INTEGRATING PATIENT PREFERENCES AND CLINICAL TRIAL DATA IN AN MCDA MODEL FOR QUANTITATIVE BENEFIT-RISK ASSESSMENT

Master thesis Henk Broekhuizen
Supervisors: Maarten J IJzerman, PhD, Karin G M Groothuis, PhD, A Brett Hauber, PhD

Introduction

Regulators at governmental agencies (the European Medicines Agency (EMA) in Europe and the Food and Drug Authority (FDA) in the United States) make decisions regarding the acceptability of pharmaceuticals. The first of these decisions called benefit-risk assessments (BRAs) is when it is decided whether or not the pharmaceutical is authorized to enter a particular market. For this, pharmaceutical companies must submit a value dossier to the agency. This dossier contains information on the characteristics of the pharmaceutical, as well as information about the target patient group. Furthermore, the dossier contains evidence about the benefits and risks of the pharmaceutical derived from (pre-)clinical studies. Based on this value dossier, the agency evaluates the benefit-risk profile of the pharmaceutical. The EMA defines benefits as the pharmaceutical's beneficial effects for the target population, and risks as the pharmaceutical's detrimental effects on the patient's health [1]. The EMA definition will be used for the remainder of this paper. After the evaluation of the benefit-risk profile, the agency decides whether or not the current evidence justifies a marketing authorisation, e.g. if the benefits outweigh the risks. The agency can also decide further evidence is required or that the pharmaceutical company should provide clarification on certain points in the value dossier.

Pharmacovigilance is the second type of decision where a BRA is done, and entails (periodic) reviews of new evidence. The evidence for this can be collected by evaluating new clinical studies that were done with the pharmaceutical and through the collection of spontaneous reports of adverse events, such as those reported through the European EUDRAVigilance[1] system. Pharmacovigilance is required because not all evidence about the benefit-risk profile of a pharmaceutical can be known when it is authorised. Examples of this are long-term risks that were not measurable in short-run pre-marketing studies that were included in the value dossier, or risks with a small incidence that were not measured because of a smaller sample size of the pre-marketing studies. With the new evidence, regulators can decide to do nothing if there is no significant change in the benefit-risk profile, withdraw the pharmaceutical if the benefits no longer outweigh the risks, or require a re-labelling when the benefit-risk profile changes for a particular group of patients.

Currently, most BRAs are done using qualitative (if any) models and informal meetings [2–6]. This obfuscates the influence of patient preferences on the eventual decision being made. The EMA is looking to improve this by increasing transparency and communicability to patients with pharmacovigilance legislation introduced in July 2012 [7–9]. The legislation requires more information about risks from pharmaceutical companies, improves methods for the collection of evidence, introduces a new procedure for urgent pharmaceutical risk situations and improves communication to stakeholders [9]. The stakeholders here are the European public and more specifically, patients. The EMA plans to improve communication to them through three means. The first is the publication of assessments,

---

[1] Acronym stands for: *European Union Pharmaceutical Regulating Authorities Pharmacovigilance*

approvals, recommendations, opinions and decisions from its committees. The second is the coordination of safety signals to the public via the EMA website. The third is the introduction of public hearings for urgent pharmaceutical risk situations. These improvements mainly serve to explain the decisions of the agency to the public, but do not give patient preferences a more explicit place in BRAs. It is still difficult to extract from the (now public) reports the tradeoffs that were made in order to come to the decision and how much influence patient preferences have had in that decision.

This is a problem for a number of reasons. First, patients are the ones experiencing the benefits and risks so their preferences should taken into account if the decisions are to be patient-centred [10]. Second, patient preferences can differ from physician and regulator preferences, so decisions from just the regulator's perspective may not necessarily represent the patient's best interest. This difference in preferences has been observed in osteoporosis, Crohn's disease and depression patients [11–13]. Third, when patient influence is unclear, patients may feel less involved in the decisions and take less responsibility for their own health [14]. This is especially a problem with chronic disease patients who's lifestyle has a large impact on their prognosis [13], [14].

The EMA organizes patient panels and 'professional' patients take part in BRA to allow for the patient's perspective, but it is unlikely that the patients who participate in the patient panels or the 'professional' patients are fully representative of a particular group of patients [14]. Also, they can not show the heterogeneity that can exist in patient groups [15–17]. Because of these reasons, other methods must be used to gather information about the preferences of that patient group. Full cooperation of a large group of patients in a BRA would require large time and money investments of the agency and of patients and as such is not practical. Another method to gather preference information from a large group of patients is by performing preference studies. The outcome of a preference study most often takes the form of numerical values indicating the importance of benefits and risks, with their associated statistical characteristics. There are two types of preference studies. The first type (revealed preference study) retrospectively looks at real life decisions that were made in order to measure patient preferences. The second type (stated preference study) presents patients with hypothetical questions from which preference can be estimated. There are different ways these studies can take shape. Patient panels can be organized where the study is performed, or questionnaires can be sent out to patients, enabling a larger possible sample size. The first would be easier to implement into the current EMA procedures, but the second better represents the target patient population because of the larger possible sample size.

Thus, with preference studies the preferences of larger groups of patients can be elicited. This can be used to give regulators in a BRA more information on which to base their decision. However, this still does not address the issue of obfuscated patient influence, e.g. how much weight does the patient perspective (or preference) have in the BRA? Methods are needed to make this influence explicit. One method of doing so is by using a quantitative model. Quantitative models require the estimation and explicitation of certain aspects of the BRA like the importance of risks and benefits, and the performance characteristics of pharmaceuticals like symptom reduction. This enables regulators to make their tradeoffs with regard to the importance of certain aspects (like the patient's preference) of the problem explicit [1], [5], [18], [19]. Also, it can help them structure the large amount of information that is required in a BRA and check the consistency of their decision [1], [6], [18], [19]. There is still a discussion

about which place (if any) quantitative models should have in BRA [1], [2], [19]. There is a wide range of involvement of quantitative models. They can have a supplementary role; this means that the output from the quantitative model is a piece of data to be considered by the regulator before making a decision [13]. They can also serve a mainly structuring role, were all available evidence is structured in the quantitative model and presented to the regulators so they can make a better decision [20]. Finally, they can have decision supporting role, where all available data is integrated in the quantitative model and regulators define decision rules which guide their decision.

We suggest integrating patient preferences and clinical trial data in a quantitative model for the following reasons. First of all, it makes all sources of data and assumptions about data that are used in the BRA explicit, which increases the BRA's transparency. Secondly, transparency is increased because tradeoffs that regulators previously made implicitly can now be made explicit. Third, it can increase communicability of the BRA because all data can be presented from within one comprehensive framework. Fourth, it is ensured that all data (preference and performance) are taken into account. This is especially important in BRAs where there is a large amount of data because of bounded rationality [21], [22]. Bounded rationality means that regulators under cognitive stress (such as when they are confronted with a large amount of information) tend to focus on only a few important aspects of the BRA. This can mean that the details of a BRA can be overlooked. When there are a large number of details this can lead to suboptimal decisions. Fifth, the use of clinical trial data instead of subjective expert opinion can make decisions more grounded in clinical evidence.

With this suggestion the question rises which of the quantitative models is suited to integrate patient preferences and clinical trial data. Guo *et al*. have reviewed twelve quantitative models that can be used to assist decision making in BRA [3]. The EMA has done a similar review of current BRA models [1]. The quantitative models differ in their approach to BRA. They differ in what kind of data they incorporate (preference, performance, or both), if they allow for tradeoffs to be made, if they can be used with multiple data sources and whether they can take uncertainty into account. They also differ with regard to their outcome measures (most commonly graphs or a numeric value with statistical properties). Guo *et al*. note that, based on academic activities at the EMA, incremental net health benefit [23] (INHB), multi criteria decision analysis [24] (MCDA) and stated preference methods [17], [25] (SPM) currently receive the most attention [3]. The EMA report considers MCDA, Bayesian statistics [26] and decision trees [27] to be the most comprehensive models [1]. The report notes that qualitative models will remain important, mainly for problem structuring [1].

With these considerations, we think MCDA to be a well suited model to integrate patient preferences and clinical trial data for the following reasons. First of all, it can incorporate both preference and performance information. Secondly, it can help make and explicate tradeoffs between the multiple criteria that are present in BRA through the process of weighting. Third, it can take stochastic uncertainty into account. Stochastic uncertainty is uncertainty surrounding preference or performance information. It can reduce the internal and external validity of BRAs if only averages and not the uncertainty (i.e. variation) around these means are explicitly taken into account.

MCDA has been used for BRA in numerous fields beside the pharmaceutical field [18], [28–30]. Mussen *et al.* developed an MCDA model that can be used for pharmaceutical BRA [24]. In their model, regulators assign weights to criteria and use their expert opinion to rate the performances of pharmaceuticals on the criteria. It has the advantages of MCDA mentioned above but it is deterministic and uncertainty is not made explicit. The model performs sensitivity analysis, but its variation is chosen by the modeller and not derived from scientific data. This can give a false sense of accuracy and decision robustness [31]. Another MCDA model named stochastic multicriteria acceptability analysis (SMAA) was adapted for BRA and can take performance uncertainty into account through probabilistic simulation [32–34] and as such improves on the disadvantage of Mussen's model. However, this model is limited with regard to the possibility to integrate uncertain preferences. This is because the model is based on inverse weight space analysis. Inverse weight space analysis was developed for decisions where decision makers were reluctant or unable to indicate their preferences, and as such weights are assumed to have a uniform distribution which conveys no additional information to the model.

So, there is currently no MCDA model that can integrate patient preferences and clinical trial data, and that can handle stochastic uncertainty surrounding both types of data. The present study will therefore make the first steps in developing and testing such an MCDA model. The rest of this article is organised as follows: after explaining MCDA, we describe the model, what its outcome measures are and what its required inputs are. After discussing the model, it will be illustrated by applying it to a case on anti-depressants and a case on renal cancer medication. We will then discuss the results and indicate possibilities for further research.
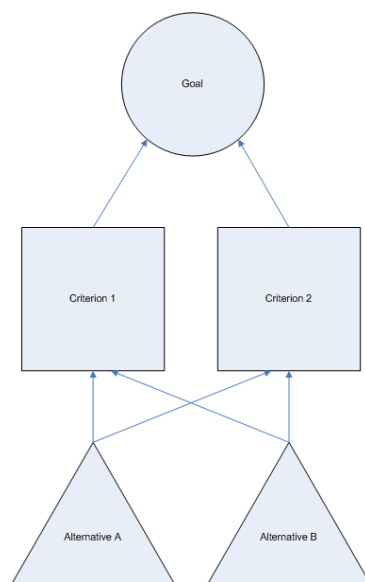
Methods

Figure 1: an MCDA model. The circle is the MCDA's goal, the squares are criteria and the triangles are alternatives.

MCDA
Figure 1 shows an MCDA structure. In an MCDA, decision makers choose between several *alternatives* in order to achieve a certain *goal*. For this, *criteria* are defined that contribute to that goal. For example, in a decision where we want to make a diagnosis, different diagnostic tests can be alternatives, and specificity and sensitivity can be criteria. Not all criteria are equally important, so they are given *weights* that indicate their importance; higher weights indicate a higher importance. Alternatives can be scored on the criteria; this is called their *performance*. For example, the performance of a certain diagnostic test on the sensitivity criterion can be 85%. To choose which of the alternatives is best suited to achieve the goal, the alternative scores on the individual criteria have to be aggregated. This is done via a value function [35]. The most commonly used value function is the additive value function, where weighted scores on criteria are added up. The weighted score is the performance of a particular alternative on a particular criterion multiplied by that criterion's weight. For example

when looking at figure 1, alternative A's total score according to a additive value function is its score on criterion 1 multiplied by criterion 1's weight, plus its score on criterion 2 multiplied by criterion 2's weight. When all alternatives are scored, the alternative with the best score is chosen. Most often this is the alternative with the highest score, but in some decisions like cost minization decisions, the alternative with the lowest score will be chosen. Our model is an MCDA model and as such shares characteristics with the MCDA model as introduced above. There are however differences which will now be discussed. For this, we will first define the model in more detail and then address the inputs to the model.
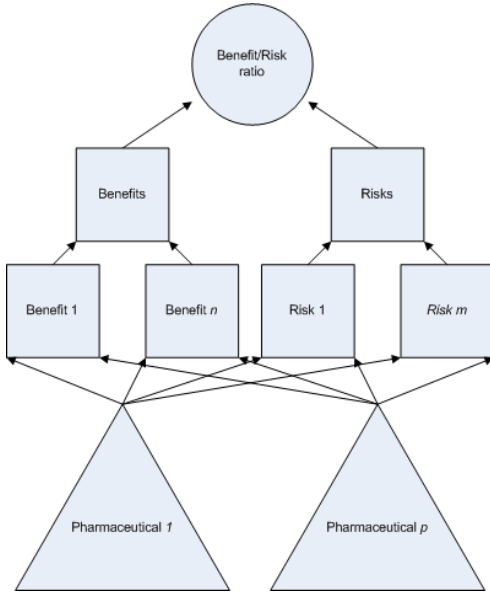

Figure 2: the structure of our MCDA model. The circle is the MCDA's goal, the squares are criteria and the triangles are alternatives.

The model

Like in the above description, our model also deals with alternatives and criteria. Our MCDA structure can be seen in figure 2. Since the model is for pharmaceutical BRAs, alternatives are most often pharmaceuticals. Criteria are divided into two categories: benefit criteria and risk criteria. In our model we consider n benefit criteria, m risk criteria and p alternatives (pharmaceuticals). The weight of a particular benefit or risk criterion k is denoted by w_{bk} and w_{rk}, respectively. The performance of pharmaceutical i on a particular criterion k is denoted with b_{ik} or r_{ik}, where the b and r denote whether k is a benefit or a risk criterion. The additive function for the weighted benefits for a particular pharmaceutical i is described by the follow formula:

$$Benefits_i = \sum_{k=1}^{n} w_{bk} \cdot b_{ik} \tag{1}$$

The additive function for the weighted risks for that pharmaceutical is:

$$Risks_i = \sum_{k=1}^{m} w_{rk} \cdot r_{ik} \tag{2}$$

And the benefit-risk ratio for that pharmaceutical is calculated by:

$$BRR_i = \frac{Benefits_i}{Risks_i} \tag{3}$$

In our model we want to make uncertainty and its influence explicit. Thus, the variables w, b and r in equations 1, 2 and 3 become stochastic variables which follow a certain stochastic distribution. Equations 1, 2 and 3 are therefore joint distributions. The calculation of these joint distributions is

difficult to solve analytically mainly because of the ratio of distributions in equation 3, but the complexity of the calculation also increases with large n, m and p. Furthermore, these analytical calculations are time-consuming and not all analytical relations between distributions are known. Because of these problems, we will not try to analytically solve the problem, but use an established method for estimating joint distributions: the use of probabilistic (Monte Carlo) simulations [36]. This means that we draw numbers a large number of times from the simple distributions of w, b and r that make up the joint distribution and then insert those in equations 1,2 and 3 to estimate the joint distributions.

To do this, the stochastic distributions of w, b and r have to be approximated. There are various types of approximations, and what approximation is useful depends on the characteristics of the data. Approximation means trying to find a standard distribution that resembles the data and adjust its parameters so that it fits the data. This means that random draws from an approximated distribution should fit the known data. There are numerous kinds of standard distributions and approximation methods. Not all distributions and approximation methods will be discussed here. For this, the reader is referred to textbooks on statistical modelling [37], [38]. Two often used approximation methods will be mentioned here: using a normal distribution and using the bootstrap method. Odds ratios are an often used outcome measure in clinical trials and systematic reviews. They follow a normal distribution in the log domain [39]. The normal distribution has two parameters it needs, and that have to be estimated from the data: a mean μ and a variation σ^2. Since odds ratios only follows a normal distribution in the log domain, performance data first have to be transformed with a logarithm before they can be used in the simulation. They then have to be transformed back after simulation to be comprehensible. The bootstrap method is a nonparametric resampling method that is most often used to approximate samples with a small sample size [40]. It is nonparametric, meaning that no parameters are assumed beforehand to fully described it. Resampling means that in order to make the distribution, the original data is resampled with Monte Carlo simulations.

Simulation

When we have approximated the distributions of w, b and r, we can run simulations to estimate the joint distributions that are described by equations 1,2 and 3 for each pharmaceutical. Our simulations were programmed and run in the statistical programming language R [41]. The simulations give us weighted benefits and weighted risks for each pharmaceutical for each simulation run. These approximate the joint distributions of equations 1, 2 and 3.

The risk-benefit plane

These summed weighted benefits and summed weighted risks are plotted per simulation run in a risk-benefit plane (figure 3). The risk-benefit plane is a plane where the x axis denotes weighted benefits, and where the y axis denotes weighted risks. The risk-benefit plane is similar to the incremental cost-effectiveness plane used in cost-effectiveness analysis, which was translated to the BRA context by Lynd et al. in 2004 [34]. They plotted simulation outputs of their discrete event model in an incremental risk-benefit plane, where the unit on the x axis was the change in probability of a particular benefit criterion and the unit on the y axis was the change in probability of a particular risk criterion. They assessed the influence of performance uncertainty by determining probabilities of certain events and then using these

probabilities to estimate patient outcomes of their discrete event simulation model. In 2010, they introduced quality adjust life years (QALY) based weighting factors to use for the data points [23], [42]. Although it possible to convert the units in the present model to QALY's based on quality of life studies, we think ours is a more versatile approach. The conversion from clinical endpoints to a QALY based unit introduces more structural uncertainty in the model (due to the extra translation step from clinical data to QALY's). Also the notion of a QALY may be difficult to understand for patients; this makes the model and thus the BRA less transparent and communicable.
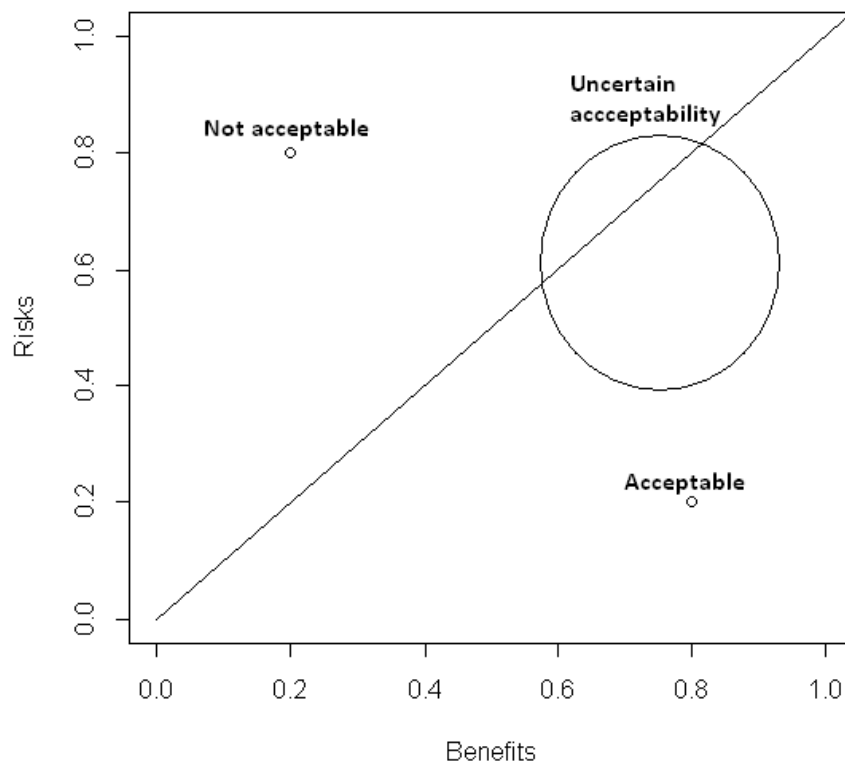

Figure 3: Risk-benefit plane with $\mu = 1$ line. In this plane the acceptability of multiple pharmaceuticals can be plotted, but comparisons between them are not done explicitly (this is done in figure 4). All points where weighted benefits > weighted risks are considered acceptable. Points below the line have weighted benefits > weighted risks. Because of uncertainty, the simulation outputs of a particular pharmaceutical are not concentrated in one point but are spread out over an area, such as the circle shown in the figure. We here show a circle as an example; with real data this takes the shape of a point cloud. In this case a probability of acceptability can be calculated by comparing the percentage of points under the line with the percentage of points above the line.

Pharmaceutical acceptability

With the joint distributions plotted in the risk-benefit plane, we can say something about a pharmaceutical's acceptability. This is interesting because this is the goal of a BRA: to decide if a pharmaceutical can be accepted (i.e. given market authorization). A regulator considers a pharmaceutical to be acceptable when its benefits outweigh its risks. When this is the case is a subjective decision made by the regulator. This decision can be approximately visualized with a line with a constant slope in the risk-benefit plane. On this line, the benefit-risk ratio is constant. The line is called the risk-benefit acceptability threshold, denoted by μ [34]. All points below the line are acceptable and points above the line are not acceptable. This is because points below the line have a higher benefit-risk ratio and points above the line have a lower benefit-risk ratio than the minimum required by the regulator. Due to the Monte Carlo simulations our model does not have one point in the risk-benefit plane per pharmaceutical, but a point cloud per pharmaceutical. Therefore, not all simulation points of a pharmaceutical may be under or above the μ line. In that case, the probability of acceptability can be calculated: the percentage of simulation points under the μ line is the probability that a particular pharmaceutical is acceptable given the current uncertainties surrounding preference and performance information. The question is, however, what level of μ is appropriate?

In the 2004 Lynd *et al.* model, μ is varied to indicate different attitudes toward what constitutes an acceptable risk for a particular amount of benefit [34]. Higher μ's indicate a higher tolerance for risk. A disadvantage of this approach is that that μ is unknown [34]. A second disadvantage of this approach is that it is difficult to account for magnitude differences in benefits and risks. Some benefits are perceived by patients to be so great that they are willing to accept high risks in order to acquire these benefits, while some risks may be so grave that only very large benefits will be able to convince patients to accept the risk. This makes it difficult to assess multiple criteria at once. We propose therefore that the units on the x and y axis are weighted odds ratios with weighting derived from elicited preference studies, so they can take patient risk attitudes into account (for example, risk averse patient tend to give risk criteria high weights). Because weighting is already done in the axis, we can set μ at 1. This means that for all points below the line the weighted benefits are greater than the weighted risks. In this way, difficulties with multiple criteria and establishing μ are avoided.
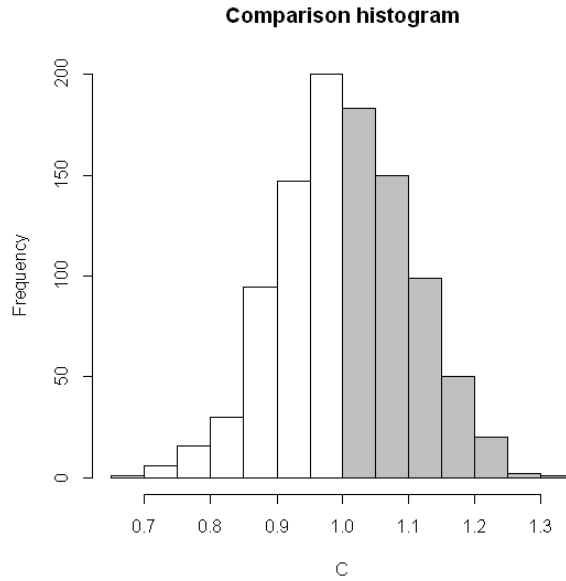
Comparison histogram

Figure 4: Comparison between two hypothetical pharmaceuticals, pharmaceutical A and pharmaceutical B. Shown here is a histogram of 1000 simulations of a possible distribution of the comparator unit C, which is defined in equation 4 to be a ratio between the benefit-risk ratios of A and B. The shaded bars are where C>1. This is when A outperforms B. The ratio of the shaded area versus the non-shaded area is the probability that A outperforms B.

Comparison between pharmaceuticals

With these plots in the risk-benefit plane, we can assess the acceptability of pharmaceuticals. However, sometimes the goal of a BRA is not to decide if pharmaceuticals are acceptable, but which of a set of pharmaceuticals in the BRA is to be reimbursed. We thus need a way to directly compare pharmaceuticals. We propose to do this by comparing the benefit-risk ratios of two pharmaceuticals. This is done with a unit-less comparator unit C:

$$C = \frac{BRR_a}{BRR_b} \tag{4}$$

If C>1, then A outperforms B. This can be made clearer by rewriting equation 4 into:

$$C = \frac{Risks_b}{Risks_a} \cdot \frac{Benefits_a}{Benefits_b} \tag{5}$$

There are two ways in which A can outperform B (i.e. in which C can be more than 1). The first is if its weighted risks are lower than B's weighted risks, and if its weighted benefits are greater than B's weighted benefits. This means that both fractions in equation 5 are more than 1, so that C will also be more than one. It is imaginable that a new pharmaceutical has more benefits, but also more risks. This is the second way A can outperform B: if its weighted risks are greater than B's weighted risks, but if this is offset by its weighted benefits which are greater than B's benefits.

Since our model has Monte Carlo simulations, there is not one value of C for each pair of pharmaceuticals. It instead has a certain joint distribution. Figure 4 shows a histogram of a possible distribution of C. The area where A outperforms B is shaded. The probability that A outperforms B can be calculated by dividing the area where $C > 1$ by the area where $c < 1$.
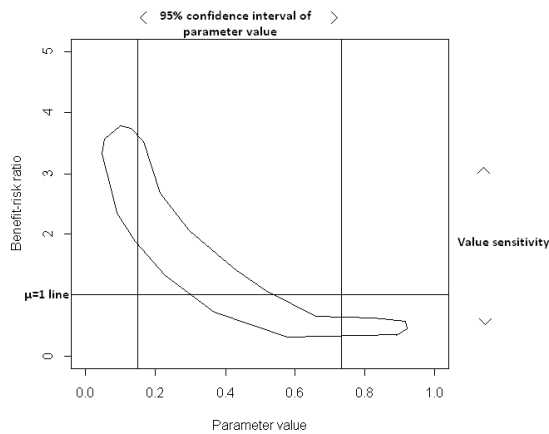

Figure 5: Graph for value sensitivity. This shows the sensitivity of a pharmaceutical's benefit-risk ratio to variation in a parameter. The area that is demarcated in the graph are simulation runs. With real data, this will take the shape of a point cloud. The 95% confidence interval of the parameter is denoted by the two vertical lines in the graph. The mean value at those lines can also be calculated. This is an indication for the value sensitivity to the parameter. If the cloud crosses the $\mu = 1$ line, the BRA is also decision sensitive to the parameter. In that case, a decision sensitivity plot like figure 6 may be more useful.
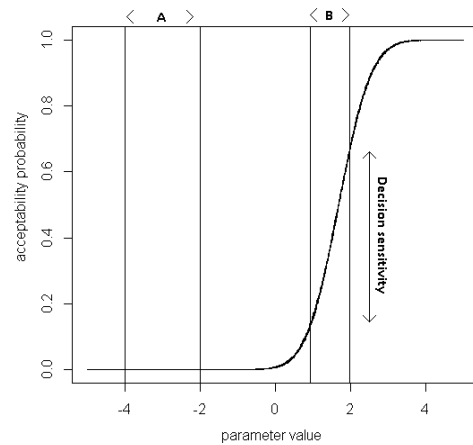
Figure 6: Decision sensitivity graph. This shows how sensitive the decision (probability of acceptance) is to changes in the parameter's value. The vertical lines in the graph denote the parameter's 95% confidence interval. Two examples are given of this as an illustration. If the parameter's 95% confidence interval is A, then the BRA is not decision sensitive to the parameter at all since changing the parameter does not change the decision, which is not to accept the pharmaceutical (probability of acceptance is always 0). If the parameter's 95% confidence interval is B, then the BRA is very decision sensitive to the parameter, since the probability of acceptability changes greatly with the parameter's currently known uncertainty.

Probabilistic sensitivity analysis

The risk-benefit plane can be used to visualize the problem and make the uncertainty surrounding preference and performance information explicit. It gives an overall view on the problem. However, sometimes regulators may want to take a more detailed look at the influence of individual parameters. This might be useful when they want to check the robustness of the decision or when they want to identify areas for further research. A more detailed look at the influence of particular parameters in our model can be given by sensitivity graphs (figure 5 and 6). There are two kinds of sensitivity graphs, a value sensitivity graph (figure 5) and a decision sensitivity graph (figure 6). The distinction between value sensitivity and decision sensitivity was explained by Felli and Hazen in 1998 [31]. Value sensitivity is the change in the magnitude of a model outcome that a parameter can achieve. In our case this is the

benefit-risk ratios of pharmaceuticals. Decision sensitivity looks only at the eventual decision being made: if variation in the parameter can mean another alternative is preferred, the model is decision sensitive to that parameter. Value sensitivity does not imply decision sensitivity, but decision sensitivity does imply value sensitivity.

Model inputs

Now that we have defined how the model works, we will describe what data is needed for the model. As described in the introduction, the goal of this model is to integrate uncertain patient preferences and clinical data. We therefore need two kinds of input: elicited preferences and clinical trial data. The elicited preferences will be used to get the weights and their distribution, and the clinical trial data will be used to get performance data and its distribution. We will now describe the two types of input data in more detail.

Elicited preferences
There is a multitude of preference elicitation methods. Two examples of methods that have been adapted for BRA are the analytical hierarchy process (AHP) [13], [43] and discrete choice experiments (DCE) like conjoint analysis [16], [44–47]. Both are examples of stated preference studies. In AHP, each respondent is asked to assess a number of pair wise criteria comparisons. In the comparisons, they indicate their preference for one of the two criteria in the pair wise comparison with numbers ranging from 1 (indifference) to 9 (extreme preference). These comparisons are put into a matrix where Eigen values determine the weight that is eventually assigned to the criteria. The outputs of an AHP study are weights for each of the criteria, with higher weights indicating a stronger preference. The sum of the weights is equal to one. In conjoint analysis, the preference respondents show for treatments are assumed to be a function of the attributes of these treatments. The respondents are asked to choose between a number of hypothetical treatments pairs which differ with regard to their attribute levels. Regression analysis is then used to determine the relative importance of attribute levels [44]. This can be compared to the weights as elicited in AHP. The difference between AHP and conjoint analysis in their outcome measures is that AHP gives a weight that is constant over performance ranges (the criteria level versus utility slope is constant), and conjoint analysis gives partial utility functions (e.g. criteria level versus utility slope can be non-constant).

Data from clinic trials
Clinical data can be derived from randomized clinical trials (RCT's). Other useful data sources are systematic reviews, which combine results from multiple RCT's. Since the data sources may vary, it is important to ensure clinical data comes from a reasonably homogeneous population. Also, there should be no difference in definitions (for example, some studies may define "survival" as one-year survival odds ratio, where others define it as two-year survival odds ratio). If all performance parameters are of the same unit such as pooled odds ratios compared to placebo, they can be used. However, if there is no single comparator across data sources, mixed treatment comparison may be useful to translate performances into a uniform unit such as odds ratio compared to a single comparator [48], [49]. Good systematic reviews should already have checked for quality measures such as those mentioned above or at least describe them, and as such are a preferred method of obtaining performance data.

Results

Now that the model and its inputs are described, the model will be illustrated by applying it to two cases: one on antidepressants and one on targeted agents for renal cancer. For each illustration we will describe the BRA context and the model definitions followed by how the preference and performance information were obtained and how the stochastic distributions for the two types of information were approximated. After information about the simulations we will present the risk-benefit plane, the comparisons between pharmaceuticals and examples of sensitivity graphs.

The antidepressants case

Major depression disorder (MDD) prevalence is rising in the modernized world and is defined as a long period of depression that interferes with normal functioning in daily life [50], [51]. One of the therapeutic options available is the use of antidepressant pharmaceuticals [52].

For our model, we will use the criteria as used by Danner *et al.* in an AHP study done in 2011 [13]. Two benefit criteria and one risk criterion were selected: response, remission and adverse events (AEs). Response is defined as a 50% reduction of acute depression symptoms. Remission is defined as a patient being below a threshold defined as constituting a major depressive episode (for example a HAMD17 score > 12) for a predefined amount of time. Adverse events are harmful side effects of the pharmaceuticals. The study by Danner *et al.* included more criteria, but these will be omitted here for clarity. The excluded criteria did not represent a clear benefit or risk aspect of the case. The alternatives that were evaluated are from a systematic review done by the German Institute for Quality and Efficiency in Health Care (IQWIQ) [53]. The pharmaceuticals that were evaluated are Duloxetine, Venlafaxine and Bupropion. Duloxetine and Venlafaxine are serotonine-norepinephrine reuptake inhibitors and Bupropion is a norepinephrine-dopamine reuptake inhibitor. The structure of our model can be seen in figure 7.
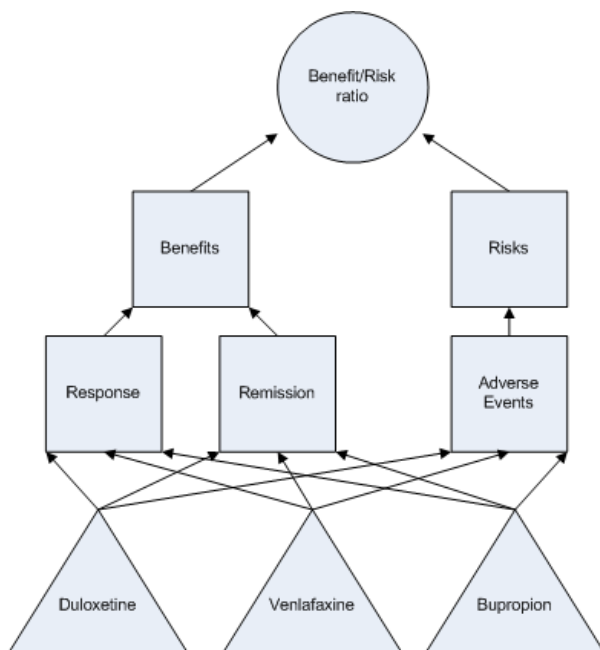

Figure 5: our MCDA model for the antidepressant case. The circle represents the MCDA goal, the squares are criteria and the triangles are alternatives.

The patient panel from which the preferences were elicited consisted of 12 patients with a previous diagnosis of depression who were currently in remission. They were asked to participate in an AHP group session where their preferences were elicited according to the AHP method described in the methods section. The normalized elicited weights can be seen in table 1.

Because of the small sample size, a nonparametric bootstrap method was used to approximate the distribution around the weights. For every run in the simulation, 12 respondents were selected with replacement at random from the group. With replacement means that a respondent can be selected more than once for a run. The mean weight

for each criterion from the 12 randomly selected respondents is then calculated. These are then used as the weights in the run.

Criterion \ Respondent	1	2	3	4	5	6	7	8	9	10	11	12
Response	0.566	0.368	0.374	0.784	0.445	0.781	0.777	0.707	0.721	0.624	0.525	0.667
Remission	0.193	0.176	0.335	0.117	0.308	0.071	0.124	0.145	0.186	0.092	0.274	0.072
Adverse Events	0.242	0.456	0.291	0.098	0.247	0.148	0.099	0.147	0.093	0.284	0.202	0.261

Table 1: Weights elicited with AHP as used for the model. Higher weights indicate a higher preference for a certain criterion. All weights are normalized per respondent, this means the sum of the weights per respondent equals 1. Source: [13]**.**

The performance data is derived from the IQWIG systematic review [53]. Pooled odds ratios for the pharmaceuticals were calculated in comparison with placebo. The distribution of pooled odds ratios was approximated by a normal distribution in the log domain. The data that was used for the performance approximations can be found in table 2.

	Pooled OR relative to placebo	Standard error
Duloxetine		
response	1.95	0.38
remission	1.91	0.35
adverse events	2.22	0.40
Venlafaxine		
response	2.04	0.44
remission	1.97	0.39
adverse events	2.47	0.56
Bupropion		
response	1.48	0.13
remission	1.46	0.11
adverse events	1.00	0.18

Table 2: performances of the three alternatives as used in the model. All pooled odds ratios are relative to placebo. A pooled odds ratio >1 means that the performance on that criterion was higher for the pharmaceutical than for placebo, which is favorable in benefit criteria but unfavorable in risk criteria.

When the distributions were approximated, the problem could be simulated. This was done with the statistical programming language R [41]. A detailed description of the code that was used can be found in appendix A. In total, 1000 simulation runs were performed for the graphs and 100,000 for the numeric values such as comparator C. The result in the risk-benefit plane can be seen in figure 8. As can be clearly seen, the simulation runs for each pharmaceutical are below the acceptability threshold, which means all pharmaceuticals have a probability of being acceptable of 1. Figure 9 show the comparisons between the pharmaceuticals' performance. From these, the probability of one pharmaceutical outperforming the other can be calculated, this is shown in table 3. Bupropion has the highest probability to outperform the other two pharmaceuticals (94.5% and 96.4%). The difference between Duloxetine and Venlafaxine is less pronounced, Duloxetine has a probability of outperforming Venlafaxine of 57.5%. An example of a value sensitivity graph and a decision sensitivity graph can be seen in figure 10.
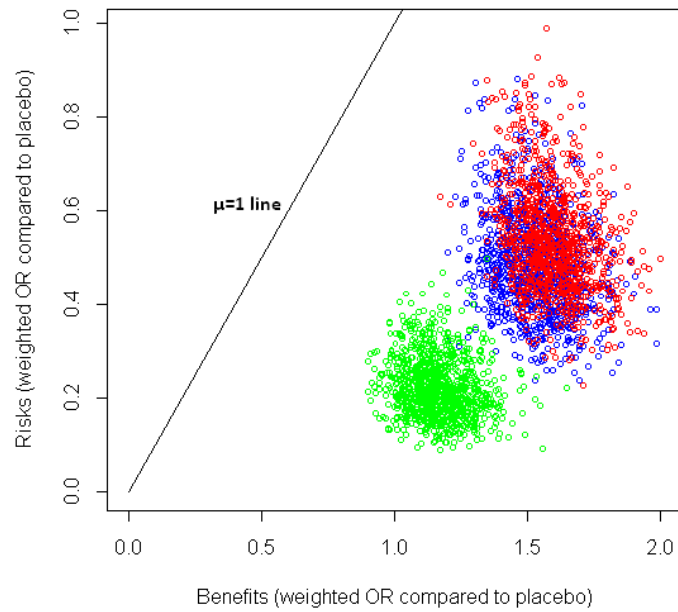
Figure 6: Risk-benefit plane of the antidepressants case. The points of a particular colour indicate simulation runs for a particular pharmaceutical. The blue points represent Duloxetine, the red ones Venlafaxine and the green ones Bupropion. All pharmaceuticals have a probability of being acceptable of 1, since all points are below the $\mu = 1$ line.

Figure 7: Performance comparisons between the three pharmaceuticals. Shown are the histograms of C, when used to compare the three pharmaceuticals. What two pharmaceuticals are compared is described in the title of each histogram and in the title of the x-axis. The line where C=1 (i.e. where no pharmaceutical outperforms the other), is shown. The shaded areas indicate C>1. BRR=benefit-risk ratio.
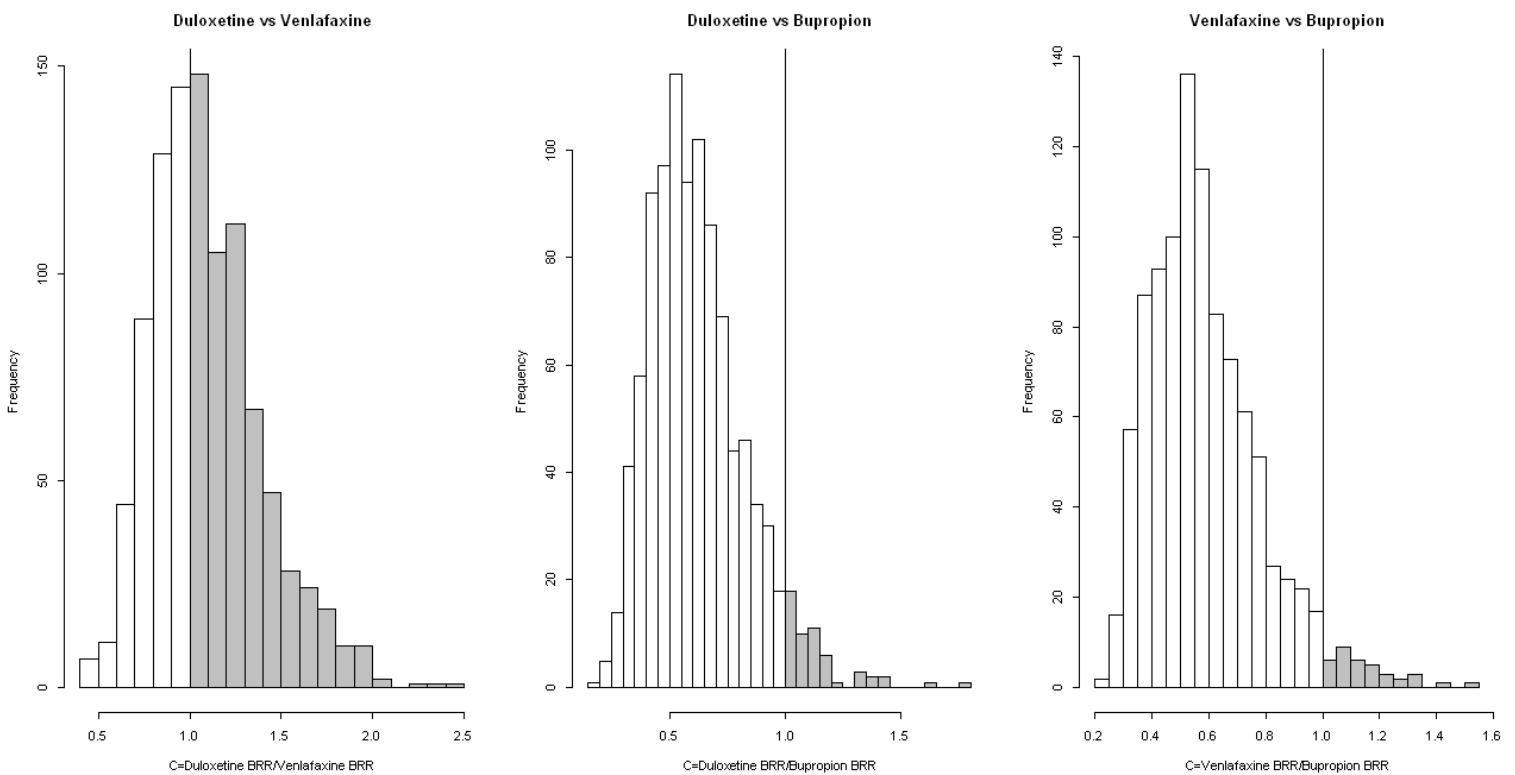
Probability that	Duloxetine outperforms…	Venlafaxine outperforms…	Bupropion outperforms…
Duloxetine	N/A	42.5%	94.5%
Venlafaxine	57.5%	N/A	96.4%
Bupropion	5.5%	3.6%	N/A

Table 3: This table shows the probabilities pharmaceuticals have to outperform other pharmaceuticals, given the current uncertainty surrounding the preference and performance parameters. n/a=not applicable.
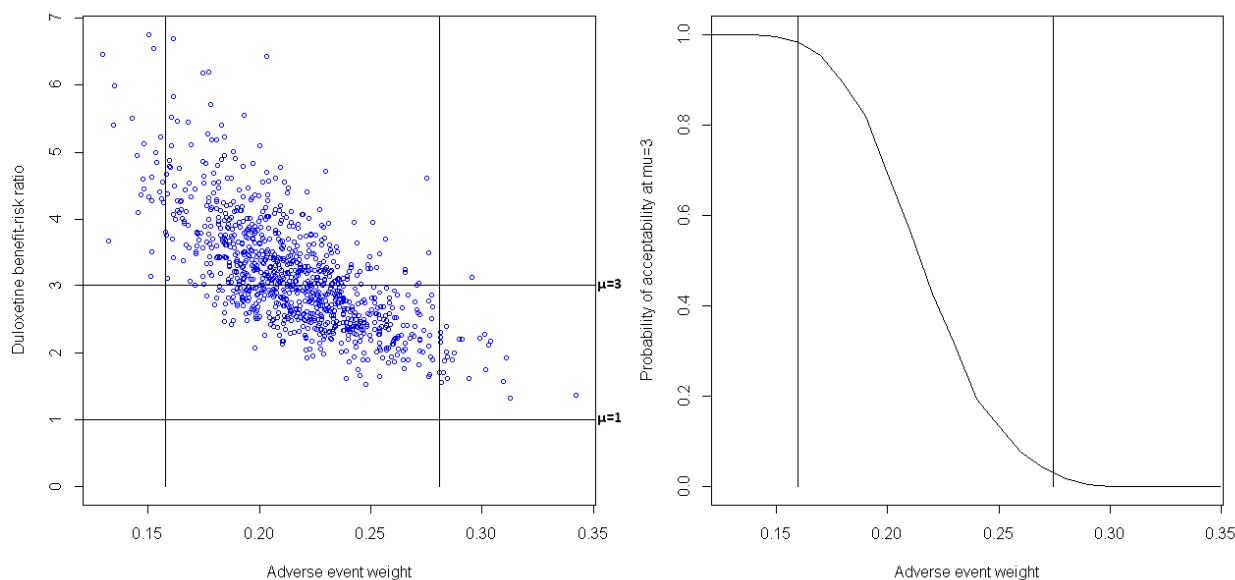

Figure 8: Sensitivity graphs for a preference parameter in the antidepressants: the weight patients give to the adverse event criterion. Shown on the left is the value sensitivity graph. On the x axis is the weight of the adverse event criterion. The two vertical lines in the graph denote the 95% confidence interval on that criterion's weight. The two horizontal lines denoted two threshold, , $\mu = 1$ and $\mu = 3$. On the right we show the change in acceptability probability at a threshold of $\mu = 3$ when the criterion's weight changes. The vertical lines again denote the 95% confidence interval of the criterion's weight. We see that the decision is sensitive to the criterion's weight, because the probability of acceptability changes considerably over the criterion weight's 95% confidence interval. Note that the $\mu = 3$ threshold is not normally used in our model, but its decision sensitivity graph is shown here because all pharmaceuticals in the antidepressants case have a probability of acceptability of 1. In that case, the graph on the right would should a line at 1, since the pharmaceutical will always be accepted.

The renal cell cancer medication case

25-30% of kidney cancer cases are a form of renal cell cancer (RCC) [45], [54]. The most common clear cell type is a by malignant neoplasm of the proximal tubular epithelial cells. Until recently the accepted RCC therapy consisted of cytokine immunotherapy that had little effect and a high adverse event incidence, giving RCC patients a bad prognosis [54], [55]. Targeted agents, pharmaceuticals that target specific molecular pathways instead of working systemically, have improved overall survival considerably [56]. However, they also have risks associated with them, something that is more important now that patients are being treated with the targeted agents for longer periods of time. It is therefore important to examine the benefit-risk profiles of these targeted agents.

The structure of the model we will use for this case is based on a preference study by Mohamed *et al.* done in 2011 [45]. They defined criteria after interviews with clinical experts and ten patients with RCC [45]. Progression free survival (PFS) is the only benefit criterion. It is defined as time until tumor

progression. This criterion is divided into three levels (sub criteria in our model): PFS at two months, PFS at 12 months and PFS at 24 months. The risk criteria are fatigue, diarrhea, hand-foot syndrome, mouth sores, liver failure (elevated liver enzymes risk) and blood cloth (cardiovascular risk). After evaluating the available evidence, we selected three alternatives. These were Sorafenib after treatment with cytokines, Pazopanib after treatment with the cytokine interferon-α (IFN-α) and Everolimus after failed vascular endothelial growth factor (VEGF) receptor inhibitor therapy. Figure 11 shows the MCDA structure for this case.
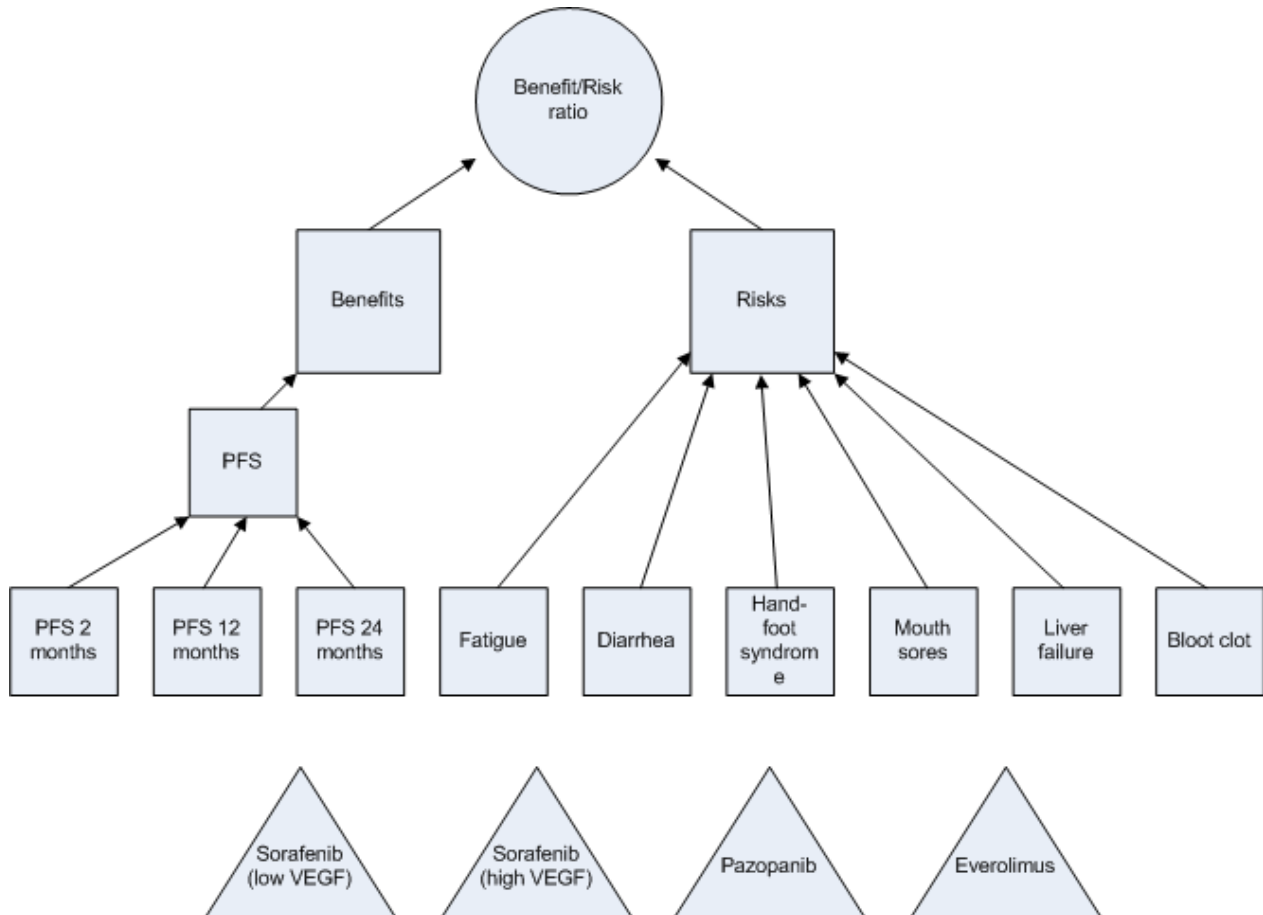

Figure 9: MCDA structure for the renal cell cancer case. The squares are criteria, the triangles are alternatives. There is a line connecting each alternative to each criterion indicating the performance, but these were omitted from this figure for clarity. The Sorafenib alternative is split into a low baseline vascular endothelial growth factor (VEGF) group and a high baseline VEGF group because of a significant difference in performance between the two groups [57]**.**

The preference study by Mohamed *et al.* was a conjoint analysis study [45]. In it, a questionnaire was sent out to adult self-reported RCC patients living in the United States. Two versions of the questionnaire were sent out, each with 12 tradeoff questions between two hypothetical pharmaceuticals that had particular attributes. 138 patients filled in the questionnaire (response rate 3%). The response rate was low because only a small portion of contacted people were eligible to enter the study. The mean weights that were elicited can be found in the third column of table 4. PFS was the most important criterion, followed by fatigue and diarrhea.

Performance data was derived from a Cochrane systematic review [55], [58]. For more details, we referred to the papers cited by this systematic review. All performances were in odds ratio compared to placebo. During the modeling process, some assumptions were made. These will now be discussed per criterion.

None of the clinical trials reported PFS of 24 months for the pharmaceuticals or for placebo. The performance on this subcriterion was therefore assumed to be equal (constant odds ratio of 1) between the alternatives and placebo. In the Everolimus trial, there was no PFS of 12 months in either the Everolimus of placebo group, so there the performance between Everolimus and placebo was also assumed to be equal. The Sorafenib trial reported a significant difference in PFS between patients with a low baseline VEGF and patients with a high baseline VEGF. Because of these differences and because no data is available for the whole Sorafenib group, the Sorafenib alternative were split up into Sorafenib (low baseline VEGF) and Sorafenib (high baseline VEGF). The risk incidence for these two groups was not different. We did not find data on the VEGF levels of the patients in the Pazopanib and Everolimus trials.

No exact data was found on the performance of Pazopanib on the criterion hand-foot syndrome. The clinical trial did mention that hand-foot syndrome occurred with "an incidence of fewer than 10%" [59]. Since the Pazopanib group had 290 patients, an incidence of less than 10% can thus mean that anywhere from 1 to 28 patients where affected. We thus approximated the performance of Pazopanib with a uniform distribution between 1 and 28. To come to the required odds ratio, data was also needed for the placebo group Pazopanib was compared with. Since no data on hand-foot syndrome was available for that group, we assumed the incidence of hand-foot syndrome would be equal to the incidence in other RCC patient groups who received placebo. This same problem with a lack of data on hand-foot syndrome data was present for Everolimus. There was no mention of hand-foot syndrome in the clinical trial of Everolimus versus placebo [60]. That there was no hand-foot syndrome incidence was confirmed after a query to the article's corresponding author. This seems to contradict other trials with Everolimus where there is hand-foot syndrome incidence, but these are of Everolimus in combination with Sorafenib [61], of Everolimus in combination with Cetuximab and Capecitabine in pancreatic cancer patients [62] and of Everolimus in combination with Capecitabine [63]. We thus assumed that the incidence of hand-foot syndrome in Everolimus was equal to that of placebo.

Mouth sores are not mentioned as such in the clinical trials. For Sorafenib, the oral mucositis incidence was used, and for Everolimus the mucosal inflammation incidence. The Pazopanib trial did not mention mouth sores, oral mucositis or mucosal inflammation incidence so performance was assumed to be equal to placebo.

Because of the rarity of hepatic failure events in clinical trials, and the tendency of clinical trials to not report adverse events that occurred in less than a preset percentage of the population (usually 1% or 10%), obtaining data on liver failure was difficult. Instead, a surrogate endpoint for liver failure was chosen: an increased serum alanine aminotransferase (ALT) concentration. Of the various indicators for liver damage, ALT was chosen because it is sensitive to liver damage and because it was a reported indicator in the included clinical trials [64], [65]. More indicators could have been chosen to be used as sub criteria to the liver failure criterion. This was not done for the following reasons. First of all, the

relative weights of the sub criteria would be unknown. This can be obtained by eliciting preferences from (renal) oncology specialists, but as far as we know no such study has yet been done. Secondly, the sub criteria were unlikely to be independent and it was probable that patients would be double-counted due to the manner in which risk data was reported in clinical trials. No patient level data was available to check for this correlation or to correct this. No data was available on the incidence of increased serum ALT for Sorafenib. A phase II clinical trial did report ALT increases, but the data is difficult to use because of the large amount of crossover patients and lack of patient level data. Because of this, we assumed Sorafenib's performance on the liver failure criterion to be equal to that of placebo.

We assumed cardiovascular adverse events such as myocardial infarction to be part of the blood clot criterion.

When the preference and performance information was gathered, and assumptions were made and described, we could approximate the distribution of the parameters for use in the simulation model. The weights of the criteria were normally distributed [45] and can be found in table 4.

Criterion	Distribution	Mean	Standard error
PFS 24m	Normal	10.00	0.85
PFS 12m	Normal	3.81	0.34
PFS 2m	Normal	0.00	0.81
Fatigue	Normal	6.19	0.95
Diarrhea	Normal	5.28	0.75
Hand-foot syndrome	Normal	4.69	0.83
Mouth sores	Normal	2.56	0.79
Liver failure	Normal	5.17	0.76
Blood clot	Normal	3.39	0.61

Table 4: parameters for the stochastic distribution for the preference parameters (weights). Higher weights indicate a greater importance. All parameters are normally distributed.

The performances of odds ratio's were approximated with a normal distribution in the log domain. The parameters for this can be seen in table 5. The standard error of the odds ratio was estimated with $\sqrt{\frac{1}{ei} + \frac{1}{nei} + \frac{1}{ec} + \frac{1}{nec}}$, where ei is the number of events in the intervention (pharmaceutical) group, nei is the number of non-events in the intervention group, ec is the number of events in the control (placebo) group and nec is the number of non-events in the control group.

	Distribution	Mean	Standard error
Sorafenib after cytokines [57], [61], [66]			
PFS			
Low baseline VEGF			
2 months	Normal in log domain	1.26	1.18

12 months	Normal in log domain	1.27	1.34
24 months	Constant	1.00	n/a
High baseline VEGF			
2 months	Normal in log domain	1.51	1.18
12 months	Normal in log domain	2.60	1.50
24 months	Constant	1.00	n/a
Fatigue	Normal in log domain	1.79	1.17
Diarrhoea	Normal in log domain	4.40	1.19
Mouth sores	Normal in log domain	2.87	1.52
HF syndrome	Normal in log domain	2.64	1.18
Liver failure (ALT)	Constant	1.00	n/a
Blood clot (cardiovascular risk)	Normal in log domain	2.99	2.27
Pazopanib after no therapy or IFN-α [59]			
PFS			
2 months	Normal in log domain	1.30	1.17
12 months	Normal in log domain	2.75	1.27
24 months	Constant	1.00	n/a
Fatigue	Normal in log domain	2.50	1.41
Diarrhoea	Normal in log domain	5.77	1.36
Mouth sores	Constant	1	n/a
HF syndrome	Normal in log domain	0.02 to 1.16	1 to 0.3
Liver failure	Normal in log domain	2.38	1.25
Blood clot	Normal in log domain	1.11	1.80
Everolimus after VEGF inhibitor therapy [60–63]			
PFS			
2 months	Normal in log domain	2.09	1.25
12 months	Constant	1.00	n/a
24 months	Constant	1.00	n/a
Fatigue	Normal in log domain	1.22	1.32
Diarrhea	Normal in log domain	5.83	1.70
Mouth sores	Constant	1.00	n/a
HF syndrome	Constant	1.00	n/a
Liver failure	Normal in log domain	4.87	1.62
Blood clot	Constant	1.00	n/a

Table 5: distributions of the performance parameters of the renal cell cancer case. The first column shows the alternatives and the criteria on which their performance was measured (indented). The clinical trials from which the performance data was obtained are noted behind the alternative names. The second column shows the type of approximation that was used and the third and fourth columns show the mean and standard errors, respectively. The fourth column shows n/a (not applicable) for constants.

With the model structure, preference distributions and performance distributions known, the model could be programmed. The simulations were programmed in R [41]. The code can be found in appendix B. A total of 1000 simulation runs were performed for the graphs in order to keep them comprehensible. For the numeric values and the decision sensitivity graphs, we performed 100,000 simulation runs to increase the accuracy of the joint distribution estimations. The summed benefits and the summed risks are plotted in the risk-benefit plane as can be seen in figure 12. Here we see that none of the pharmaceuticals pass the $\mu = 1$ line. This means all pharmaceuticals have a probability of being acceptable of zero. When comparing the pharmaceuticals with comparator C (figure 13 and table 6), both Sorafenib in low VEGF patients and Everolimus have high probabilities of outperforming Pazopanib and Sorafenib in high VEGF patients. When comparing Sorafenib in low VEGF patients and Everolimus, Sorafenib in low VEGF patients outperforms Everolimus with a probability of 66.3%.
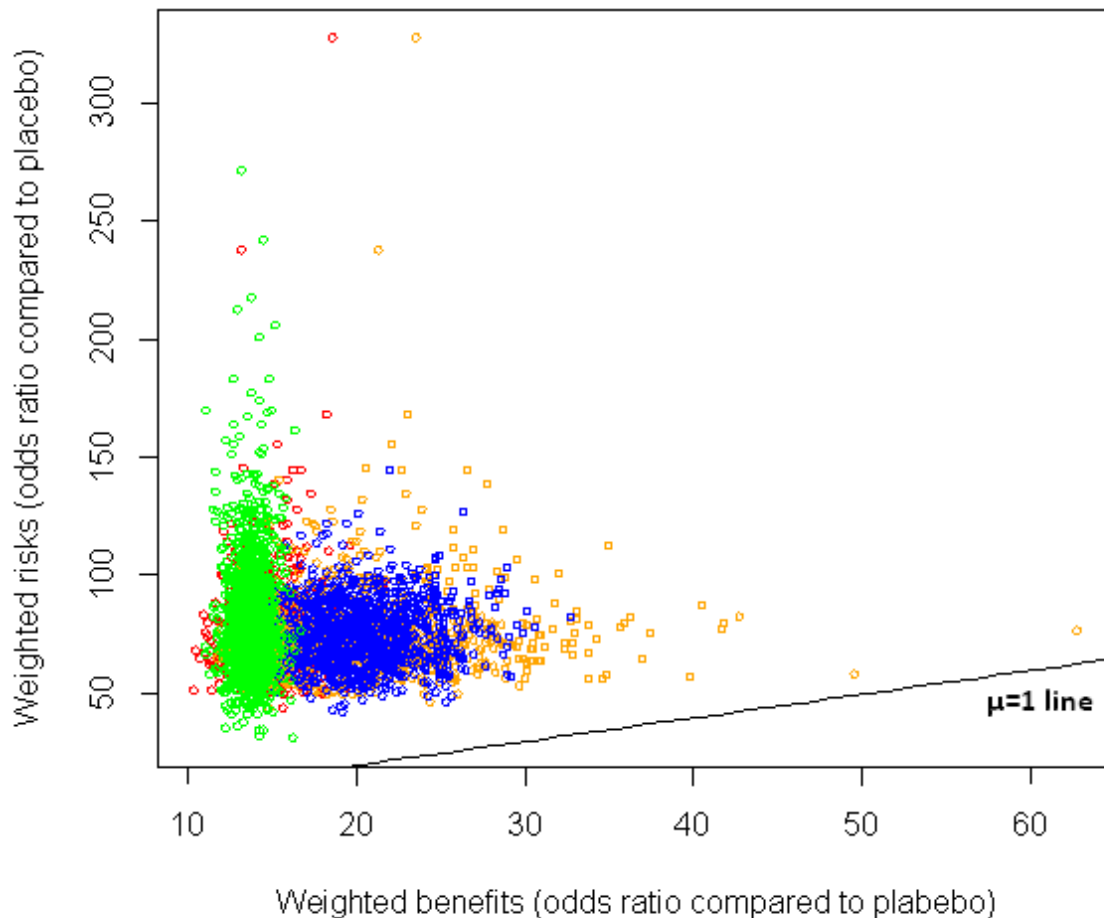

Figure 10: Risks-benefit plane of the renal cell cancer case. The red points are simulation outputs for Sorafenib in low baseline VEGF patients. The orange points are simulation outputs for Sorafenib in high baseline VEGF patients. Pazopanib's simulation outputs are the blue points, and Everolimus' the green points.
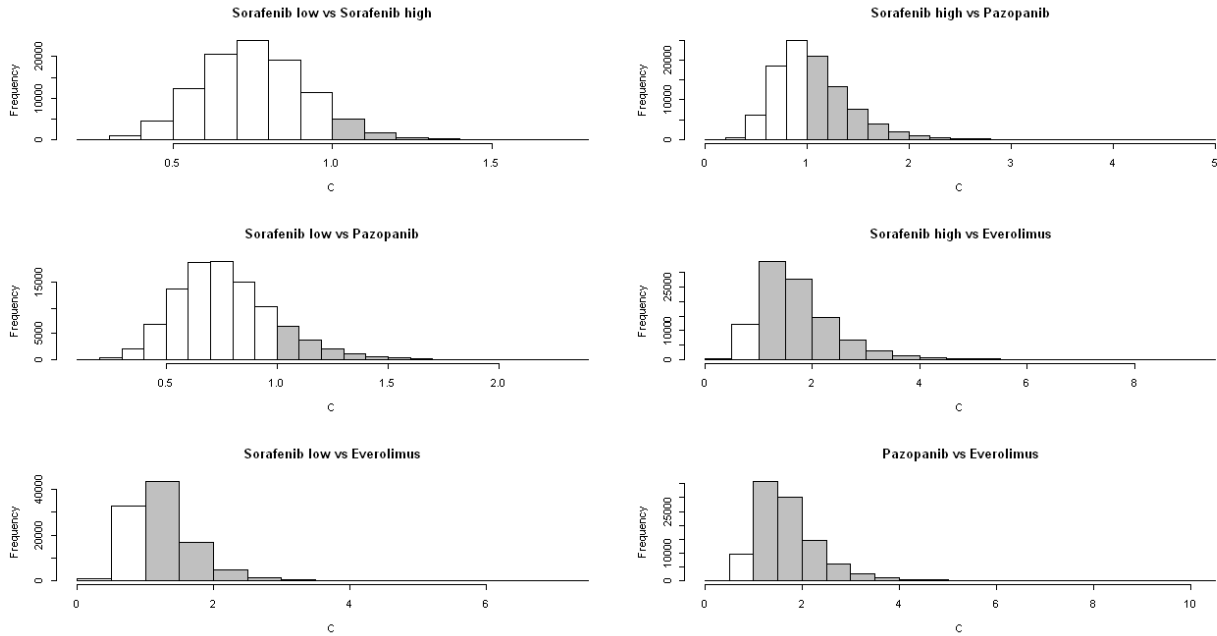
Figure 11: Comparisons between the pharmaceuticals. Shown are frequency histograms of the distribution of comparator variable C, which is defined as $\frac{BRR_a}{BRR_b}$. So for example, the bottomright graph shows the distribution of $\frac{BRR_{Pazopanib}}{BRR_{Everolimus}}$. Where C>1, the graph is shaded grey. From this distribution, the probability that a pharmaceutical outperforms another pharmaceutical can be calculated by comparing the amount of data on both sides of the C=1 line, see table 6.

Probability that…	Sorafenib low baseline VEGF outperforms…	Sorafenib high baseline VEGF outperforms…	Pazopanib outperforms…	Everolimus outperforms…
… Sorafenib low baseline VEGF	n/a	92.6%	85.7%	33.6%
Sorafenib high baseline VEGF	7.4%	n/a	50.1%	12.2%
Pazopanib	14.2%	49.9%	n/a	9.6%
Everolimus	66.3%	87.8%	90.4%	n/a

Table 6: Comparisons between pharmaceuticals. This table shows the probability that particular pharmaceuticals have to outperform another pharmaceutical (i.e. the percentage of simulation runs where its benefit-risk ratio was higher). n/a=not applicable.
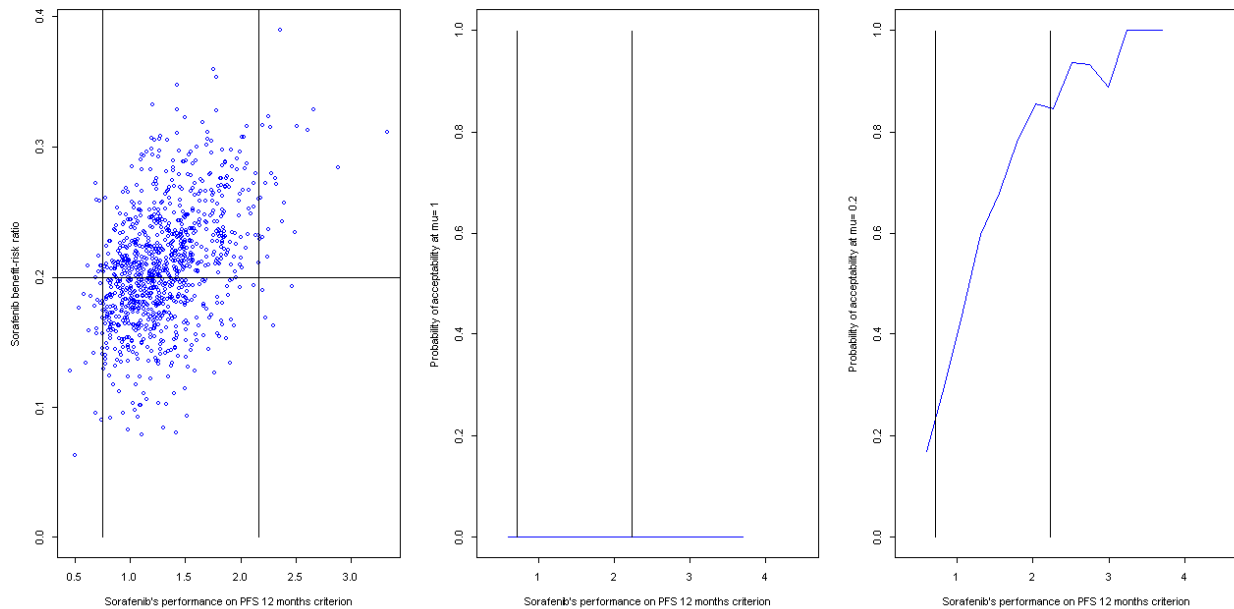
Figure 12: Example of sensitivity graphs for the renal cell cancer case. Shown is the sensitivity of Sorafenib (high baseline VEGF) benefit-risk ratio to its performance on the PFS 12 months criterion. The left graph shows the value sensitivity graph. The centre graph shows the decision sensitivity for a $\mu = 1$ acceptability threshold. The decision is not sensitive to the parameter, as the probability of acceptability remains 0 irrespective of the parameter's value. This is logical when looking at this case's risk-benefit plane (figure 12), where we see that all simulation outputs are above the $\mu = 1$ line. The right graph shows the decision sensitivity graph for a $\mu = 0.2$ threshold. Here, the decision is sensitive to the value of the performance of Soranefib (high baseline VEGF) on the PFS 12 months criterion. In all the graphs, the vertical lines denote the 95% confidence interval of the parameter on the x axis.

Discussion

With the above two cases, we aimed to illustrate the possibilities of our model. They show an integration of patient preferences and clinical data is possible, taking into account the uncertainty surrounding these types of data. Interesting about the antidepressant case is the difference between Bupropion and the two other alternatives. Although the other alternatives have more benefits, these do not weigh up to the added risks and Bupropion is considered to outperform them. The sensitivity graphs show a characteristic shape for the weights of benefit and risk criteria. When benefit criteria have more weight, the benefit-risk ratio and probability acceptability of the alternatives improves. On the other hand, when risk criteria have a higher weight, the benefit-risk ratio and probability of acceptability deteriorates (figure 10). The renal cancer medication case shows an interesting disparity between the model and clinical practice. Although all the evaluated alternatives had a probability of acceptability of 0, they are being widely used in health care with great satisfaction. A reason for this can be found in Mohamed *et al.*'s study: overall survival is not taken into account as a seperate criterion, while for patients this is one of the reasons they consider the targeted agents to be preferable over other kinds of therapy [56]. Apart from the described assumptions, two other factors may have negatively influenced the results of the renal cancer case. First of all, although the Sorafenib trial showed a correlation between baseline VEGF

level and PFS, the baseline VEGF levels of the Pazopanib and Everolimus trial were unknown. This may have confounded the results. Secondly, there was a low incidence of adverse events in the trials. The estimator for the standard error of an odds ratio that was used may give a biased estimate if the incidence in one of the groups is low.

Our model has some inherent limitations which we will now discuss. First of all, for a valid application of the model considerable thought needs to go into how the BRA is translated to the model and this can be time consuming. It is important to clearly define criteria and performance measures, and use inputs that comply with these definitions as much as possible. If this is not possible, the parameters should be translated or it should be mentioned in the report that certain assumptions exists. With the whole modelling proces described, including assumptions, there can be a more grounded discussion of the model's validity. In this phase, the input of clinical experts is important.

The second limitation, linked to the first, is that structural uncertainty (uncertainty about whether or not the model represents reality) is not taken into account in the model itself. Clinical experts and regulators taking part in the model building proces can help to reduce the structural uncertainty. An example of the structural uncertainty in our model is the assumption of criterion weight independence. The weights vary according to their own distribution without regard to the weight of other criteria, while in reality the weights may be correlated. For example, the weights of risk criteria may be correlated because of the risk attitude of patients.

Third, there is some hesitation in the BRA field to use the benefit-risk ratio as an aggregate measure to compare pharmaceuticals with. We argue that the benefit-risk ratio can be a useful and consistent aggregate measure when the benefits and risks are weighted according to preference elicitation studies that incorporate tradeoffs. With $\mu = 1$, acceptable benefit-risk ratios are consistent with the preferences of the patients who's preferences were elicited. However, the validity of this approach depends on the structure of the criteria in the preference study. The hesitation of regulators to use the benefit-risk ratio is linked to a hesitation to use quantitative models in a decision supporting role. We consider the best role of our model is a supplementary role if the modeling is done outside the agency, and a structuring role if the modeling is done in or together with experts working with the agency.

The fourth limitation is that it is unclear what probability of acceptability or probability of outperforming is considered conclusive. This limitation can be illustrated by the antidepressants case. The probability that Bupropion outperforms Duloxetine or Venlafaxine is 94.5% and 96.4% respectively. Here, it is clear that Bupropion outperforms the other two pharmaceuticals. The distinction between Duloxetine and Venlafaxine is less clear (probability that Duloxetine outperforms Venlafaxine is 57.5%).

There are several things that further research can do to increase the value of our model. First, although the sensitivity graphs are visually informative, research should be done to quantify these sensitivity measures. This can give regulators a better idea of the sensitivity of parameters when these are not directly clear by looking at the sensitivity graphs. A way of doing this may be using value of information methods [31], [67]. This can not only help give regulators a better idea of the sensitivity, but it may also indicate areas for further research. Parameters with a high value of information may call for new research into those parameters. For example, if a BRA is most decision sensitive to the performance of a

particular pharmaceutical, a clinical trial can be set up to gather more data on this performance [67], [68]. To estimate the impact of new research on the decision, Bayesian methods can be used [68]. In order to be able to use value of information measures for preference studies, research should be done on the likelihood ratio (≈predictive power) of different types of preferences studies so they can be used in Bayesian probability revision methods. Secondly, the model could be developed to incorporate the preference data from multiple preference studies and of different types of preference studies. Linked to this is the integration of various types of performance parameters (dichotomous, discrete and continous data) in one model. This can potentially be done by using partial value functions such as in the Mussen model [24]. Third, a method should be sought to include correlation between parameters in the model. Finally, more research should be done into the notion of 'value' and units in BRA. There are several aspects to this. Value and units should theoretically fit into a model so conclusions of the model are internally valid, and the units and value should be understandable and usable by regulators and patients. If they are not, the model can run the risk of being overly theoretical, reducing application.

Conclusion

This study shows that it is possible to integrate elicited patient preferences and clinical data in a quantitative model for benefit-risk assessment. The model can handle preference information from AHP and conjoint analysis studies and performance information from clinical trials and systematic reviews. It can assess the impact of the uncertainty around parameters by simulating the problem with approximated joint distributions. The graphs that can be produced with the model can give an overview of the characteristics of the problem, structuring the available evidence for regulators. We tested the model in two cases which served to illustrate our model. It is important that modellers, regulators and clinical experts collaborate in real-life applications of quantitative models like this. This is especially important during the first stages of modelling, when the model structure is decided, data is selected and assumptions are made.

References cited

[1]   EMA, "Work package 2 report: Applicability of current tools and processes for regulatory benefit-risk," 2011.

[2]   EMA, "Benefit-Risk Balance for Marketed Drugs: Evaluating Safety Signals. Report of CIOMS Working Group IV," Geneva, 1998.

[3]   J. J. Guo, S. Pandey, J. Doyle, B. Bian, Y. Lis, and D. W. Raisch, "A review of quantitative risk-benefit methodologies for assessing drug safety and efficacy-report of the ISPOR risk-benefit management working group.," *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research*, vol. 13, no. 5, pp. 657–66, Aug. 2010.

[4]   W. Holden, "Benefit-risk analysis: a brief review and proposed quantitative approaches," *Drug Safe*, vol. 26, pp. 853–862, 2003.

[5]   Z. Yuan, B. Levitan, and J. A. Berlin, "Benefit-risk assessment : to quantify or not to quantify , that is the question," *Pharmacoepidemiology and Drug Safety*, no. April, pp. 653–656, 2011.

[6]   N. Zafiropoulos, L. Phillips, F. Pignatti, and X. Luria, "An Agency perspective," vol. 9, no. 6, 2012.

[7]   EMA, "Regulation (EU)1235/2010," *Official journal, L series*, no. 348, pp. 1–16, 2010.

[8]   EMA, "Directive (EU) 2010/84," *Official journal, L series*, no. 348, pp. 74–99, 2010.

[9]   EMA, "Plan for implementation of the pharmacovigilance legislation by the European Medicines Agency," 2012. [Online]. Available: http://www.ema.europa.eu/docs/en_GB/document_library/Other/2012/02/WC500121837.pdf.

[10]   K. Facey, A. Boivin, J. Gracia, H. P. Hansen, A. Lo Scalzo, J. Mossman, and A. Single, "Patients' perspectives in health technology assessment: a route to robust evidence and fair deliberation.," *International journal of technology assessment in health care*, vol. 26, no. 3, pp. 334–40, Jul. 2010.

[11]   E. W. de Bekker-Grob, M. L. Essink-Bot, W. J. Meerding, B. W. Koes, and E. W. Steyerberg, "Preferences of GPs and patients for preventive osteoporosis drug treatment: A discrete choice experiment," *PharmacoEconomicsPharmacoEconomics*, vol. 27, no. 3, pp. 211–219, 2009.

[12]   F. R. Johnson, A. B. Hauber, S. Özdemir, C. A. Siegel, S. Hass, and B. E. Sands, "Are gastroenterologists less tolerant of treatment risks than patients? Benefit-risk preferences in Crohn's disease management," *Journal of managed care pharmacy*, vol. 16, no. 8, pp. 616–628, 2010.

[13]   M. Danner, J. M. Hummel, F. Volz, J. G. van Manen, B. Wiegard, C.-M. Dintsios, H. Bastian, A. Gerber, and M. J. Ijzerman, "Integrating patients' views into health technology assessment: Analytic hierarchy process (AHP) as a method to elicit patient preferences.," *International journal of technology assessment in health care*, vol. 27, no. 4, pp. 369–75, Oct. 2011.

[14]   J. F. P. Bridges and C. Jones, "Patient-based health technology assessment: a vision of the future.," *International journal of technology assessment in health care*, vol. 23, no. 1, pp. 30–5, Jan. 2007.

[15]   N. Guo, C. A. Marra, J. M. Fitzgerald, R. K. Elwood, A. H. Anis, and F. Marra, "Patient Preference for Latent Tuberculosis Infection Preventive Treatment : A Discrete Choice Experiment," *Value in Health*, vol. 14, no. 6, pp. 937–943, 2011.

[16]   F. R. Johnson, G. Van Houtven, S. Ozdemir, S. Hass, J. White, G. Francis, D. W. Miller, and J. T. Phillips, "Multiple sclerosis patients' benefit-risk preferences: serious adverse event risks versus treatment efficacy.," *Journal of neurology*, vol. 256, no. 4, pp. 554–62, Apr. 2009.

[17]   F. R. Johnson, S. Ozdemir, B. Hauber, and T. L. Kauf, "Women's willingness to accept perceived risks for vasomotor symptom relief.," *Journal of women's health (2002)*, vol. 16, no. 7, pp. 1028–40, Sep. 2007.

[18]   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.

[19]   L. D. Phillips, B. Fasolo, N. Zafiropoulos, and A. Beyer, "Is quantitative benefit–risk modelling of drugs desirable or possible?," *Drug Discovery Today: Technologies*, vol. 8, no. 1, pp. e3–e10, Mar. 2011.

[20]   B. S. Levitan, E. B. Andrews, a Gilsenan, J. Ferguson, R. a Noel, P. M. Coplan, and F. Mussen, "Application of the BRAT framework to case studies: observations and insights.," *Clinical pharmacology and therapeutics*, vol. 89, no. 2, pp. 217–24, Mar. 2011.

[21]   H. Simon, "A Behavioral Model of Rational Choice," in *Models of Man, Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting*, New York: Wiley, 1957.

[22]   K. D. Miller, "Simon and Polanyi on Rationality and Knowledge," *Organization Studies*, vol. 29, no. 7, pp. 933–955, May 2008.

[23]   L. P. Garrison, A. Towse, and B. W. Bresnahan, "Assessing a structured, quantitative health outcomes approach to drug risk-benefit analysis.," *Health affairs (Project Hope)*, vol. 26, no. 3, pp. 684–95, 2007.

[24]   F. Mussen, S. Salek, and S. Walker, "A quantitative approach to benefit-risk assessment of medicines – part 1 : The development of a new model using multi-criteria decision analysis y," *wiley interscience*, no. June, pp. 2–15, 2007.

[25]   F. R. Johnson and A. B. Hauber, "Quantifying Patient Benefit-Risk Tradeoff Preferences : A Brief Introduction," *RTI White paper*, 2008.

[26]   D. Ashby and A. Smith, "Evidence-based medicine as Bayesian decision-making," *Statistics in medicine*, vol. 19, no. 23, pp. 3291–3305, 2000.

[27]   P. Goodwin and G. Wright, *Decision Analysis for Management Judgment.* Chichester: John Wiley & Sons, 2009.

[28]   I. Linkov, F. K. Satterstrom, J. Steevens, E. Ferguson, and R. C. Pleus, "Multi-criteria decision analysis and environmental risk assessment for nanomaterials," *Journal of nanoparticle research*, vol. 9, no. 4, pp. 543–554, 2007.

[29]   S. R. Haynes, T. G. Kannampallil, and L. L. Larson, "Optimizing Anti-Terrorism Resource Allocation," vol. 56, no. 3, pp. 299–309, 2005.

[30]   I. B. Huang, J. Keisler, and I. Linkov, "Science of the Total Environment Multi-criteria decision analysis in environmental sciences : Ten years of applications and trends," *Science of the Total Environment, The*, vol. 409, no. 19, pp. 3578–3594, 2011.

[31]   J. Felli, "Sensitivity analysis and the expected value of perfect information," *Medical Decision Making*, vol. 18, no. 1, pp. 95–109, 1998.

[32]   T. Tervonen, G. van Valkenhoef, E. Buskens, H. L. Hillege, and D. Postmus, "A stochastic multicriteria model for evidence-based decision making in drug benefit-risk analysis.," *Statistics in medicine*, vol. 30, no. 12, pp. 1419–28, May 2011.

[33]   G. van Valkenhoef, T. Tervonen, J. Zhao, B. de Brock, H. L. Hillege, and D. Postmus, "Multicriteria benefit-risk assessment using network meta-analysis.," *Journal of clinical epidemiology*, vol. 65, no. 4, pp. 394–403, Apr. 2012.

[34]   L. D. Lynd and B. J. O'brien, "Advances in risk-benefit evaluation using probabilistic simulation methods: an application to the prophylaxis of deep vein thrombosis.," *Journal of clinical epidemiology*, vol. 57, no. 8, pp. 795–803, Aug. 2004.

[35]   R. Schäfer and D. Gmbh, "Rules for Using Multi-Attribute Utility Theory for Estimating a User ' s Interests," 1999.

[36]   G. Fishmann, *Monte Carlo. Concepts, Algorithms, and Applications*. Springer-Verlag New York, Inc., 1996.

[37]   W. J. Krzanowski, *An Introduction to Statistical Modelling*, 1st ed. London: Hodder headline group, 2002.

[38]   B. J. T. Morgan, *Applied Stochastic Modelling*, 2nd ed. Boca Raton: Taylor & Francis group, 2009.

[39]   J. H. A. Biswas, "Odds ratio for a single 2 × 2 table with correlated binomials for two margins," pp. 483–497, 2008.

[40]   T. Hesterberg, D. S. Moore, S. Monaghan, A. Clipson, and R. Epstein, "Bootstrap Methods and Permutation Tests," in *Introduction to the Practice of Statistics*, 5th ed., D. S. Moore and G. P. McCabe, Eds. New York: W.H. Freeman and company, pp. 1–70.

[41] R Development Core Team, "R: A Language and Environment for Statistical Computing," 2012. [Online]. Available: http://www.r-project.org.

[42] L. D. Lynd, C. A. Marra, M. Najafzadeh, and M. Sadatsafavi, "A quantitative evaluation of the regulatory assessment of the benefits and risks of rofecoxib relative to naproxen : an application of the incremental net-benefit framework y," no. July, pp. 1172–1180, 2010.

[43] T. Saaty and L. Vargas, *Prediction, projection and forecasting; applications of the Analytic Hierarchy Process in economics, finance, politics, games and sports*. Boston, Dordrecht, London: Kluwer Academic Publishers, 1991.

[44] F. R. Johnson, A. Brett Hauber, and C. M. Poulos, "A Brief Introduction to the Use of Stated-Choice Methods to Measure Preferences for Treatment Benefits and Risks," no. September, 2009.

[45] A. Mohamed, A. Hauber, and M. Neary, "Patient Benefit-Risk Preferences for Targeted Agents in the Treatment of Renal Cell Carcinoma," *PharmacoEconomics*, vol. 29, no. 11, pp. 977–988, 2011.

[46] a. Mohamed, J. C. Yang, a. B. Hauber, Z. Liu, M. Wong, J. Rogerio, and C. Garay, "PCN155 Patient Benefit-Risk Preferences for Advanced Renal Cell Carcinoma Treatments: Results From a Conjoint Analysis Study," *Value in Health*, vol. 15, no. 4, p. A236, Jun. 2012.

[47] J. F. P. Bridges, A. F. Mohamed, H. W. Finnern, A. Woehl, and a B. Hauber, "Patients' preferences for treatment outcomes for advanced non-small cell lung cancer: A conjoint analysis.," *Lung cancer (Amsterdam, Netherlands)*, vol. 77, no. 1, pp. 224–31, Jul. 2012.

[48] G. Lu and a E. Ades, "Combination of direct and indirect evidence in mixed treatment comparisons.," *Statistics in medicine*, vol. 23, no. 20, pp. 3105–24, Oct. 2004.

[49] G. Salanti, J. P. T. Higgins, a E. Ades, and J. P. a Ioannidis, "Evaluation of networks of randomized trials.," *Statistical methods in medical research*, vol. 17, no. 3, pp. 279–301, Jun. 2008.

[50] B. H. Hidaka, "Depression as a disease of modernity: Explanations for increasing prevalence.," *Journal of affective disorders*, vol. 140, no. 3, pp. 205–14, Nov. 2012.

[51] "Depression," 2008.

[52] T. Baghai and H. Volz, "Antidepressive pharmacotherapy - State of the art and recent developments [Antidepressive Pharmakotherapie: Aktueller Stand und neue Entwicklungen]," *Journal für neurologie, neurochirurgie und psychiatrie*, vol. 12, no. 1, pp. 70–81, 2011.

[53] "Selective serotonin and noradrenalin reuptake inhibitors (SNRI) with depression patients [Selektive Serotonin- und Wiederaufnahmehemmer (SNRI) bei Patienten mit Depressionen]," Cologne, 2010.

[54] E. Rubin, F. Gorstein, R. Rubin, R. Schwarting, and D. Strayer, *Rubin's pathology: clinicopathologic foundations of medicine*, 4th ed. Baltimore: Lippincott Williams & Wilkins, 2005, pp. 882–884.

[55]   C. Coppin, L. Le, W. TJ, and C. Kollmannsberger, "Targeted therapy for advanced renal cell carcinoma (Review)," *Cochrane systematic reviews*, no. 2, 2008.

[56]   T. Eisen, C. N. Sternberg, C. Robert, P. Mulders, L. Pyle, S. Zbinden, H. Izzedine, and B. Escudier, "Targeted therapies for renal cell carcinoma: review of adverse event management strategies.," *Journal of the National Cancer Institute*, vol. 104, no. 2, pp. 93–113, Jan. 2012.

[57]   B. Escudier, T. Eisen, W. M. Stadler, C. Szczylik, S. Oudard, M. Staehler, S. Negrier, C. Chevreau, A. a Desai, F. Rolland, T. Demkow, T. E. Hutson, M. Gore, S. Anderson, G. Hofilena, M. Shan, C. Pena, C. Lathia, and R. M. Bukowski, "Sorafenib for treatment of renal cell carcinoma: Final efficacy and safety results of the phase III treatment approaches in renal cancer global evaluation trial.," *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, vol. 27, no. 20, pp. 3312–8, Jul. 2009.

[58]   C. Coppin, C. Kollmannsberger, L. Le, F. Porzsolt, and T. J. Wilt, "Targeted therapy for advanced renal cell cancer (RCC): a Cochrane systematic review of published randomised trials.," *BJU international*, vol. 108, no. 10, pp. 1556–63, Nov. 2011.

[59]   C. N. Sternberg, I. D. Davis, J. Mardiak, C. Szczylik, E. Lee, J. Wagstaff, C. H. Barrios, P. Salman, O. a Gladkov, A. Kavina, J. J. Zarbá, M. Chen, L. McCann, L. Pandite, D. F. Roychowdhury, and R. E. Hawkins, "Pazopanib in locally advanced or metastatic renal cell carcinoma: results of a randomized phase III trial.," *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, vol. 28, no. 6, pp. 1061–8, Feb. 2010.

[60]   R. J. Motzer, B. Escudier, S. Oudard, T. E. Hutson, C. Porta, S. Bracarda, V. Grünwald, J. a Thompson, R. a Figlin, N. Hollaender, G. Urbanowitz, W. J. Berg, A. Kay, D. Lebwohl, and A. Ravaud, "Efficacy of everolimus in advanced renal cell carcinoma: a double-blind, randomised, placebo-controlled phase III trial.," *Lancet*, vol. 372, no. 9637, pp. 449–56, Aug. 2008.

[61]   R. J. Amato, A. L. Flaherty, and M. Stepankiw, "Phase I trial of everolimus plus sorafenib for patients with advanced renal cell cancer.," *Clinical genitourinary cancer*, vol. 10, no. 1, pp. 26–31, Mar. 2012.

[62]   S. Kordes, D. J. Richel, H.-J. Klümpen, M. J. Weterman, A. J. W. M. Stevens, and J. W. Wilmink, "A phase I/II, non-randomized, feasibility/safety and efficacy study of the combination of everolimus, cetuximab and capecitabine in patients with advanced pancreatic cancer.," *Investigational new drugs*, pp. 1–7, Feb. 2012.

[63]   M. Deenen, H. Klümpen, D. Richel, R. Sparidans, M. Weterman, J. Beijnen, J. Schellens, and J. Wilmink, "Phase I and pharmacokinetic study of capecitabine and the oral mTOR inhibitor everolimus in patients with advanced solid malignancies," *Investigational new drugs*, pp. 1–9, 2011.

[64]   P. Coates, "Liver function tests.," *Australian family physician*, vol. 40, no. 3, pp. 113–5, Mar. 2011.

[65]   D. Gotthardt, C. Riediger, K. H. Weiss, J. Encke, P. Schemmer, J. Schmidt, and P. Sauer, "Fulminant hepatic failure: etiology and indications for liver transplantation.," *Nephrology, dialysis,*

transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association, vol. 22 Suppl 8, pp. viii5–viii8, Sep. 2007.

[66] M. J. Ratain, T. Eisen, W. M. Stadler, K. T. Flaherty, S. B. Kaye, G. L. Rosner, M. Gore, A. a Desai, A. Patnaik, H. Q. Xiong, E. Rowinsky, J. L. Abbruzzese, C. Xia, R. Simantov, B. Schwartz, and P. J. O'Dwyer, "Phase II placebo-controlled randomized discontinuation trial of sorafenib in patients with metastatic renal cell carcinoma.," *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, vol. 24, no. 16, pp. 2505–12, Jun. 2006.

[67] K. Claxton, "Bayesian approaches to the value of information: implications for the regulation of new pharmaceuticals," *Health Economics*, vol. 274, no. December 1998, pp. 269–274, 1999.

[68] A. E. Ades, G. Lu, and K. Claxton, "Expected value of sample information calculations in medical decision modeling.," *Medical decision making : an international journal of the Society for Medical Decision Making*, vol. 24, no. 2, pp. 207–27, 2004.

Appendix A: source code antidepressants case

This code can be run with R, available from www.r-project.org/. There are four scripts below. The first one is the model itself. It is run 100,000 times to calculate the simulation points. The second script generates the risk-benefit plane and calculates acceptability chances. The third script can be used to make comparison plots between pharmaceuticals. The fourth script can be used to make sensitivity graphs. #'s indicate comment lines.

The model

```
# reset all variables
rm(list = ls())

## number of iterations that we want to run
iterations = 100000

respondents=12

# weights from source data (expert opinion elicited with ahp), normalized
weightRespons  =  c(0.566,0.368,0.374,0.784,0.445,0.781,0.777,0.707,0.721,0.624,0.525,0.667)
weightRemission = c(0.193,0.176,0.335,0.117,0.308,0.071,0.124,0.145,0.186,0.092,0.274,0.072)
weightAE =        c(0.242,0.456,0.291,0.098,0.247,0.148,0.099,0.147,0.093,0.284,0.202,0.261)

# initialize weights matrix
weights=matrix(1:(3*iterations),ncol=3)

### preference distribution method: bootstrap
# for every simulation run...
for(j in 1:iterations){

# sample 12 respondents (with replace)
sampler=sample(1:respondents,replace=TRUE)

# reset the local weight vectors
W1=1:respondents
W2=1:respondents
W3=1:respondents

# for every respondent that was sampled...
for(i in 1:respondents){
# add the weights of the respondent to the local weight vector
W1[i] = weight[sampler[i],1]
W2[i] = weight[sampler[i],2]
W3[i] = weight[sampler[i],3]
} # end inner loop (j)

# the mean of the sampled weights is calculated
```

```
W1 = sum(W1)/length(W1)
W2 = sum(W2)/length(W2)
W3 = sum(W3)/length(W3)

# the weights are output to the 'weights' vector which is multiplied in the next model step by the
performance vector to get eventual scores
weights[j,]=c(W1,W2,W3)
#weights[,1]=seq(from=0+(1/iterations),to=1,by=(1/iterations))

}# end outer loop (i)

### 3. Running the simulation
###       The simulation runs for 'iterations' times.
###       This is the general procedure per pharmaceutical per model iteration:
###       I. Get the performances by getting random numbers from the normal distribution with the
defined mean and standard deviation in the log domain, then transform this back
###       II. Multiply this performance by the weights
###       III. Sum all benefits and all risks seperately
###       IV. plot the summed benefits and summed risks in the risk-benefit plane
###       V. Calculate the acceptability chance for both pharmaceuticals and output this to the console

z = 1.96

# Duloxetine
duloxetineRespons  = weights[,1]*(exp(rnorm(iterations,log(1.95),(log(1.95)-log(1.61))/z)))
duloxetineRemission = weights[,2]*(exp(rnorm(iterations,log(1.91),(log(1.91)-log(1.56))/z)))
duloxetineBenefits = duloxetineRespons+duloxetineRemission
duloxetineAE       = weights[,3]*(exp(rnorm(iterations,log(2.22),(log(2.22)-log(1.55))/z)))
duloxetineRisks    = duloxetineAE

# Venlafaxine
venlafaxineRespons  = weights[,1]*(exp(rnorm(iterations,log(2.04),(log(2.04)-log(1.74))/z)))
venlafaxineRemission = weights[,2]*(exp(rnorm(iterations,log(1.97),(log(1.97)-log(1.64))/z)))
venlafaxineBenefits = venlafaxineRespons+venlafaxineRemission
venlafaxineAE       = weights[,3]*(exp(rnorm(iterations,log(2.47),(log(2.47)-log(1.81))/z)))
venlafaxineRisks    = venlafaxineAE

# Bupropion
bupropionRespons  = weights[,1]*(exp(rnorm(iterations,log(1.48),(log(1.48)-log(1.2))/z)))
bupropionRemission = weights[,2]*(exp(rnorm(iterations,log(1.46),(log(1.46)-log(1.18))/z)))
bupropionBenefits = bupropionRespons+bupropionRemission
bupropionAE       = weights[,3]*(exp(rnorm(iterations,log(1),(log(1)-log(0.61))/z)))
bupropionRisks    = bupropionAE
Risk-benefit plane and acceptabilities
# define the threshold and make the function for the risk-benefit
# threshold line so it can be plotted
threshold=1
rbtx=0:iterations
```

```
rbty=rbtx/threshold

benefitAxis =
c(min(0,duloxetineBenefits,venlafaxineBenefits,bupropionBenefits),max(duloxetineBenefits,venlafaxineB
enefits,bupropionBenefits))
riskAxis =
c(min(0,duloxetineRisks,venlafaxineRisks,bupropionRisks),max(duloxetineRisks,venlafaxineRisks,bupropi
onRisks))

# plot the performances of both pharmaceuticals
plotSize=0.4
plot(duloxetineBenefits,duloxetineRisks, xlab="Benefits (weighted OR compared to placebo)",ylab="Risks
(weighted OR compared to placebo)", type="p", xlim=benefitAxis, ylim=riskAxis,
cex=plotSize,col=rgb(0,0,1))
points(venlafaxineBenefits,venlafaxineRisks,cex=plotSize,col=rgb(1,0,0))
points(bupropionBenefits,bupropionRisks,cex=plotSize,col=rgb(0,1,0))
lines(rbtx, rbty)

## acceptability chances
# duloxetine
duloxetine_brr=duloxetineBenefits/duloxetineRisks
duloxetine_acceptability=sum(duloxetine_brr>threshold)/sum(duloxetine_brr!=0)

# venlafaxine
venlafaxine_brr=venlafaxineBenefits/venlafaxineRisks
venlafaxine_acceptability=sum(venlafaxine_brr>threshold)/sum(venlafaxine_brr!=0)

# bupropion
bupropion_brr=bupropionBenefits/bupropionRisks
bupropion_acceptability=sum(bupropion_brr>threshold)/sum(bupropion_brr!=0)
```

Pharmaceutical to pharmaceutical comparison

```
### head to head comparisons ###
# create a 1 row,3 column graphical window
par(mfcol=c(1,3))

# create the histograms and the C=1 lines
hist(duloxetine_brr/venlafaxine_brr,25,xlab="C=Duloxetine BRR/Venlafaxine BRR",main="Duloxetine vs
Venlafaxine")
lines(c(1,1),c(0,1000))
hist(duloxetine_brr/bupropion_brr,25,xlab="C=Duloxetine BRR/Bupropion BRR",main="Duloxetine vs
Bupropion")
lines(c(1,1),c(0,1000))
hist(venlafaxine_brr/bupropion_brr,25,xlab="C=Venlafaxine BRR/Bupropion BRR",main="Venlafaxine vs
Bupropion")
lines(c(1,1),c(0,1000))

# chances that a particular pharmaceutical outperforms another pharmaceutical. The name of the
```

```
# variable indicates which chance it is. The syntax is XBX, where X is a pharmaceutical.
# D=duloxetine, V=Venlafaxine, B=Bupropion
chanceDBV=sum(((duloxetine_brr/venlafaxine_brr)>1)/iterations)
chanceVBD=sum(((venlafaxine_brr/duloxetine_brr)>1)/iterations)
chanceDBB=sum(((duloxetine_brr/bupropion_brr)>1)/iterations)
chanceBBD=sum(((bupropion_brr/duloxetine_brr)>1)/iterations)
chanceBBV=sum(((bupropion_brr/venlafaxine_brr)>1)/iterations)
chanceVBB=sum(((venlafaxine_brr/bupropion_brr)>1)/iterations)

compare=matrix(c(0,chanceDBV,chanceDBB,chanceVBD,0,chanceVBB,chanceBBD,chanceBBV,0),ncol=3)
```

Sensitivity graphs

```
## Sensitivity graph maker ##
x=weights[,1] # we will check how sensitive parameter y is to this parameter
xname="Response weight" #name of the parameter
y=duloxetine_brr # change this to the parameter whose sensitivity is under analysis
yname="Duloxetine BRR"
colour="blue"

plot(x,y,cex=0.1,col=colour,xlab=xname, ylab=yname) # the plot
lines(c(quantile(x,.025),quantile(x,.025)),c(0,100)) # vertical line LL
lines(c(quantile(x,.975),quantile(x,.975)),c(0,100)) # vertical line UL
lines(c(0,1),c(approxfun(x,y)(quantile(x,.025)),approxfun(x,y)(quantile(x,.025)))) # y value at LL
lines(c(0,1),c(approxfun(x,y)(quantile(x,.975)),approxfun(x,y)(quantile(x,.975)))) # y value at UL
```

Appendix B: sourcecode renal cancer case

The code for the second case differs considerably from the first. The model does not change, but the code is structured in a different way. This is done primary by making the code primarily matrix based instead of variable based. The new structure allows for greater flexibility of the model.

As in the antidepressants case code, #'s indicate comment lines.

```
############################################
##  MCDA for risk-benefit assessments        ##
##        by Henk Broekhuizen                 ##
##  h.broekhuizen@student.utwente.nl          ##
##    RCC medication case model              ##
##    requires gregmisc package               ##
############################################

## model parameters
t=100000 # amount of simulation runs

## preference matrix, which is a 2D matrix:
## t\W w1  w2   w3   (...)
## 1   v11 v21  v31
## 2   v12 v22  v32
##(...)
## Here, v is a value that is drawn from the distribution of
## a particular weight. The matrix expands in the t direction
## for every added iteration.
preference=matrix(c(
#1 PFS 2 months
rnorm(t,0,0),
#2 PFS 12 months
rnorm(t,3.81,.34),
#3 PFS 24 months
rnorm(t,10,.85),
#4 Fatigue
rnorm(t,6.19,.95),
#5 Diarrhea
rnorm(t,5.28,.75),
#6 Mouth sores
rnorm(t,2.56,.79),
#7 Hand-foot syndrome
```

```
rnorm(t,4.69,.83),
#8 Liver failure (raised serum ALT)
rnorm(t,5.17,.76),
#9 Blood clot (cardiovascular risk)
rnorm(t,3.39,.61)),
ncol=9)


## Performance matrix, which is a 2D matrix:
## t\p p1  p2  p3 (...)
## 1   p11 p21 p31
## 2   p12 p22 p32
##(...)
## Here, p is a particular performance parameter, for example:
## "Pazopanib's performance on the mouth sores criterion"


# this next small amount of code is for Pazopanib > Hand-foot syndrome,
# for more information see the results section of the thesis.
ei=runif(t,1,28) # events intervention group
nei=290-ei # non events intervention group
ic=18.325 # events control group
nic=145-ic # non events control group


performance=matrix(c(


## Sorafenib after cytokines ##
#1 Sorafenib low baseline VEGF > PFS 2 months
exp(rnorm(t,.231,.162)),
#2 Sorafenib low baseline VEGF > PFS 12 months
exp(rnorm(t,.236,.289)),
#3 Sorafenib low baseline VEGF > PFS 24 months
rnorm(t,1,0),
#4 Sorafenib high baseline VEGF > PFS 2 months
exp(rnorm(t,.410,.166)),
#5 Sorafenib high baseline VEGF > PFS 12 months
exp(rnorm(t,.955,.404)),
#6 Sorafenib high baseline VEGF > PFS 24 months
rnorm(t,1,0),
#7 Sorafenib > Fatigue
exp(rnorm(t,.584,.160)),
#8 Sorafenib > Diarrhea
exp(rnorm(t,1.481,.172)),
#9 Sorafenib > Mouth sores
```

exp(rnorm(t,1.054,.416)),
#10 Sorafenib > Hand-foot syndrome
exp(rnorm(t,.972,.169)),
#11 Sorafenib > Liver failure (ALT)
rnorm(t,1,0),
#12 Sorafenib > Blood clot (cardiovascular risk)
exp(rnorm(t,1.096,.819)),

Pazopanib after nil or IFNalpha
#13 Pazopanib > PFS 2 months
exp(rnorm(t,.260,.156)),
#14 Pazopanib > PFS 12 months
exp(rnorm(t,1.010,0.241)),
#15 Pazopanib > PFS 24 months
rnorm(t,1,0),
#16 Pazopanib > Fatigue
exp(rnorm(t,.916,.346)),
#17 Pazopanib > Diarrhea
exp(rnorm(t,1.752,.306)),
#18 Pazopanib > Mouth sores
rnorm(t,1,0),
#19 Pazopanib > Hand-foot syndrome
exp(rnorm(t,(ei/nei)/(ic/nic),sqrt(1/ei+1/nei+1/ic+1/nic))),
#20 Pazopanib > Liver failure (ALT)
exp(rnorm(t,.865,.219)),
#21 Pazopanib > Blood clot (cardiovascular risk)
exp(rnorm(t,.108,.586)),

Everolimus after VEGFR inhibitor therapy
#22 Everolimus > PFS 2 months
exp(rnorm(t,.739,.223)),
#23 Everolimus > PFS 12 months
rnorm(t,1,0),
#24 Everolimus > PFS 24 months
rnorm(t,1,0),
#25 Everolimus > Fatigue
exp(rnorm(t,.201,.274)),
#26 Everolimus > Diarrhea
exp(rnorm(t,1.763,.532)),
#27 Everolimus > Mouth sores
rnorm(t,1,0),
#28 Everolimus > Hand-foot syndrome

```
rnorm(t,1,0),
#29 Everolimus > Liver failure (ALT)
exp(rnorm(t,1.583,.481)),
#30 Everolimus > Bloot clot (cardiovascular risk)
rnorm(t,1,0)),
ncol=30)
```

```
## Benefits, which is a 2D matrix
## t\B B1  B2  B3 (...)
## 1   B11 B21 B31
## 2   B12 B22 B32
##(...)
## Here, Bij is the summed and weighted benefit of pharmaceutical
## i in simulation run j.
benefits=matrix(c(
#1 Sorafenib low baseline VEGF
performance[,1]*preference[,1]+
performance[,2]*preference[,2]+
performance[,3]*preference[,3],
#2 Sorafenib high baseline VEGF
performance[,4]*preference[,1]+
performance[,5]*preference[,2]+
performance[,6]*preference[,3],
#3 Pazopanib
performance[,13]*preference[,1]+
performance[,14]*preference[,2]+
performance[,15]*preference[,3],
#4 Everolimus
performance[,22]*preference[,1]+
performance[,23]*preference[,2]+
performance[,24]*preference[,3]),
ncol=4)
```

```
## Risks, which is a 2D matrix
## t\R R1  R2  R3 (...)
## 1   r11 r21 r31
## 2   r12 r22 r32
##(...)
## Here, rij is the summed and weighted risk of pharmaceutical
## i in simulation run j.
risks=matrix(c(
#1 Sorafenib low baseline VEGF (equal to high baseline risks)
```

```
preference[,4]*performance[,7]+
preference[,5]*performance[,8]+
preference[,6]*performance[,9]+
preference[,7]*performance[,10]+
preference[,8]*performance[,11]+
preference[,9]*performance[,12],
#2 Sorafenib high baseline VEGF (equal to low baseline risks)
preference[,4]*performance[,7]+
preference[,5]*performance[,8]+
preference[,6]*performance[,9]+
preference[,7]*performance[,10]+
preference[,8]*performance[,11]+
preference[,9]*performance[,12],
#3 Pazopanib
preference[,4]*performance[,16]+
preference[,5]*performance[,17]+
preference[,6]*performance[,18]+
preference[,7]*performance[,19]+
preference[,8]*performance[,20]+
preference[,9]*performance[,21],
#4 Everolimus
preference[,4]*performance[,25]+
preference[,5]*performance[,26]+
preference[,6]*performance[,27]+
preference[,7]*performance[,28]+
preference[,8]*performance[,29]+
preference[,9]*performance[,30]),
ncol=4)

## outcome measures
## risk-benefit plane
size=.75
threshold=1 # mu
plot(benefits[,1],risks[,1],xlim=c(min(benefits),max(benefits)),ylim=c(min(risks),max(risks)),xlab="Weight
ed benefits (odds ratio compared to plabebo)",ylab="Weighted risks (odds ratio compared to
placebo)",col="red",cex=size)
points(benefits[,2],risks[,2],col="orange",cex=size)
points(benefits[,3],risks[,3],col="blue",cex=size)
points(benefits[,4],risks[,4],col="green",cex=size)
rbtx=0:t
rbty=rbtx/threshold
lines(rbtx, rbty)
```

```
## acceptability
acceptability=matrix(c(
#1 Sorafenib low baseline VEGF
sum(benefits[,1]/risks[,1]>threshold)/t,
#2 Sorafenib high baseline VEGF
sum(benefits[,2]/risks[,2]>threshold)/t,
#3 Pazopanib
sum(benefits[,3]/risks[,3]>threshold)/t,
#4 Everolimus
sum(benefits[,4]/risks[,4]>threshold)/t),
ncol=4)

## comparisons between pharmaceuticals
comparisons=matrix(c(
#1 Sorafenib low baseline VEGF vs Sorafenib high baseline VEGF
(benefits[,1]/risks[,1])/(benefits[,2]/risks[,2]),
#2 Sorafenib low baseline VEGF vs Pazopanib
(benefits[,1]/risks[,1])/(benefits[,3]/risks[,3]),
#3 Sorafenib low baseline VEGF vs Everolimus
(benefits[,1]/risks[,1])/(benefits[,4]/risks[,4]),
#4 Sorafenib high baseline VEGF vs Pazopanib
(benefits[,2]/risks[,2])/(benefits[,3]/risks[,3]),
#5 Sorafenib high baseline VEGF vs Everolimus
(benefits[,2]/risks[,2])/(benefits[,4]/risks[,4]),
#6 Pazopanib vs Everolimus
(benefits[,3]/risks[,3])/(benefits[,4]/risks[,4])),
ncol=6)

# this makes a 3 by 2 panel graph showing the histograms of all C's
par(mfcol=c(3,2))
hist(comparisons[,1],main="Sorafenib low vs Sorafenib high",xlab="C")
hist(comparisons[,2],main="Sorafenib low vs Pazopanib",xlab="C")
hist(comparisons[,3],main="Sorafenib low vs Everolimus",xlab="C")
hist(comparisons[,4],main="Sorafenib high vs Pazopanib",xlab="C")
hist(comparisons[,5],main="Sorafenib high vs Everolimus",xlab="C")
hist(comparisons[,6],main="Pazopanib vs Everolimus",xlab="C")
par(mfcol=c(1,1))

# this creates a table with probabilities of outperformance
dominate=matrix(1:16,ncol=4)
for(i in 1:4){
```

```
for(j in 1:4){
dominate[i,j]=sum(((((benefits[,j]/risks[,j])/(benefits[,i]/risks[,i]))>1)/t)
}
}
```

```
# outputs the acceptability matrix and outperformance matrix to the console
acceptability
dominate
```