A stochastic multicriteria model for evidence-based decision making in drug benefit-risk analysis

Tervonen, Tommi; van Valkenhoef, Gert; Buskens, Erik; Hillege, Hans L.; Postmus, Douwe

Published in:
Statistics in Medicine

DOI:
10.1002/sim.4194

Research Article

Statistics in Medicine

Received 15 February 2010 Accepted 20 December 2010 Published online 26 January 2011 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.4194

A stochastic multicriteria model for evidence-based decision making in drug benefit-risk analysis

Tommi Tervonen,[a*†] Gert van Valkenhoef,[b,c] Erik Buskens,[c] Hans L. Hillege[c,d] and Douwe Postmus[c]

Drug benefit-risk (BR) analysis is based on firm clinical evidence regarding various safety and efficacy outcomes. In this paper, we propose a new and more formal approach for constructing a supporting multi-criteria model that fully takes into account the evidence on efficacy and adverse drug reactions. Our approach is based on the stochastic multi-criteria acceptability analysis methodology, which allows us to compute the typical value judgments that support a decision, to quantify decision uncertainty, and to compute a comprehensive BR profile. We construct a multi-criteria model for the therapeutic group of second-generation antidepressants. We assess fluoxetine and venlafaxine together with placebo according to incidence of treatment response and three common adverse drug reactions by using data from a published study. Our model shows that there are clear trade-offs among the treatment alternatives. Copyright © 2011 John Wiley & Sons, Ltd.

Keywords: clinical pharmacology; decision analysis; benefit-risk analysis; stochastic multi-criteria acceptability analysis (SMAA)

1. Introduction

Drug benefit-risk (BR) analysis is a daily business for health-care professionals. Health authorities, prescribing physicians, pharmacists, reimbursement policy makers, and employees of insurance companies all more or less explicitly evaluate the safety and efficacy of different medicinal compounds. Although the exact scope of the analyses conducted by these evaluators is different (e.g. in clinical practice, the decision concerns an individual patient, whereas in policy making the general population or a subset of the population that has some particular characteristics is considered), they all must examine and weigh the clinical evidence regarding the magnitudes of benefit and risk, taking into account the quality and precision with which these magnitudes have been estimated.

The benefit/risk ratio (which is calculated from the difference in risk and difference in benefit between therapies) has been proposed as a simple aggregate measure of the BR trade-off for a single efficacy criterion and a single risk criterion. Although such a measure is easy to interpret and implement in clinical practice, drug BR analysis typically includes multiple benefit and risk criteria and consequently must include value judgments [1–3]. In such a setting, the use of multi-criteria decision analysis (MCDA) is more appropriate as it provides a framework for systematic and replicable analyses of complex decision problems involving value trade-offs.

The use of MCDA in the context of drug BR analysis was first proposed by Mussen *et al.* [4]. Their work includes a general framework for constructing a multi-criteria decision model for BR

[a] *Econometric Institute, Erasmus University Rotterdam, The Netherlands*
[b] *Faculty of Economics and Business, University of Groningen, The Netherlands*
[c] *Department of Epidemiology, University Medical Center Groningen, The Netherlands*
[d] *Department of Cardiology, University Medical Center Groningen, The Netherlands*
[*] *Correspondence to: Tommi Tervonen, Econometric Institute, Erasmus University Rotterdam, PO Box 1738, 3000DR Rotterdam, The Netherlands.*
[†] *E-mail: tervonen@ese.eur.nl*

assessment of new drugs by regulatory authorities. Although it is an important seminal work in the field, they score alternative drugs on the different benefit and risk criteria solely based on point estimates. Thus, uncertainty associated with sampling variation inherent to criteria measurements obtained in experimental or observational studies is ignored. In addition, the approach suggested by Mussen *et al.* [4] requires decision makers (DMs) to provide exact weights for describing the relative importance of the different criteria. Although detailed weight elicitation during model construction can help the DMs to understand the problem, in many real-life situations, DMs are not able to (or do not want to) give exact preference information. Also, a group of DMs may not reach a consensus about the weights [5]. Felli *et al.* [6] provided a similar application of MCDA in drug BR analysis. Instead of using continuous measurements, they proposed to use categorical value scales for all BR attributes included in the model. Although it makes the model easier to apply in different contexts, there is a substantial risk of losing information by mapping measurements from a continuous scale to ordinal categories.

To overcome the limitations of the two previous approaches, we propose to use stochastic multicriteria acceptability analysis (SMAA) [5, 7, 8] as a new and more elaborate approach to drug BR analysis. Our choice of the SMAA methodology is supported by its proven applicability in risk assessment [9, 10] and reported real-life analyses [11–17] alike. To demonstrate its applicability in drug BR analysis, we will apply the SMAA-2 method [8] to evaluate the potential benefits and risks of two commonly prescribed second-generation antidepressants in the setting of a published placebo-controlled trial [18].

2. SMAA

SMAA-2 [8] considers a discrete, multi-criteria decision problem consisting of a set of m alternatives that are evaluated in terms of n criteria. The vector of criteria values corresponding to alternative i is denoted by $\mathbf{C}^i=(C_1^i,\ldots,C_n^i)$, where C_k^i represents the performance of alternative i on criterion k. Instead of using deterministic values, the criteria values are assumed to be random variables with joint density function $f_{\mathbf{C}^i}(\mathbf{c}^i)$ in the evaluation space $X\subseteq R^n$.

It is assumed that the DM's preferences for any point $\mathbf{c}\in X$ can be represented by the real-valued value function $u(\mathbf{c},\cdot)$. Although SMAA-2 can be applied with any type of value function, it is generally assumed that the criteria satisfy the independence conditions [19] for applying the additive value function:

$$u(\mathbf{c},\mathbf{w})=w_1\cdot u_1(c_1)+\cdots+w_n\cdot u_n(c_n).$$

The additive value function is normalized by $u(\mathbf{c}',\mathbf{w})=0$ and $u(\mathbf{c}'',\mathbf{w})=1$ for arbitrarily chosen $\mathbf{c}',\mathbf{c}''\in X$, such that $(c''_{\{k\}},\mathbf{c}'_{\{\bar{k}\}})\succ(c'_{\{k\}},\mathbf{c}'_{\{\bar{k}\}})\ \forall k\in\{1,\ldots,n\}$. The symbol $\succ$ denotes the strict preference relation, and $(c_Y,c_{\bar{Y}})$ refers to the partition of c according to a subset Y of the criteria and its complement $\bar{Y}$. For example, if $n=5$ and $Y=\{1,3,5\}$, $c_Y=(c_1,c_3,c_5)$ and $c_{\bar{Y}}=(c_2,c_4)$. The weights w_k, normalized so that they sum to one, rescale the values of the partial value functions, normalized by $u_k(c'_k)=0$ and $u_k(c''_k)=1$, in such a way that a unit increase in the scaled function (i.e. the swing from c'_k to c''_k) indicates the importance of the criterion [20]. For example, $w_s>w_t$ implies that if the DM is currently at $\mathbf{c}'$ and could choose between moving to $(c''_{\{s\}},\mathbf{c}'_{\{\bar{s}\}})$ or $(c''_{\{t\}},\mathbf{c}'_{\{\bar{t}\}})$, he or she would rather move to $(c''_{\{s\}},\mathbf{c}'_{\{\bar{s}\}})$.

Instead of using the value function to rank the alternatives for an elicited weight vector $\mathbf{w}$, which is the traditional approach in multi-attribute value theory, the SMAA methodology has been developed for situations where the weights are random variables with a joint density function $f_{\mathbf{W}}(\mathbf{w})$ in the feasible weight space

$$\Omega=\left\{\mathbf{w}\in R^n:\mathbf{w}\geqslant 0 \text{ and } \sum_{j=1}^{n} w_j=1\right\}.$$

Total lack of preference information is represented by a uniform weight distribution in Ω, i.e. $f_{\mathbf{W}}(\mathbf{w})=1/\mathrm{vol}(\Omega)$. In practice, it may be possible to elicit some preference information from the DM, such as a partial or complete ranking of the criteria. This information can easily be incorporated into the model by restricting the feasible weight space accordingly [21].

Define $\Xi=(\mathbf{C}^1,\ldots,\mathbf{C}^m)$, and let $f_\Xi(\xi)$ denote the joint density function of Ξ. For given realizations ξ of Ξ and $\mathbf{w}$ of $\mathbf{W}$, the rank of each alternative is defined as an integer from the best rank ($=1$) to the worst rank ($=m$) by means of a ranking function

$$\mathrm{rank}(i,\xi,\mathbf{w})=1+\sum_{k=1}^{m}\rho(u(\mathbf{c}^k,\mathbf{w})>u(\mathbf{c}^i,\mathbf{w})),$$

where $\rho(\mathrm{true})=1$ and $\rho(\mathrm{false})=0$. SMAA-2 is then based on analyzing the stochastic sets of favorable rank weights

$$\Omega_i^r(\xi)=\{\mathbf{w}\in\Omega:\mathrm{rank}(i,\xi,\mathbf{w})=r\}.$$

Any weight $\mathbf{w}\in\Omega_i^r(\xi)$ results in such values for the different alternatives that alternative i obtains rank r.

The main decision-aiding measure in SMAA-2 is the *rank acceptability index*, denoted by b_i^r. It describes the share of all possible values of the weight vector $\mathbf{W}$ and the joint random vector Ξ for which alternative i is ranked at place r. Its value can be interpreted as the probability that alternative i is ranked at place r, where 0 indicates that the alternative will never obtain rank r and 1 indicates that alternative i will always obtain rank r. The rank acceptability index b_i^r is computed numerically as a multidimensional integral over the criteria distributions and the favorable rank weights as

$$b_i^r=\int_{\xi\in\Xi}f_\Xi(\xi)\int_{\mathbf{w}\in\Omega_i^r(\xi)}f_\mathbf{W}(\mathbf{w})\,d\mathbf{w}\,\mathrm{d}\xi.$$

The preferred (best) alternatives are those with high acceptabilities for the best ranks.

In addition to the rank acceptability indices, the SMAA methods allow to describe the typical preferences of a DM supporting each efficient alternative (i.e. all alternatives with a non-zero first rank acceptability index). These so-called *central weight vectors* can be presented to the DM to help him or her understand what kind of weights would favor a certain alternative, without providing factual preference information. The central weight vector of an alternative is defined as the expected center of gravity of all possible weight vectors that rank the alternative at the first place. It is computed numerically as a multidimensional integral over the criteria distributions and the favorable first rank weights using

$$\mathbf{w}_i^c=\int_{\xi\in\Xi}f_\Xi(\xi)\int_{\mathbf{w}\in\Omega_i^1(\xi)}f_\mathbf{W}(\mathbf{w})\mathbf{w}\,\mathrm{d}\mathbf{w}\,\mathrm{d}\xi/b_i^1.$$

The *confidence factor* p_i^c is the probability for an alternative to obtain the first rank when the central weight vector is chosen. The confidence factor is computed as a multidimensional integral over the criteria distributions using

$$p_i^c=\int_{\xi\in\Xi:\mathrm{rank}(i,\xi,\mathbf{w}_i^c)=1}f_\Xi(\xi)\,\mathrm{d}\xi.$$

Confidence factors can similarly be calculated for any given weight vector. The confidence factors indicate whether the criteria values are sufficiently accurate to discern the efficient alternatives. Alternatives with low first rank acceptability indices and low confidence factors for their central weight vectors are unlikely to be considered the most preferred one by any DM. In contrast, a very high confidence factor indicates that if a DM finds his or her preferences to correspond to an alternative's central weight vector, the alternative is almost certainly the one with highest preference [22]. Central weights of alternatives with low confidence factors (<0.50) should be interpreted with care, as even when a DM finds his central weight vector to correspond with his preferences, there might be other alternatives that achieve higher first rank acceptability with those weights.

If there is no preference information, the decision making is aided mainly through central weight vectors and confidence factors. When preference information is incorporated, the rank acceptability indices can be used to find the 'best' alternative and to quantify the risks related to uncertainties surrounding outcomes.

3. A multi-criteria model for the therapeutic group of antidepressants

To demonstrate the applicability of SMAA in drug BR analysis, we constructed a model for the therapeutic group of antidepressants using efficacy and safety data from a published study [18]. If patients are not harmed by deferral of therapy, it is important to have a non-active control included in the analysis to put the relative performances of the different active compounds into context with what is seen without a treatment [23]. For depressive disorder, there is no evidence that treatment delay or assignment to placebo results in permanent harm [24]. Placebo was therefore explicitly included as one of the alternatives in the constructed BR model.

3.1. Criteria

The original placebo-controlled trial compared efficacy and safety of venlafaxine and fluoxetine [18]. From this study, we selected treatment response, defined as an improvement from baseline of at least 50 per cent on the Hamilton Depression Rating Scale (HAM-D), as our benefit criterion. To obtain our risk criteria, we asked an expert in the field of antidepressants to select three adverse drug reactions (ADRs) that she considered to be most relevant from a drug safety perspective. The resulting criteria for evaluating the two drugs and placebo are summarized in Table I, and the data reported in the original study are shown in Table II. There is a certain overlap between efficacy and insomnia, because improved efficacy can lead to less insomnia. For the sake of simplicity, we disregarded this possible source of double-counting and assumed the criteria to be independent.

3.2. Probability distributions of the criteria values

The observed incidences r_k^i/n_k^i of treatment response and ADRs can be considered to be realizations from binomially distributed variables with success probability C_k^i (i.e. $r_k^i \sim \text{Bin}(n_k^i, C_k^i)$). Assuming independence of the $m \cdot n$ success probabilities, we modeled $C_k^i \sim \text{Beta}(a_k^i, b_k^i)$. Following a Bayesian approach with a flat Beta(1, 1) prior, the Beta parameters a_k^i and b_k^i were set equal to $r_k^i + 1$ and $n_k^i - r_k^i + 1$, respectively. The resulting parameter values are summarized in Table III.

Table I. Criteria, preference directions, and scaling vectors. All criteria are measured as incidences.

Name	Preference direction	c_k'	c_k''
Efficacy	↑	0.28	0.63
Nausea ADRs	↓	0.50	0.04
Insomnia ADRs	↓	0.31	0.08
Anxiety ADRs	↓	0.17	0.00

Table II. Incidence rates of HAM-D responders and three ADRs as reported in the original study [18], with their risk differences (RD) versus placebo (calculated by the authors based on the original data).

Criterion	Placebo	Fluoxetine	RD (95 per cent CI)	Venlafaxine	RD (95 per cent CI)
Efficacy	37/101	45/100	0.08 (−0.05, 0.22)	51/96	0.16 (0.03, 0.30)
Nausea ADRs	8/102	22/102	0.14 (0.04, 0.23)	40/100	0.32 (0.21, 0.43)
Insomnia ADRs	14/102	15/102	0.01 (−0.09, 0.11)	22/100	0.08 (−0.02, 0.19)
Anxiety ADRs	1/102	7/102	0.06 (0.01, 0.11)	10/100	0.09 (0.03, 0.15)

Table III. Beta distributions of the criteria values (parameters are given as a_k^i, b_k^i).

Criterion	Venlafaxine	Fluoxetine	Placebo
Efficacy	52, 46	46, 56	38, 65
Nausea ADRs	41, 61	23, 81	9, 95
Insomnia ADRs	23, 79	16, 88	15, 89
Anxiety ADRs	11, 91	8, 96	2, 102

Table IV. Rank acceptability indices from the analysis without preference information.

Drug	b_i^1	b_i^2	b_i^3
Venlafaxine	0.08	0.14	0.78
Fluoxetine	0.17	0.71	0.12
Placebo	0.75	0.16	0.09

3.3. Partial value functions

It is generally helpful to limit the region $Z \subseteq X$ over which preferences must be assessed to as small a region as possible, taking into account the observed ranges in the criteria values [19]. Our approach was therefore to bound Z by the interval hulls (the interval hull of k intervals is defined as the smallest possible interval that contains all these k intervals) of the 95 per cent probability intervals of the m success probabilities associated with each of the criteria. This ensures that even if the underlying independence assumptions are not valid for the complete range of theoretically achievable values in X, the additive value function will still be a good approximation for the subset Z in which the criteria values are most likely to fall. The two points c' and c'' required to scale the (partial) value function(s) were set equal to the least and most preferable values in Z, respectively, and are listed in Table I. The partial value functions $u_k(c_k)$ were assumed to be linear, meaning that they were defined as $u_k(c_k) = (c_k - c_k')/(c_k'' - c_k')$ if the preference direction is increasing and $u_k(c_k) = (c_k' - c_k)/(c_k' - c_k'')$ if the preference direction is decreasing. For example, for nausea $c_{\text{nau}}' = 0.50$, $c_{\text{nau}}'' = 0.04$, and the preference direction is decreasing, so $u_{\text{nau}}(c_{\text{nau}}) = (0.50 - c_{\text{nau}})/0.46$.

3.4. Preference information

We performed three analyses: one without preference information, and two scenarios with criteria rankings elicited from an expert in the field of antidepressants. For the scenario-based analyses, we considered a scenario of mild depression and a scenario of severe depression. For both scenarios, we asked the expert to identify the criterion that she considered to be most important, i.e. would foremost increase from the worst to the best value, given the range of the scales as depicted in Table I. Then we asked for the second one, etc. This process is similar to swing weighting in multi-attribute value theory [25]. However, since no exact weights are elicited, it requires less effort from the DM.

Let us denote by $\succ$ the strict preference relation for unit increases in the partial value functions of the criteria. The elicitation process resulted in the following ranking for mild depression: Nausea $\succ$ Anxiety $\succ$ Efficacy $\succ$ Insomnia. For severe depression the ranking was similar with the exception of efficacy being the most preferred criterion (i.e. Efficacy $\succ$ Nausea $\succ$ Anxiety $\succ$ Insomnia).

3.5. Analyses

The three analyses were conducted using the open source JSMAA software [26] v0.8 for Monte Carlo estimation of SMAA models. All analyses were executed with 10 000 Monte Carlo iterations, thereby giving the results sufficient accuracy (95 per cent confidence error margins of ± 0.01) [21].

The rank acceptability indices resulting from the analysis without preference information are listed in Table IV and visualized as a column chart in Figure 1. These indices show that each of the drugs is the preferred one given some preferences. Thus, all of them should be considered for further analysis. In a situation like this, the decision can be aided through the central weight vectors (see Table V and Figure 2). By looking at the central weights, we can see clear trade-offs among the three alternatives. For example, if the DM displays an *a priori* preference for venlafaxine, then based on the BR profiles expressed through the central weights, apparently efficacy has the highest relative importance. If the DM accepts the independence conditions underlying the additive model, he or she should find increasing efficacy from the worst scale value (0.28) to the best one (0.63) more important than improving any of the ADR criteria from their worst to best scale values.

By contrasting a DM's preferences for scale swings (Table I) with the central weights presented in Table V, the DM can quickly decide which drug is preferable in the current situation. For example, if the only preference information available is that the DM considers the scale swing of anxiety (0.17–0.00) less important than the scale swing of insomnia (0.31–0.08, see Table I), then he or she should prefer fluoxetine as it is the only alternative for which the central weight of anxiety is considerably lower than

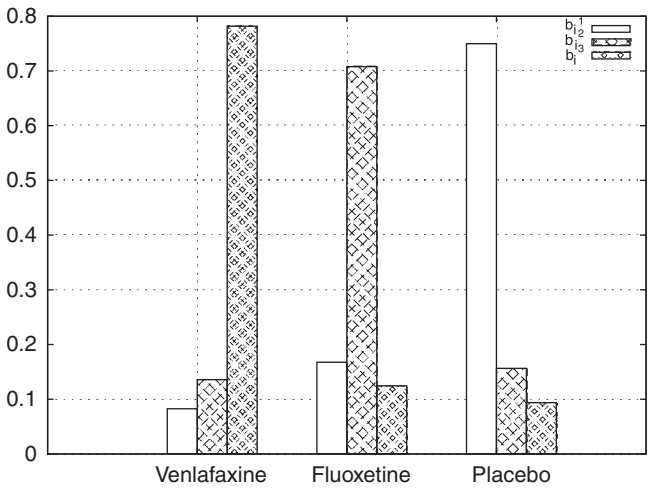

Figure 1. Rank acceptability indices for the model without preference information.

Table V. Central weights and corresponding confidence factors from the analysis without preference information.

		w_i^c			
Drug	p_i^c	Efficacy	Nausea	Insomnia	Anxiety
Venlafaxine	0.48	0.58	0.11	0.15	0.15
Fluoxetine	0.35	0.37	0.16	0.30	0.17
Placebo	0.96	0.18	0.28	0.25	0.29

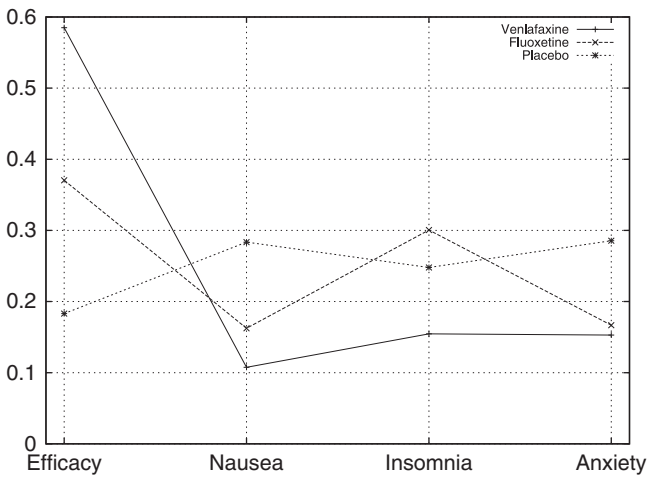

Figure 2. Central weight vectors for the model without preference information.

the central weight of insomnia. In addition, the confidence factors (Table V) quantify the risk associated with the decision. For example, if a DM finds fluoxetine's central weight vector to correspond with his or her preferences, the confidence factor (0.35) shows that the clinical data is too uncertain to make a truly informed decision.

Rank acceptability indices from the scenario of mild (severe) depression are presented in Table VI (Table VII) and illustrated in Figure 3 (Figure 4). Placebo obtains a very high first rank acceptability for the scenario of mild depression, and it obtains a reasonable rank profile for the scenario of severe depression. The rank profiles of fluoxetine and venlafaxine, in contrast, are very sensitive to the preferences as both of them obtain extremely low ($\leqslant$0.01) first rank acceptabilities for the scenario of mild depression, but reasonable ones (0.25 and 0.29) for the scenario of severe depression.

Table VI. Rank acceptability indices for the scenario of mild depression.

Drug	b_i^1	b_i^2	b_i^3
Venlafaxine	0.00	0.02	0.97
Fluoxetine	0.01	0.96	0.02
Placebo	0.99	0.01	0.00

Table VII. Rank acceptability indices for the scenario of severe depression.

Drug	b_i^1	b_i^2	b_i^3
Venlafaxine	0.25	0.25	0.50
Fluoxetine	0.29	0.48	0.23
Placebo	0.46	0.28	0.27

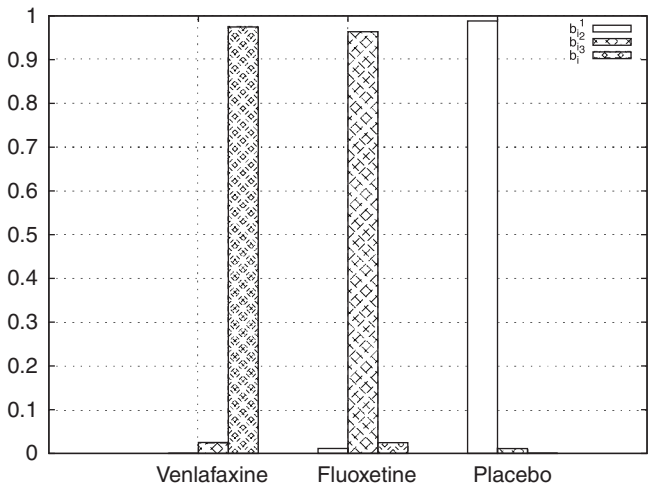

Figure 3. Rank acceptability indices from the scenario of mild depression.

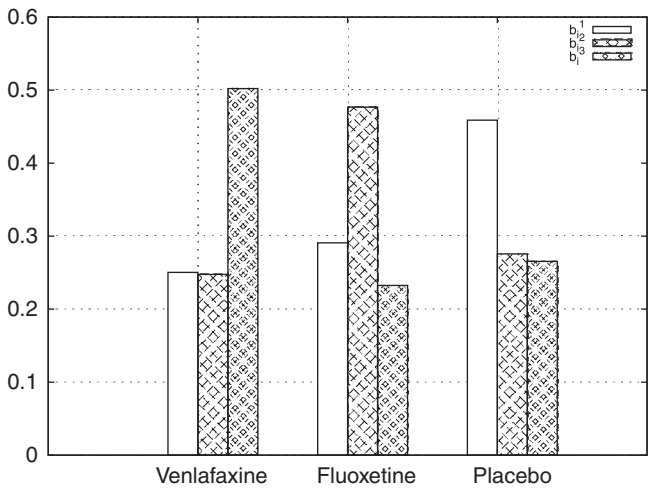

Figure 4. Rank acceptability indices from the scenario of severe depression.

4. Discussion

Drug BR analysis has multiple uses, ranging from regulatory decision making to supporting decisions of a practicing physician. The MCDA-based approach suggested in this paper can be adapted for most

contexts. We constructed a stochastic multi-attribute model for the therapeutic group of antidepressants by using data from a published placebo-controlled trial. Despite the fact that few of the differences among the three alternatives were significant from a frequentist perspective, our results show that there are clear trade-offs among the two active compounds and placebo when the uncertainty regarding the criteria measurements is taken into account. This can be seen from the central weight vectors of the analysis without preference information, and from the rank acceptability indices for the scenarios of mild and severe depression that differed only in the preference rank of the efficacy criterion relative to the risk criteria.

Compared to the MCDA-based approaches proposed by Mussen *et al.* [4] and Felli *et al.* [6], the use of SMAA has two main advantages. The first advantage of the SMAA methodology is the possibility to include the sampling variation that is inherent in criteria measurements that are based on clinical trials. Ignoring the uncertainty surrounding the criteria values, as is done by Mussen *et al.* [4] and Felli *et al.* [6], makes it difficult to assess how much the different drugs differ on the selected criteria. For example, a systematic review of 10 second-generation antidepressants [27] concluded that when looking at the point estimates and the corresponding 95 per cent confidence intervals, the drugs 'probably do not differ substantially for the treatment of major depressive disorder' and that choosing the most appropriate treatment is therefore difficult. However, a more recent review [28] was able to provide more concrete results through a network meta-analysis, a Bayesian approach to evidence synthesis that fully takes into account the uncertainty in the effect estimates. In addition to performing all possible comparisons, the authors provided rank probability plots that clearly showed that some drugs are 'better' than others on specific criteria. Unfortunately, rank probability plots for individual criteria provide little guidance when more than two criteria are considered. As our results have shown, applying the SMAA method enables one to clearly assess the existing trade-offs. In addition, the ability of our approach to propagate uncertainty to the results (in terms of rank acceptability indices and confidence factors) allows one to quantify the risks associated with any decision that is based on the results of the BR analysis.

The second advantage of our approach over the two existing ones is the possibility to characterize typical trade-offs supporting a drug BR profile without knowing or eliciting the (exact numerical) preferences beforehand. The possibility to use our model without any preferences as well as with scenario-based ordinal preferences lowers the effort required to apply the model in different situations, and also increases the transparency of the decision-making process. An analysis without preference information is useful when it is not feasible to elicit preferences or when the potential merits of a drug have to be assessed across a wide range of preferences. This latter situation occurs, for example, in policy decision making, where the policy maker's decision affects the complete target population. The central weight vectors could then be used to see whether there are likely scenarios (in terms of criteria weights) that will lead to the selection of a certain drug. The selection scenario could be, for example, a prescription decision, and the actual decision being aided is whether the drug should be granted a marketing authorization. The scenario-based rank acceptability indices can be used in operational support of decisions depending on drug BR analysis. For example, if the BR analysis is used for aiding a prescription decision for a patient with severe depression, our results show that both venlafaxine and fluoxetine are viable choices because of their relatively high acceptabilities for the best ranks. Also, if due to external factors (local reimbursement policy, patient profile including allergies, etc) a drug with a low first rank acceptability is prescribed (such as either of the active compounds in case of mild depression), the prescriber should be sensitive to changes in the external environment as drugs with 'better' BR profiles may have become available.

Although our example of the second-generation antidepressants clearly demonstrated the usefulness of the proposed approach, in some decision making contexts, other approaches might be more appropriate. When the DMs have the time and motivation to engage in decision conferencing [29], a traditional multi-attribute value/utility theory approach can be more suitable. However, although such a facilitated environment might help the DMs to explore the problem in more detail, it can also introduce additional bias as the preference elicitation is heavily guided by the facilitator. In any case, we acknowledge that social aspects play an important role in group decision making, and future research should explore the applicability of our model in real-life pharmaceutical group decision-making contexts, such as policy decision making.

Instead of having a different model for each therapeutic group, one could also consider constructing a more generic model by using the dimensions of an existing utility instrument, such as the EQ-5D or the Health Utilities Index. Although such instruments are suitable for calculating QALYs in the context of cost-effectiveness analysis, there is an important drawback when using them for drug BR

analysis: their dimensions are defined in terms of generic health attributes, such as physical functioning, social functioning, and vitality, and may therefore not be very sensitive and responsive to the disease of interest. So, although our results have shown that there are clear trade-offs among the considered alternatives, the relative differences in safety and efficacy may not be large enough to significantly change a patient's health status when this is measured in terms of generic health attributes.

The results from our example should be interpreted with care for three reasons. First, ideally evidence from all existing studies should be taken into account, rather than just a single trial. Future research should therefore consider our model together with evidence synthesis methods. As discussed previously, an appropriate method in such cases would be network meta-analysis (also known as the Mixed Treatment Comparisons model) [30, 31] as it allows to take into account all evidence simultaneously. If a full network meta-analysis were performed, the random samples from the full joint posterior distribution of the effect estimates could be fed directly into the BR model. In this case, however, the possible inconsistencies in the network of trials would have to be evaluated, which brings additional level of complexity to the model. Second, the model is relevant only with respect to the data within the trial. For decisions depending on comprehensive BR profiles (e.g. drug marketing authorization decision), it can serve only as a starting point for further discussion as there can be additional qualitative information that is not included in the model. For example, our model excludes drug–drug interactions that might differ among the alternatives. Finally, the preferences were elicited from a single expert, and might not represent consensus among a larger group of experts.

To conclude, we presented a new MCDA-based approach to drug BR analysis with an example application to the therapeutic group of second-generation antidepressants. In contrast to previous models, our model is based on the SMAA methodology, which allows us to take into account the sampling variation that is inherent in criteria measurements that are based on clinical trials and/or observational studies. In addition, by making the trade-offs among the analyzed drugs explicit, we separated clinical data from subjective judgments, thereby increasing the transparency of the decision-making process. Finally, the constructed model is specific to the therapeutic group of antidepressants. It would appear that the underlying concepts are general, but future research should assess the applicability of the SMAA methodology to other therapeutic groups.

Acknowledgements

This study was partly supported by the Escher project (T6-202), which is a project of the Dutch Top Institute Pharma. We thank Barbara van Zwieten for her expert opinion and information about antidepressants. In addition, we acknowledge Kit Roes, Hein Fennema, Hans van Leeuwen, and Marcel Hekking of Schering-Plough, as well as the anonymous referees, for their valuable remarks and suggestions. The authors declare no conflicts of interest.

References

1. McGregor M, Caro JJ. QALYs: are they helpful to decision makers? *Pharmacoeconomics* 2006; **24**(10):947–952.
2. Claycamp HG. Rapid benefit-risk assessment: no escape from expert judgments in risk management. *Risk Analysis* 2006; **26**(1):147–156.
3. Garrison Jr LP, Towse A, Bresnahan BW. Assessing a structured, quantitative health outcomes approach to drug risk-benefit analysis. *Health Affairs* 2007; **26**(3):684–695.
4. Mussen F, Salek S, Walker S. A quantitative approach to benefit-risk assessment of medicines—part 1: the development of a new model using multi-criteria decision analysis. *Pharmacoepidemiology and Drug Safety* 2007; **16**(Suppl. I): S12–S15.
5. Tervonen T, Figueira JR. A survey on stochastic multicriteria acceptability analysis methods. *Journal of Multi-Criteria Decision Analysis* 2008; **15**(1–2):1–14. DOI: 10.1002/mcda.407.
6. Felli JC, Noel RA, Cavazzoni PA. A multiattribute model for evaluating the benefit-risk profiles of treatment alternatives. *Medical Decision Making* 2009; **29**(1):104–115.
7. Lahdelma R, Hokkanen J, Salminen P. SMAA—stochastic multiobjective acceptability analysis. *European Journal of Operational Research* 1998; **106**(1):137–143. DOI: 10.1016/S0377-2217(97)00163-X.
8. Lahdelma R, Salminen P. SMAA-2: stochastic multicriteria acceptability analysis for group decision making. *Operations Research* 2001; **49**(3):444–454. DOI: 10.1287/opre.49.3.444.11220.
9. Tervonen T, Linkov I, Steevens J, Chappell M, Figueira JR, Merad M. Risk-based classification system of nanomaterials. *Journal of Nanoparticle Research* 2009; **11**(4):757–766. DOI: 10.1007/s11051-008-9546-1.
10. Tervonen T, Figueira JR, Lahdelma R, Almeida Dias J, Salminen P. A stochastic method for robustness analysis in sorting problems. *European Journal of Operational Research* 2009; **192**(1):236–242. DOI: 10.1016/j.ejor.2007.09.008.
11. Hokkanen J, Lahdelma R, Salminen P. A multiple criteria decision model for analyzing and choosing among different development patterns for the Helsinki cargo harbor. *Socio-Economic Planning Sciences* 1999; **33**:1–23.

12. Tervonen T, Hakonen H, Lahdelma R. Elevator planning with stochastic multicriteria acceptability analysis. *Omega* 2008; **36**(3):352–362. DOI: 10.1016/j.omega.2006.04.017.
13. Kangas A, Kangas J, Lahdelma R, Salminen P. Using SMAA-2 method with dependent uncertainties for strategic forest planning. *Forest Policy and Economics* 2006; **9**:113–125. DOI: 10.1016/j.forpol.2005.03.012.
14. Kangas J, Hokkanen J, Kangas A, Lahdelma R, Salminen P. Applying stochastic multicriteria acceptability analysis to forest ecosystem management with both cardinal and ordinal criteria. *Forest Science* 2003; **49**(6):928–937.
15. Hokkanen J, Lahdelma R, Miettinen K, Salminen P. Determining the implementation order of a general plan by using a multicriteria method. *Journal of Multi-Criteria Decision Analysis* 1998; **7**(5):273–284. DOI: 10.1002/(SICI)1099-1360(199809)7:5<273::AID-MCDA198>3.0.CO;2-1.
16. Kangas J, Kangas A. Multicriteria approval and SMAA-O method in natural resources decision analysis with both ordinal and cardinal criteria. *Journal of Multi-Criteria Decision Analysis* 2003; **12**(1):3–15. DOI: 10.1002/mcda.344.
17. Lahdelma R, Salminen P, Hokkanen J. Locating a waste treatment facility by using stochastic multicriteria acceptability analysis with ordinal criteria. *European Journal of Operational Research* 2002; **142**(2):345–356. DOI: 10.1016/S0377-2217(01)00303-4.
18. Nemeroff CB, Thase ME. A double-blind, placebo-controlled comparison of venlafaxine and fluoxetine treatment in depressed outpatients. *Journal of Psychiatric Research* 2007; **41**:351–359.
19. Keeney R, Raiffa H. *Decisions with Multiple Objectives*: *Preferences and Value Tradeoffs*. Wiley: New York, 1976.
20. von Nitzch R, Weber M. The effect of attribute ranges on weights in multiattribute utility measurement. *Management Science* 1993; **39**:937–943.
21. Tervonen T, Lahdelma R. Implementing stochastic multicriteria acceptability analysis. *European Journal of Operational Research* 2007; **178**(2):500–513. DOI: 10.1016/j.ejor.2005.12.037.
22. Lahdelma R, Salminen P. Classifying efficient alternatives in SMAA using cross confidence factors. *European Journal of Operational Research* 2006; **170**(1):228–240. DOI: 10.1016/j.ejor.2004.07.039.
23. Temple R, Ellenberg S. Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 1: ethical and scientific issues. *Annals of Internal Medicine* 2000; **133**(6):455.
24. Walsh BT, Seidman SN, Sysko R, Gould M. Placebo response in studies of major depression. *Journal of the American Medical Association* 2002; **287**(14):1840–1847. DOI: 10.1001/jama.287.14.1840.
25. Belton V, Stewart TJ. *Multiple Criteria Decision Analysis—An Integrated Approach*. Kluwer Academic Publishers: Dordrecht, 2002.
26. Tervonen T. JSMAA an open source software for SMAA computations. In *Proceedings of the 25th Mini-EURO Conference on Uncertainty and Robustness in Planning and Decision Making* (*URPDM2010*), Coimbra, Portugal, Henggeler Antunes C, Rios Insua D, Dias L (eds). 2010.
27. Hansen RA, Gartlehner G, Lohr KN, Gaynes BN, Carey T. Efficacy and safety of second-generation antidepressants in the treatment of major depressive disorder. *Annals of Internal Medicine* 2005; **143**(6):415–426.
28. Cipriani A, Furukawa TA, Salanti G, Geddes JR, Higgins JP, Churchill R, Watanabe N, Nakagawa A, Omori IM, McGuire H, Tansella M, Barbui C. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. *The Lancet* 2009; **373**(9665):746–758. DOI: 10.1016/S0140-6736(09)60046-5.
29. Phillips LD. Decision conferencing. In *Advances in Decision Analysis*: *From Foundations to Applications*, Edwards W, Miles Jr R, von Winterfeldt D (eds). Cambridge University Press: Cambridge, 2007.
30. Salanti G, Higgins JPT, Ades AE, Ioannidis JPA. Evaluation of networks of randomized trials. *Statistical Methods in Medical Research* 2008; **17**(3):279–301. DOI: 10.1177/0962280207080643.
31. Sutton AJ, Higgins JPT. Recent developments in meta-analysis. *Statistics in Medicine* 2008; **27**(5):625–650. DOI: 10.1002/sim.2934.