patient preference data can be used to weigh clinical evidence using MCDA. The model takes into account uncertainty in preferences and clinical outcomes. The model can support decision makers during the aggregation step of the MCDA process and provides a first step toward preference-based personalized medicine, yet requires further testing regarding its appropriate use in real-world settings.

## Key Points for Decision Makers

Healthcare decisions require an assessment of the value treatments provide for patients. Such assessments are made under uncertainty and there is no consensus about how to account for patient preferences in making these assessments.

This study applies a multi-criteria decision analysis model where clinical evidence is weighted with patient preferences. In this way, patient-weighted treatment values can be estimated in a representative manner while building on the existing the clinical evidence.

The probabilistic approach adopted in the model allows for the simultaneous modelling of measurement uncertainty and patient-specific preference variation. Scenario analyses show that the impact of these different types of uncertainty on decision uncertainty is substantially different in a simplified case on HIV treatments.

**Electronic supplementary material** The online version of this article (doi:10.1007/s40273-016-0467-z) contains supplementary material, which is available to authorized users.

✉ Henk Broekhuizen  
h.broekhuizen@utwente.nl

<sup>1</sup> Department of Health Technology and Services Research, MIRA institute, University of Twente, PO Box 217, 7500 AE Enschede, The Netherlands

<sup>2</sup> RTI Health Solutions, Research Triangle Park, NC, USA

## 1 Introduction

Assessing the relative effectiveness of treatments is an essential component in regulatory and reimbursement decisions and is substantiated by strong methodology and clinical evidence development guidelines. However, interpretation of clinical evidence is a largely subjective and opinion-based process and such judgments are rarely formally included in decision-making processes [1]. Nevertheless, subjective value judgments of stakeholders (in particular those of patients [1–4]) are an essential part of what finally determines treatment value [5]. For instance, in decisions considering two or more equally effective drugs, the (subjective) relative severity of the associated adverse events may dominate the final decision. Currently, subjective value judgments such as the relative severity of adverse events are mostly considered implicitly in healthcare policy decisions. Patient engagement is increasingly promoted [5–7], and several mechanisms of patient engagement have been used, including patient panels and patient-reported outcomes. Whereas decision makers can use such implicit viewpoints, these can still be inaccurate or biased [8, 9]. A potentially more representative and transparent approach is the use of results from survey-based stated preference studies [10]. These studies yield numeric estimates of the relative importance that respondents place on attributes of medical services such as clinical outcomes or other treatment characteristics [1, 2].

Although elicited preferences can be used as a piece of information in the deliberative process of assessing the relative treatment value, a more formal approach would actually use such preferences to weigh clinical endpoints and, thus, prioritize interventions by explicitly mapping the benefits and risks. One method that could be used to structure and analyze decision problems and to explicitly include patient preferences is multi-criteria decision analysis (MCDA). The application of MCDA allows decision makers to structure their decision problem and work on it in a transparent and consistent way [11, 12]. Importantly, MCDA is flexible in that it allows multiple stakeholder groups, including patients, to assign preference weights to clinical outcomes (which are referred to as criteria in MCDA). An MCDA process typically comprises several steps, including (1) definition of the decision problem and alternatives; (2) identification of the decision criteria; (3) weighting of the criteria; (4) identification of the performance of the alternatives; (5) scoring the performance; (6) aggregating the results and dealing with uncertainty; and (7) reporting the results [13].

The use of MCDA in healthcare has grown over the last few years [14, 15], which has led to the International Society For Pharmacoeconomics and Outcomes Research

(ISPOR) MCDA Emerging Good Practices Taskforce publishing guidance on the basic concepts and implementation of MCDA in healthcare recently [13, 16].

The most commonly used method for aggregating the criterion weights and performance scores in MCDA is the linear additive value function. By combining the criterion weights derived from a group of patients and the actual clinical performance scores of the therapeutic alternatives, one can estimate patient-weighted treatment values [17]. This is a relatively straightforward approach but it only reflects the mean clinical treatment performance for an average patient and uncertainty in neither the criteria weights nor performance scores is evaluated.

There are, however, several solutions to include uncertainty in the MCDA model. For instance, Wen et al. [18] and Chim et al. [19] use probability distributions for clinical evidence and point estimates for preferences. Kalso et al. [20] estimate preference uncertainty by defining preference subgroups and coupling these to point estimates for clinical performance. Lynd et al. [21] used point estimates from a discrete choice study to assign utilities to events in a (probabilistic) discrete event simulation. In two other methods, probability distributions for clinical evidence is combined with uniform distributions for criterion weights as a non-informative prior [22, 23]. One common problem of all these methods is that they do not simultaneously combine parameter uncertainty and random variation in both patient preferences and clinical evidence. Here, parameter uncertainty is uncertainty around an estimated quantity (such as a group mean) which can be reduced with more measurements [24, 25]. Random preference variation is the systematic variation of preferences across the population. This variation can, by definition, not be reduced by repeated measurements but can only be better characterized [24, 25]. The importance of including preference variation is recently recognized in the development of personalized medicine based on (variation in) preferences [26, 27].

The aim of this study was to demonstrate an application of probabilistic MCDA that would allow a joint analysis of patient preferences and clinical evidence for treatments, taking into account uncertainty in both preferences and clinical evidence. The proposed model can be applied during the aggregation step of a value-based MCDA and is able to handle three sources of uncertainty: random preference variation, parameter uncertainty in preferences, and parameter uncertainty in clinical evidence. The model is designed to yield value distributions that enable explicit probabilistic statements about which treatments are preferred. The model is illustrated using a simplified case study of highly active antiretroviral therapies (HAART) for HIV patients.

## 2 A Proposed Multi-Criteria Decision Analysis (MCDA) Method for Capturing the Value of Treatments Under Uncertainty

### 2.1 Building the Value Function and Defining Sources of Uncertainty

We adopted a value-based MCDA method. The value of a particular treatment  $i$ , denoted  $V_i$ ,  $i = 1, \dots, n$  was assumed to be a linear additive function of  $K$  criteria (Eq. 1):

$$V_i = \sum_{k=1}^K \beta_k X_{ki} \quad (1)$$

where  $\beta_k$  denotes the preference weight of criterion  $k$ , and  $X_{ki}$  is the performance of treatment  $i$  on criterion  $k$ . We assumed that patients prefer treatments with a higher value to treatments with a lower value, that preference weights and the clinical performances were not correlated, and that the clinical performances were measured on an interval scale.

To introduce random preference variation, suppose each patient  $q$  has his/her own preference weight  $\beta_{kq}$  for each criterion  $k$  (Eq. 2):

$$V_{iq} = \sum_{k=1}^K \beta_{kq} X_{ki} \quad (2)$$

The term  $\beta_{kq}$  in this equation is composed of population mean preference weight  $\beta_k$  and a patient-specific random effect  $\theta_{kq}$  (Eq. 3):

$$\beta_{kq} = \beta_k + \theta_{kq} \quad (3)$$

with (Eq. 4):

$$\theta_{kq} \sim N(0, \sigma_k^2). \quad (4)$$

### 2.2 Parameter Estimators

There are different methods for obtaining preference weights, such as swing weighting. However, in this study the required parameter estimators  $\hat{\beta}_k$  and  $\hat{\sigma}_k^2$  were obtained by analyzing patient-level data from stated preference studies. Parameter uncertainty in these estimators across all criteria is reflected in the covariance matrices  $\sum_{\hat{\beta}}$  and  $\sum_{\hat{\sigma}^2}$ . The estimator for clinical performance  $X_{ki}$  is denoted  $\hat{x}_{ki}$  and can be obtained from clinical trial reports.

### 2.3 Sampling Framework

We assumed that the vectors  $\hat{\beta} = (\hat{\beta}_1 \dots \hat{\beta}_K)$  and  $\hat{\sigma} = (\hat{\sigma}_1 \dots \hat{\sigma}_K)$  were both distributed according to multivariate normal distributions. We denoted the value distribution of treatment  $i$  across the population with  $\psi_i$ , assuming that the

individual values  $V_{iq}$  were identically distributed for patients  $q$ . In a probabilistic model all uncertain parameters are varied at the same time, and this implies that the probability distribution  $\psi_i$  is a complex and analytically challenging combination of the distributions for  $\hat{\beta}_k$ ,  $\hat{\sigma}_k$  and  $\hat{x}_{ki}$ . We therefore approximated  $\psi_i$  with Monte-Carlo simulations (Fig. 1). The Monte-Carlo simulations were programmed as follows. In each simulation run  $t$ , we first sampled a population mean preference weight  $\beta_t = (\beta_{1t} \dots \beta_{Kt})$  from  $MVN(\hat{\beta}, \sum_{\hat{\beta}})$  and a standard deviation  $\sigma_t = (\sigma_{1t} \dots \sigma_{Kt})$  from  $MVN(\hat{\sigma}, \sum_{\hat{\sigma}^2})$ . From that we obtained a respondent  $q_t$  using Eqs. 3 and 4 (Eq. 5):

$$\beta_{q_t} \sim MVN\left(\beta_t, \sum_{\sigma_t^2}\right), \quad (5)$$

with  $\sum_{\sigma_t^2}$  a diagonal covariance matrix with  $\sigma_t^2 = (\sigma_{1t}^2, \dots, \sigma_{Kt}^2)$  on the diagonal. This means that in every simulation run  $t$  a hypothetical patient  $q_t$  with a particular vector of preferences  $\beta_{q_t} = (\beta_{1q_t} \dots \beta_{Kq_t})$  was obtained. The probability distribution of  $\hat{x}_{ki}$  was denoted with  $F_{ki}$  and chosen for each criterion depending on what best modeling practices recommend for that type of clinical performance [24]. From  $F_{ki}$  we sampled  $X_{kit}$  in each simulation run. The preference sample was combined with samples from the probability distributions  $F_{ki}$  for each criterion  $k$  to calculate the value for each treatment  $i$  in the simulation run with (Eq. 6):

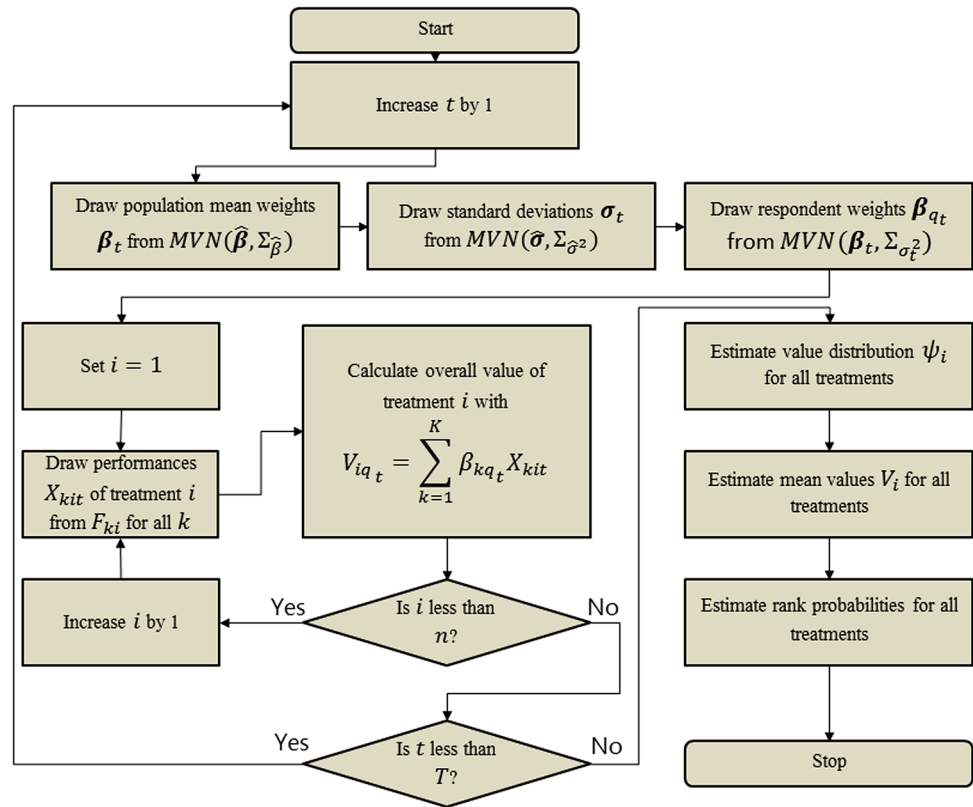
$$V_{iq_t} = \sum_{k=1}^K \beta_{kq_t} X_{kit}. \quad (6)$$

The Monte-Carlo simulation process was repeated a large number of times  $T$  and was programmed in R [28].

### 2.4 Model Outcomes

From the Monte-Carlo simulations the mean value of each treatment in the population was obtained, calculated as the posterior mean,  $V_i = \frac{\sum_{i=1}^T V_{iq_t}}$ . This metric can be interpreted as the mean perceived value of the treatment's clinical performance according to patients. Since we assumed that patients prefer treatments with a higher value to treatments with a lower value, the treatment value should only be interpreted relative to the value of other treatments. The degree to which the value distributions of two treatments overlap indicates how uncertain we are about selecting the treatment with the highest value. More concretely, to make probabilistic statements about the degree to which we are sure about which treatment has the

**Fig. 1** Overview of the Monte-Carlo simulation method used in the model.  $i = 1 \dots n$  the treatment, and  $t = 1 \dots T$  the Monte Carlo simulation run



highest value in each simulation,  $t$  treatments were ranked from the first (highest value) rank to the last (lowest value) rank based on their respective  $v_{iq_t}$ . The rank achieved by treatment  $i$  in simulation run  $t$  was denoted with  $r_{it}$ . We denoted the rank of treatment  $i$  with  $R_i$ . The probability that this rank equals  $y$  was estimated with the percentage of simulations treatment  $i$  had rank  $y$  (Eqs. 7 and 8):

$$\frac{\sum_{t=1}^T 1_y(r_{it})}{T} \quad (7)$$

where

$$1_y(r_{it}) = \begin{cases} 1, & \text{if } r_{it} = y \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

Since we assumed patients prefer treatments with a higher value to treatments with a lower value, the treatment with the highest estimated probability of being ranked first was considered the most preferred treatment. One minus the probability that the most preferred treatment is ranked first is the rank reversal probability for the first rank. The rank reversal probability for the first rank is used as a measure of decision uncertainty, because it is the probability that the most preferred treatment (according to our model) turned out not to be the most valuable treatment.

## 2.5 An Illustration of the Proposed Approach on a Simplified HIV Case

To illustrate the proposed model, it is applied to the case of HIV treatments. The MCDA was designed according to the guidelines as proposed by the ISPOR taskforce [13]. Although the data used comes from real studies, the main purpose of this paper is to illustrate the modeling approach. The simulation results presented in this paper can therefore not be used to inform clinical HIV decisions. All data used (with references) as well as the R script can be found in the Electronic Supplementary Material to replicate our results.

## 2.6 Identification of Treatment Alternatives and Decision Criteria

The example is about the comparison of the relative value of HAARTs for HIV-positive patients from a regulatory perspective. Treatments under consideration are those recommended by the US National Institutes of Health (NIH) for treatment-naïve patients [29]. A HAART consists of an active drug component and one of two backbones (abacavir/lamivudine [AL] or tenofovir/emtricitabine [TE]). The treatments included in the case study were four combination treatments

(dolutegravir + AL/TE,<sup>1</sup> efavirenz + AL, raltegravir + TE, atazanavir/ritonavir + TE, elvitegravir/cobicistat + TE, and darunavir/ritonavir + TE) plus two backbone-only treatments (AL and TE). For simplicity we denote combination treatments only by their active drug component (e.g., dolutegravir instead of dolutegravir + AL/TE, etc.) in the remainder of this paper.

In the MCDA model we used the same criteria as identified in an earlier preference study [30], namely probability of virologic failure ( $p_{\text{viro}}$ ), probability of allergic reaction ( $p_{\text{all}}$ ), probability of bone damage ( $p_{\text{bone}}$ ), and probability of kidney damage ( $p_{\text{kid}}$ ) [30]. All probabilities were all defined over a 52-week time horizon.

## 2.7 Measuring Performance of the Alternatives

The clinical performance of the treatments on the included criteria was obtained from clinical trial reports (for an overview of the included trials see the NIH guideline [29]). All criteria were represented as probabilities. For each treatment we therefore retrieved from the clinical trial reports the number of events ( $e_k^+$ ) and number of non-events ( $e_k^-$ ) per criterion  $k$ . The performance estimates for each criterion  $k$  were then calculated with  $\frac{e_k^+}{e_k^+ + e_k^-}$ . For studies that reported time horizons other than 52 weeks, we used linear extrapolation to obtain performance estimates for a 52-week time horizon. We used a linear partial value function for each criterion. No rescaling of the performances was needed for the partial value functions since probabilities are already naturally constrained between 0 and 1. For criteria measured on other scales, such a rescaling to the  $[0, 1]$  range using a predefined lower and upper level would be required [31].

Plasma HIV RNA of more than 50 copies/mL 52 weeks after treatment start was considered a virologic failure event. Reported incidence of rash was used as a surrogate measure for the TE-induced allergic reaction event. Reported fractures in the clinical trials were considered to constitute bone damage events. However, this was not reported for TE; therefore, decreases in bone mineral density of more than 6% were used as a surrogate endpoint. Because of the diagnostic and treatment options available, we limited our model to include only treatable bone damage [29, 32]. Reported cases of renal failure were considered to be kidney damage events, and we therefore limited our model to include only non-treatable kidney damage [33].

<sup>1</sup> Either AL or TE.

## 2.8 Input for Criteria Weights: Results from a Previously Published Discrete Choice Experiment

The present methodological study did not elicit criteria weights itself, but used the results from a previous stated preference study to inform the preference weights. In that study by Hauber et al. [30], 147 treatment-naïve HIV-positive African Americans gave their preferences for four criteria relevant for HIV treatment. The study employed a discrete choice study design with 24 choice tasks. In the choice tasks the criteria were defined as the probability of the event in the next 52 weeks. An overview of the preference data is given in Table 1.

## 2.9 Aggregating Scores and Performance

In the base-case analysis we assumed an additive value function. The value of treatment  $i$  for patient  $q$  (Eq. 2) could therefore be specified as Eq. 10:

$$V_{iq} = p_{\text{viro},i} \beta_{\text{viro},q} + p_{\text{all},i} \beta_{\text{all},q} + p_{\text{bone},i} \beta_{\text{bone},q} + p_{\text{kid},i} \beta_{\text{kid},q}. \quad (10)$$

The mean population preferences and the random preference variations were assumed to be normally distributed. The parameter estimates for these distributions were obtained from a (mixed logit [34]) analysis of the used preference study (Table 1). The performances of each treatment are probabilities and were therefore modeled with beta distributions [24]. Beta distributions require two input parameters ( $\alpha_1$  and  $\alpha_2$ ) and these were estimated from the clinical trials cited in the NIH guideline. A complete overview of the clinical trials used in this study can be found in the Electronic Supplementary Material. We used the number of reported events  $e_k^+$  as  $\alpha_1$  and the number of non-events  $e_k^-$  as  $\alpha_2$ . In total, 100,000 Monte-Carlo simulations were performed.

## 2.10 Handling Uncertainty: Scenario Analyses

A scenario analysis was employed to illustrate the effect of the different sources of uncertainty on the treatment value distributions. The base-case scenario described previously considered all types of uncertainty simultaneously, i.e., parameter uncertainty in preferences, random preference variation, and parameter uncertainty in clinical evidence. Each of the other scenarios tested for one specific source of uncertainty. This means that each scenario included one source of uncertainty while the other two sources were assumed to have no uncertainty and were thus fixed at a particular value. In the first scenario, only parameter uncertainty in preferences (as parameterized by the



**Table 1** Preference data used from Hauber et al. [30]. All  $\hat{\beta}$  are per cent point probability of the event occurring in the next 52 weeks, i.e., the partial value of a 2% probability of allergic reaction in the next 52 weeks is  $-0.12$ . Note that the covariance  $\sum_{\hat{\beta}}$

and  $\sum_{\hat{\sigma}^2}$  are not presented here for brevity but can be found in the Electronic Supplementary Material. Both  $\hat{\beta}_k$  and  $\hat{\sigma}_k$  are assumed to be distributed with a multivariate normal distribution

	Levels used in DCE (%) [21]	$\hat{\beta}$	SE( $\hat{\beta}$ )	$\hat{\sigma}_k$	SE( $\hat{\sigma}_k$ )
Virological failure prevented	96.0, 85.0, 79.0	-0.05	0.01	0.05	0.02
Allergic reaction	0.0, 1.0, 8.0, 12.0	-0.06	0.01	0.06	0.02
Bone damage (treatable)	0.0, 1.0, 5.0, 10.0	-0.01	0.02	0.06	0.06
Kidney damage (not treatable)	0.0, 1.0, 5.0, 10.0	-0.22	0.04	0.21	0.05

DCE discrete choice experiment

**Table 2** Clinical evidence used. For references to all included clinical trials, see the National Institutes of Health guideline [29] and the Electronic Supplementary Material. All performances were

defined over a 52-week time horizon and assumed to be distributed with beta distributions

HAART regimen	Probability of virological failure (95% CI)	Probability of allergic reaction (95% CI)	Probability of bone damage (95% CI)	Probability of kidney damage (95% CI)
Dolutegravir + TE/AL	7.61% (6.10–9.25)	0.24% (0.01–0.90)		0.09% (0.00–0.34)
Atazanavir/ritonavir + TE	5.25% (3.94–6.77)	1.72% (0.74–3.10)	1.29% (0.48–2.5%)	0.43% (0.05–1.21)
Elvitegravir/cobicistat + TE	13.52% (10.17–17.34)		2.02% (0.83–3.77%)	0.87% (0.18–2.12)
Efavirenz + AL	13.67% (12.32–15.13)	1.86% (1.21–2.66)	2.12% (1.41–2.98%)	0.29% (0.11–0.56)
AL (backbone)	16.3% (13.74–18.99)	2.32% (1.01–4.18)	4.57% (3.02–6.38%)	2.50% (1.47–3.75)
TE (backbone)	20.41% (17.58–23.39)	1.01% (0.28–2.19)	2.19% (1.18–3.52%)	3.84% (2.56–5.36)

AL abacavir/lamivudine, CI confidence interval, HAART highly active antiretroviral therapy, TE tenofovir/emtricitabine

covariance matrix  $\sum_{\hat{\beta}}$  was taken into account. Therefore,  $\theta_{kq}$  was set to zero for all criteria and individuals, and all clinical performances were set to the mean clinical performance of treatments as found in Table 2. In the second scenario, only random preference variation was included. This means that the  $\sum_{\hat{\beta}}$  was set to be a zero matrix, all clinical performances were set to the mean clinical performance of treatments as found in Table 2, and  $\theta_{kq}$  was distributed as in the base case. In the third scenario, only parameter uncertainty in clinical performance was included. This means that  $\theta_{kq}$  was set to zero for all criteria and individuals,  $\sum_{\hat{\beta}}$  was set to be a zero matrix, and performances  $X_{ki}$  were distributed with  $F_{ki}$  as in the base case.

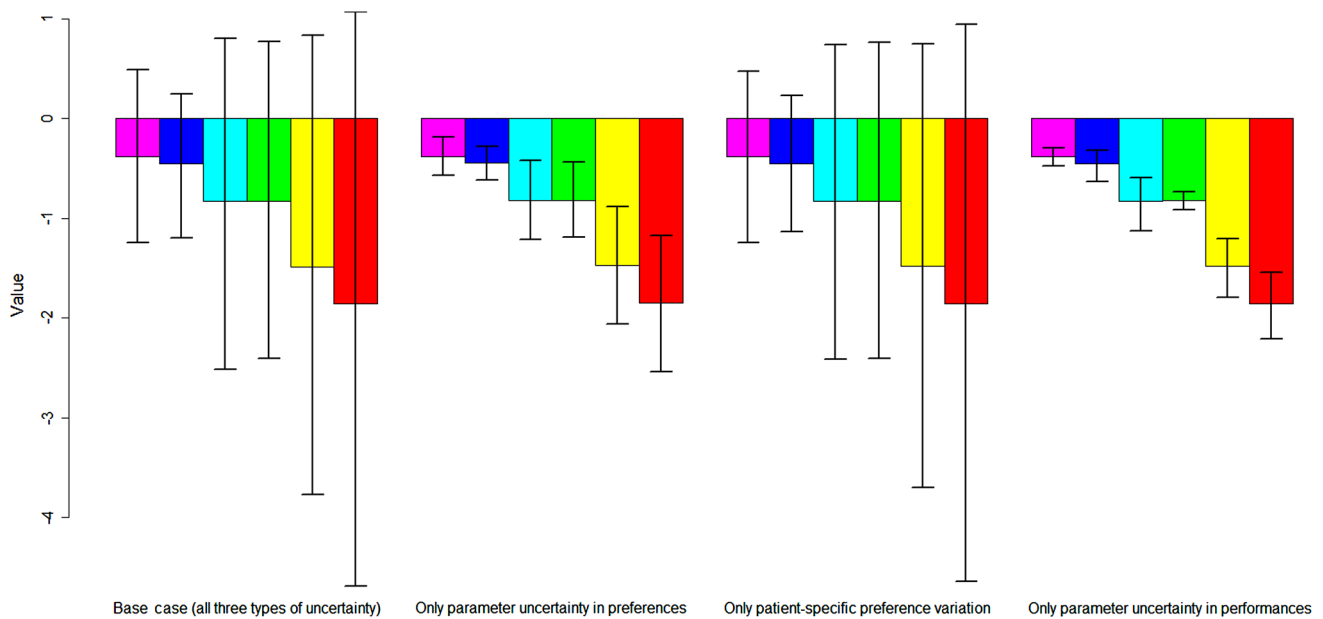
### 3 Examination and Discussion of Findings

#### 3.1 Outcomes of Case Study

The model outcome was a value distribution for each included HAART (Fig. 2). In the base case, dolutegravir has the highest patient-weighted estimated mean value of  $-0.39$  with an empirical 95% confidence interval (CI) running from  $-1.25$  to  $0.48$ . The backbone-only treatments

AL and TE had the lowest mean values ( $-1.49$  and  $-1.86$ , respectively). In all scenario analyses, mean treatment values were similar to the base-case results (Fig. 2; Table 3) and the most likely rank order of treatments did not change. The width of the CIs did vary between the base case and the scenarios. When only random preference variation was considered in the second scenario, the CIs were only slightly narrower than those of the base case. In the two other scenarios that considered parameter uncertainty the CIs were substantially narrower.

In 49.1% of the simulations, dolutegravir was ranked first (Table 4). This implies that in 50.9% of simulations, another treatment was preferred. Atazanavir/ritonavir had the highest probability of being ranked second (40.0%) and efavirenz had the highest probability of being ranked third (40.3%). The narrower value distributions in the scenarios are reflected in the rank probabilities. The rank probabilities in the scenario that considered only parameter uncertainty in preferences were higher: the probability of dolutegravir attaining first rank and atazanavir/ritonavir attaining second rank were both more than 90%. This means that we were confident about the ranking of treatments for the first and second rank. Similarly, the first and second rank probabilities for dolutegravir and atazanavir/ritonavir are more than 75% in the scenario that considers



**Fig. 2** Barplots of the regimen values with 95% confidence intervals across the four analysis scenarios. *Purple* dolutegravir, *dark blue* elvitegravir/cobicistat, *light blue* atazanavir/ritonavir, *green* efavirenz, *yellow* abacavir/lamivudine, *red* tenofovir/emtricitabine

**Table 3** Values (with 95% confidence intervals) for the included highly active antiretroviral therapy regimens across the four analyses

HAART regimen	Base case	Scenario analyses		
	All three types of uncertainty	Only parameter uncertainty in preferences	Only patient-specific preference variation	Only parameter uncertainty in performances
Dolutegravir + TE/AL	-0.39 (-1.25 to 0.48)	-0.38 (-0.57 to -0.19)	-0.39 (-1.24 to 0.47)	-0.38 (-0.48 to -0.30)
Atazanavir/ritonavir + TE	-0.46 (-1.2 to 0.24)	-0.45 (-0.62 to -0.28)	-0.45 (-1.14 to 0.23)	-0.45 (-0.64 to -0.32)
Elvitegravir/cobicistat + TE	-0.83 (-2.52 to 0.8)	-0.82 (-1.22 to -0.42)	-0.83 (-2.41 to 0.74)	-0.83 (-1.13 to -0.59)
Efavirenz + AL	-0.83 (-2.4 to 0.77)	-0.82 (-1.19 to -0.44)	-0.83 (-2.41 to 0.76)	-0.82 (-0.92 to -0.73)
AL backbone	-1.49 (-3.77 to 0.83)	-1.47 (-2.06 to -0.89)	-1.48 (-3.7 to 0.75)	-1.48 (-1.79 to -1.20)
TE backbone	-1.86 (-4.68 to 1.07)	-1.85 (-2.54 to -1.17)	-1.85 (-4.64 to 0.94)	-1.86 (-2.21 to -1.55)

AL abacavir/lamivudine, HAART highly active antiretroviral therapy, TE tenofovir/emtricitabine

only parameter uncertainty in clinical performances. From the fourth until last rank, there is slightly more decision uncertainty in the scenario that considers parameter uncertainty in preferences than there is in the scenario that considers parameter uncertainty in clinical performances.

### 3.2 Implications of the Modeling Framework for Personalized Medicine

The main objective of this study was to develop a novel methodology. The HIV case was used to demonstrate the concepts and cannot be used for guidance on HIV treatment decisions. However, the modelling framework does allow an exploration of its usefulness for other applications: assessing the impact of uncertain parameters (clinical evidence and preferences) on decision uncertainty and

personalizing treatment based on preferences. The assessment of the impact of the uncertainty in model parameters on decision uncertainty is important for identifying the value of additional research: it is most worthwhile to further investigate model parameters where more information would most likely reduce decision uncertainty. In the presented case, there is a clear difference between the impact of the sources of uncertainty. When only parameter uncertainty in either preferences or performances was considered, one treatment was clearly the most valuable treatment. However, when random preference variation was considered, there was considerable overall decision uncertainty. The value of additional clinical research may therefore not be high but decision aids (that help patients think about their preferences, reducing their individual preference uncertainty) may be valuable.

**Table 4** Ranking probabilities for all included regimens across the four analyses

	Dolutegravir + TE/AL (%)	Atazanavir/ ritonavir + TE (%)	Elvitegravir/ cobicistat + TE (%)	Efavirenz + AL (%)	AL backbone (%)	TE backbone (%)
Base case						
All three types of uncertainty						
1	49.07	34.21	4.95	4.28	1.92	5.57
2	36.12	39.98	10.53	7.54	4.08	1.75
3	5.56	8.69	40.25	40.31	3.27	1.92
4	3.46	8.47	39.96	39.00	5.69	3.42
5	4.60	3.52	3.76	4.96	69.52	13.64
6	1.19	5.13	0.55	3.91	15.52	73.70
Scenario analyses						
1: Only parameter uncertainty in preferences						
1	90.42	9.58	0.00	0.00	0.00	0.00
2	9.58	90.26	0.14	0.02	0.00	0.00
3	0.00	0.11	46.26	53.63	0.00	0.00
4	0.00	0.05	53.60	46.35	0.00	0.00
5	0.00	0.00	0.00	0.00	100.00	0.00
6	0.00	0.00	0.00	0.00	0.00	100.00
2: Only patient-specific preference variation						
1	53.28	32.25	2.34	5.32	1.59	5.22
2	33.77	46.18	8.50	6.43	3.93	1.19
3	4.68	5.81	42.71	42.69	2.07	2.04
4	2.76	8.52	44.4	37.84	3.69	2.79
5	4.41	2.46	1.90	3.49	79.51	8.23
6	1.10	4.78	0.15	4.23	9.21	80.53
3: Only parameter uncertainty in performances						
1	77.2	22.79	0.01	0.00	0.00	0.00
2	22.8	76.41	0.68	0.11	0.00	0.00
3	0.00	0.74	51.00	48.26	0.00	0.00
4	0.00	0.06	48.19	51.63	0.12	0.00
5	0.00	0.00	0.12	0.00	95.2	4.68
6	0.00	0.00	0.00	0.00	4.68	95.32

AL abacavir/lamivudine, TE tenofovir/emtricitabine, TE/AL either tenofovir/emtricitabine or abacavir/lamivudine

This relates to the second application for which our work may be used: personalized medicine. Although most research in that field has focused on personalizing treatment based on clinically measurable patient characteristics such as genetic differences, there is an increasing interest in personalizing treatment based on patient preferences [26]. Our model could help decision makers to make the first steps toward such a personalization by showing to what extent (uncertainty in) patient preferences influences the choice of treatment. A next step to formalize the assessment of the extent to which differences in preferences are relevant would be to calculate metrics such as the value of heterogeneity [35]. This metric shows the marginal population-wide value gained from having patients choose between more than one treatment. This value will

be low if there is a clear most valuable treatment for all patients, but it will be high if there is clinical equipoise and/or much patient-specific preference variation. An important outstanding research issue here is how values derived from an MCDA can be contrasted with financial costs [36]. Further research could also be directed toward the integration of patient-specific clinical outcome measures; that is, introducing a patient-specific performance distribution that yields estimates of the performance of a specific treatment for a specific patient. For the HIV case this could, for example, be operationalized by estimating the treatability of kidney damage based on respondent characteristics such as level of kidney functioning at treatment start. Such a holistic view of patient-specific variation in both preferences and clinical outcomes would



be a step toward combining the two current viewpoints on personalized medicine [26].

#### 4 Strengths, Weaknesses, and a Comparison to the Existing Literature

The first strength of the current study is that it has demonstrated a methodological approach for combining preference data and clinical data into one value metric, which may contribute to the ongoing attempts to integrate patient preference research in health technology assessment and market approval. A second strength of the study is that the developed model allows for the simultaneous consideration of the impact of three sources of uncertainty, i.e., random preference variation and parameter uncertainty in both clinical and preference estimates.

The developed model uses results from stated preference studies to inform value judgements in policy decisions from one of the most important stakeholders in healthcare: the patient. By including stated preference studies in the model, the patient preferences are incorporated in an explicit, structured and representative manner. Many different types of preference elicitation methods exist. In this study we used results from a discrete choice study, but in theory our model is able to handle preference weights obtained from a wide range of preference elicitation methods as long as a value function can be constructed and probability distributions can be assigned to weights. Finally, the possibility of including preference studies with large sample sizes allows for the investigation of variation and uncertainty in preferences. The impact of these was investigated by assigning informative probability distributions, which sets our study apart from earlier studies that have used point estimates [18–21] or non-informative (uniform) distributions [22, 23].

The decision to use a probabilistic approach in the present study follows from a recent review identifying five approaches to deal with uncertainty in MCDA [25]. The approach adopted in this study seems most advantageous for our aims for a number of reasons. It is the approach that is best suited for dealing with the preferences of a group of stakeholders and is most able to consider multiple uncertain parameters [25]. Another advantage is that it is possible to implement Monte-Carlo simulations as a flexible method that can combine all types of parametric and non-parametric probability distributions [24]. For these reasons, decision makers could apply the method during the aggregation and uncertainty steps of the MCDA process described in the recent ISPOR taskforce report [13]. This would be especially advantageous in policy decisions where the patient perspective is considered explicitly and

where various sources of uncertainty are relevant to consider.

Even though the model may be appropriate given our aims, a limitation of all studies that use results from stated preference research is that a person's stated preference may not be the same as their revealed preference [37]. A limitation specific to the illustrative case study was that treatment value is assumed to be linearly related to the clinical performance. This assumption implies that we assumed we could extrapolate linearly beyond the performance levels originally included in the preference study to calculate partial values. The linearity specification could not be rejected in the preference study but this could also have been due to the sample size [30]. A related limitation is that we assumed that we could extrapolate performance measures to conform to the 52-week time horizon used in the preference study. Furthermore, another limitation is that random preference variation was assumed to be normally distributed. Although this is practical and a commonly made assumption in patient preference research [38], in our study it resulted in a small percentage of Monte-Carlo simulations having sign reversals for the preference weights, mainly for the virologic failure criterion. In a larger patient sample it may have been possible to estimate a more specific functional form for the value function [39]. Performance samples for virologic failure all fell outside of the range for which preferences were elicited, which may have biased the value estimates.

#### 5 Conclusion

In an attempt to explore new approaches for increasing patient engagement in healthcare policy decisions, the current paper presents a probabilistic MCDA model in which treatment values were estimated by weighting clinical trial evidence with results from a patient preference study. The model outcomes were patient-weighted probability distributions of relative treatment value and the respective rank probabilities. The developed model was illustrated using a simplified case study. The adopted probabilistic approach integrates random preference variation and parameter uncertainty in patient preferences with parameter uncertainty in clinical evidence using a Monte-Carlo simulation method. Further research about the use of the modelling approach for non-simplified cases and the match to decision maker needs is required.

#### Compliance with Ethical Standards

**Author contributions** All authors contributed to the study conception and design. HB collected the data from literature and ran the analyses. HB, ABH, and CGMG-O developed the model. HB, MJJ,

and CGMG-O drafted the manuscript. All authors contributed to the critical revision of intellectual content in the final manuscript. All authors approved the final version for submission. CG is the guarantor for the overall content.

**Funding** No funding from external sources was provided for this work.

**Conflicts of interest** Henk Broekhuizen, Maarten IJzerman, Brett Hauber, and Catharina Groothuis-Oudshoorn declare no conflicts of interest.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

1. Hauber AB, Fairchild AO, Reed Johnson F. Quantifying benefit-risk preferences for medical interventions: an overview of a growing empirical literature. *Appl Health Econ Health Policy*. 2013;11:319–29.
2. Weernink MGM, Janus SIM, van Til JA, Raisch DW, van Manen JG, IJzerman MJ. A systematic review to identify the use of preference elicitation methods in healthcare decision making. *Pharmaceut Med*. 2014;28:175–85.
3. van Til JA, IJzerman MJ. Why should regulators consider using patient preferences in benefit-risk assessment? *Pharmacoeconomics*. 2013;32:1–4.
4. MDIC PCBR project group members. A framework for incorporating information of patient preferences regarding benefit and risk into regulatory assessments of new medical technology. Minneapolis: Medical Device Innovation Consortium; 2015.
5. Eddy D. Anatomy of a decision. *JAMA*. 1990;263:441–3.
6. Hofmann B, Cleemput I, Bond K, Kronen T, Droste S, Sacchini D, et al. Revealing and acknowledging value judgments in health technology assessment. *Int J Technol Assess Health Care*. 2015;30:579–86.
7. EMA. Reflection paper on benefit-risk assessment methods in the context of the evaluation of marketing authorisation applications of medical products for human use: report of the CHMP working group on benefit-risk assessment methods. London; 2008. [http://www.emea.europa.eu/docs/en\\_GB/document\\_library/Regulatory\\_and\\_procedural\\_guideline/2010/01/WC500069634.pdf](http://www.emea.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2010/01/WC500069634.pdf). Accessed 30 June 2016.
8. de Bekker-Grob EW, Essink-Bot ML, Meerding WJ, Koes BW, Steyerberg EW. Preferences of GPs and patients for preventive osteoporosis drug treatment: a discrete choice experiment. *Pharmacoeconomics*. 2009;27:211–9.
9. Johnson FR, Hauber AB, Özdemiş S, Siegel CA, Hass S, Sands BE. Are gastroenterologists less tolerant of treatment risks than patients? Benefit-risk preferences in Crohn's disease management. *J Manag Care Pharm*. 2010;16:616–28.
10. Reed Johnson F, Lancsar E, Marshall D, Kilambi V, Mühlbacher A, Regier DA, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. *Value Health*. 2013;16:3–13.
11. Keeney R, Raiffa H. Decisions with multiple objectives. Cambridge: Cambridge University Press; 1976.
12. Belton V, Stewart TJ. Multiple criteria decision analysis: an integrated approach. 2nd ed. Dordrecht: Kluwer Academic; 2002.
13. Marsh K, IJzerman M, Thokala P, Baltussen R, Boysen M, Kalo Z, et al. Multiple criteria decision analysis for health care decision making—emerging good practices: report 2 of the ISPOR MCDA Emerging Good Practices Task Force. *Value Health*. 2016;19:1–13.
14. Diaby V, Campbell K, Goeree R. Multi-criteria decision analysis (MCDA) in health care: a bibliometric analysis. *Oper Res Health Care*. 2013;2:20–4.
15. Marsh K, Lanitis T, Neasham D, Orfanos P, Caro J. Assessing the value of healthcare interventions using multi-criteria decision analysis: a review of the literature. *Pharmacoeconomics*. 2014;32:1–21.
16. Thokala P, Devlin N, Marsh K, Baltussen R, Boysen M, Kalo Z, et al. Multiple criteria decision analysis for health care decision making—an introduction: report 1 of the ISPOR MCDA Emerging Good Practices Task Force. *Value Health*. 2015;19:1–13.
17. Hummel JM, Volz F, van Manen JG, Danner M, Dintsios CM, IJzerman MJ, et al. Using the analytic hierarchy process to elicit patient preferences. *Patient*. 2012;5:1–13.
18. Wen S, Zhang L, Yang B. Two approaches to incorporate clinical data uncertainty into multiple criteria decision analysis for benefit-risk assessment of medicinal products. *Value Health*. 2014;17:619–28.
19. Chim L, Salkeld G, Stockler MR, Mileshekin L. Weighing up the benefits and harms of a new anti-cancer drug: a survey of Australian oncologists. *Intern Med J*. 2015;45:834–42.
20. Kaltoft M, Turner R, Nielsen J, Cunich M, Salkeld G, Dowie J. Addressing preference heterogeneity in public policy by combining cluster analysis and multi-criteria decision analysis. *Health Econ Rev*. 2014;5:1–11.
21. Lynd LD, Najafzadeh M, Colley L, Byrne MF, Willan AR, Sculpher MJ, et al. Using the incremental net benefit framework for quantitative benefit-risk analysis in regulatory decision-making—a case study of alosetron in irritable bowel syndrome. *Value Health*. 2009;13:1–7.
22. Tervonen T, van Valkenhoef G, Buskens E, Hillege HL, Postmus D. A stochastic multicriteria model for evidence-based decision making in drug benefit-risk analysis. *Stat Med*. 2011;30:1419–28.
23. Caster O, Norén GN, Ekenberg L, Edwards IR. Quantitative benefit-risk assessment using only qualitative information on utilities. *Med Decis Making*. 2012;32:E1–15.
24. Briggs AH, Weinstein MC, Fenwick EA, Karnon J, Sculpher MJ, Paltiel AD; ISPOR-SMDM Modeling Good Research Practices Task Force. Model parameter estimation and uncertainty: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-6. *Value Health*. 2012;15:835–42.
25. Broekhuizen H, Groothuis-Oudshoorn CGM, van Til JA, Hummel JM, IJzerman MJ. A review and classification of approaches for dealing with uncertainty in multi-criteria decision analysis for healthcare decisions. *Pharmacoeconomics*. 2015;33:445–55.
26. Rogowski W, Payne K, Schnell-Inderst P, Manca A, Rochau U, Jahn B, et al. Concepts of “personalization” in personalized medicine: implications for economic evaluation. *Pharmacoeconomics*. 2015;33:49–59.
27. Ho MP, Gonzalez JM, Lerner HP, Neuland CY, Whang JM, McMurphy-Heath M, et al. Incorporating patient-preference evidence into regulatory decision making. *Surg Endosc*. 2015;29:2984.
28. R Development Core Team. R: a language and environment for statistical computing. Vienna: R Development Core Team; 2015. <http://www.r-project.org>. Accessed 30 June 2016.

29. Panel on Antiretroviral Guidelines for Adults and Adolescents. Guidelines for the use of antiretroviral agents in HIV-infected adults and adolescents. Department of Health and Human Services; 2014. <http://aidsinfo.nih.gov/ContentFiles/AdultandAdolescentGL.pdf>. Accessed 26 May 2014.
30. Hauber AB, Mohamed AF, Watson ME, Johnson FR, Hernandez JE. Benefits, risk, and uncertainty: preferences of antiretroviral-naïve african americans for HIV treatments. *AIDS Patient Care STDS*. 2009;23:1–6.
31. Dodgson J, Spackman M, Pearman A, Phillips L. Multi-criteria analysis: a manual. London; 2009. [http://eprints.lse.ac.uk/12761/1/Multi-criteria\\_Analysis.pdf](http://eprints.lse.ac.uk/12761/1/Multi-criteria_Analysis.pdf). Accessed 30 June 2016.
32. Walker Harris V, Brown TT. Bone loss in the HIV-infected patient: evidence, clinical implications, and treatment strategies. *J Infect Dis*. 2012;205(Suppl):S391–8.
33. Molitoris B. Acute kidney injury. In: Goldman L, Schafer A, editors. *Goldman's Cecil medicine*. 24th ed. Philadelphia: Saunders Elsevier; 2011.
34. McFadden D, Train K. Mixed MNL models for discrete response. *J Appl Econometrics*. 2000;15:447–70.
35. Basu A, Meltzer D. Value of information on preference heterogeneity and individualized care. *Med Decis Making*. 2009;27:112–27.
36. Thokala P, Duenas A. Multiple criteria decision analysis for health technology assessment. *Value Health*. 2012;15:1172–81.
37. Kjør T. A review of the discrete choice experiment - with emphasis on its application in health care. Syddansk Universitet (Health Economics Papers; Nr. 1). 2005. [http://findresearcher.sdu.dk/portal/da/publications/a-review-of-the-discrete-choice-experiment-with-emphasis-on-its-application-in-health-care\(07c6f140-e09c-11db-9628-000ea68e967b\).html](http://findresearcher.sdu.dk/portal/da/publications/a-review-of-the-discrete-choice-experiment-with-emphasis-on-its-application-in-health-care(07c6f140-e09c-11db-9628-000ea68e967b).html). Accessed 7 Nov 2016.
38. Hauber AB, González JM, Groothuis-Oudshoorn CG, Prior T, Marshall DA, Cunningham C, et al. Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR Conjoint Analysis Good Research Practices Task Force. *Value Health*. 2016;19:300–15.
39. Van Der Pol M, Currie G, Kromm S, Ryan M. Specification of the utility function in discrete choice experiments. *Value Health*. 2014;17:297–301.