Tommi Tervonen, Huseyin Naci, Gert van Valkenhoef, Anthony E. Ades, Aris Angelis, Hans L. Hillege and Douwe Postmus

Applying multiple criteria decision analysis to comparative benefit-risk assessment: choosing among statins in primary prevention

Article (Accepted version) (Refereed)

Original citation:
Tervonen, Tommi, Naci, Huseyin, van Valkenhoef, Gert, Ades, Anthony E. , Angelis, Aris, Hillege, Hans L. and Postmus, Douwe (2015) Applying multiple criteria decision analysis to comparative benefit-risk assessment: choosing among statins in primary prevention. Medical Decision Making . ISSN 1552-681X
DOI: 10.1177/0272989X15587005

Applying Multiple Criteria Decision Analysis (MCDA) to Comparative Benefit-Risk Assessment: Choosing Among Statins in Primary Prevention

Authors: Tommi Tervonen PhD (1, *), Huseyin Naci PhD (2), Gert van Valkenhoef PhD (3), Anthony E. Ades PhD (4), Aris Angelis MSc (2), Hans L. Hillege PhD (3), Douwe Postmus PhD (3)

(1) Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, The Netherlands

(2) LSE Health, London School of Economics and Political Science, London, UK

(3) Department of Epidemiology, University Medical Center Groningen, University of Groningen, The Netherlands

(4) School of Social and Community Medicine, University of Bristol, UK

(*) Corresponding author. Email: tommi@smaa.fi

Word count: 5402

Abstract

Decision makers in different health care settings need to weigh the benefits and harms of alternative treatment strategies. Such health care decisions include marketing authorization by regulatory agencies, practice guideline formulation by clinical groups, and treatment selection by prescribers and patients in clinical practice. Multiple criteria decision analysis (MCDA) is a family of formal methods that help make explicit the trade-offs decision makers accept between the benefit and risk outcomes of different treatment options. Despite the recent interest in MCDA, certain methodological aspects are poorly understood. This paper presents seven guidelines for applying MCDA in benefit-risk assessment, and illustrates their use in the selection of a statin drug for the primary prevention of cardiovascular disease. We provide guidance on the key methodological issues of how to define the decision problem, how to select a set of non-overlapping evaluation criteria, how to synthesize and summarize the evidence, how to translate relative measures to absolute ones that permit comparisons between the criteria, how to define suitable scale ranges, how to elicit partial preference information from the decision makers, and how to incorporate uncertainty in the analysis. Our example on statins indicates that fluvastatin is likely to be the most preferred drug by our decision maker, and that this result is insensitive to the amount of preference information incorporated in the analysis.

Keywords: Decision aids; Multi-attribute utility function; Decision analysis

1 Introduction

Many decisions in health care involve assessing the balance of favorable and unfavorable effects of alternative treatment regimens, taking into account the associated uncertainties. For example, to choose among alternative treatment options, doctors and patients need comparative evidence to assess whether a new compound is expected to have a more favorable benefit-risk profile than existing alternatives. While subjectivity in the assessment of the benefit-risk balance of alternative treatments cannot be avoided, the decision making process itself can be made more transparent by describing the underlying value judgments in a formal and consistent manner.

A recent systematic review (1) identified Multi-Criteria Decision Analysis (MCDA) and Stochastic Multicriteria Acceptability Analysis (SMAA) to be among the most promising methods for conducting a quantitative benefit-risk assessment. MCDA provides a framework for systematic and replicable analyses of complex decision problems involving value trade-offs (2). MCDA based on multi-attribute value- or utility models has been proposed for use in benefit-risk assessments (3-5). SMAA allows applying these models in cases where exact information on the decision maker preferences is not available. Although previous studies have demonstrated applicability of these approaches in relative treatment effect assessment (2, 6), certain methodological aspects are still poorly understood in the health care research community. The structured process of arriving at a multi-criteria benefit-risk decision regarding a particular medication is not trivial, and various potential pitfalls lie on the analyst's path.

To enable wider application of MCDA in drug benefit-risk assessment, this paper provides guidance on seven important phases of the assessment process: how to define the decision problem, how to select a set of non-overlapping evaluation criteria, how to synthesize and summarize the available data, how to translate relative measures obtained through evidence synthesis to absolute scales that permit comparisons between the criteria, how to define suitable scale ranges, how to elicit preference information, and how to incorporate uncertainty into the analysis. Using a running example on a widely used class of cholesterol-lowering drugs, statins, we make recommendations about potential ways to address these key methodological challenges.

2 Data

Our running example considers a set of six statins for which there was evidence available from three recently conducted systematic reviews (7-9): atorvastatin, fluvastatin, lovastatin, pravastatin, rosuvastatin, and simvastatin. The focus of the first study was on determining the comparative tolerability and harms of the individual statins, and it included data on the number of participants who experienced myalgia, elevations in hepatic transaminases, elevations in creatine kinase (CK), and discontinuations because of adverse events. The second and third studies focused on assessing the comparative benefits of statins and included data on all-cause mortality and major coronary events and on major cerebrovascular events, respectively. All open-label and double-blind randomized, controlled trials comparing one statin with another at any dose or with control that had more than 50 participants per trial arm, lasted longer than four weeks, and reported any of the outcomes of interest were eligible for inclusion. In total, this resulted in the inclusion of 184 randomized controlled trials with 260,630 individuals with or without cardiovascular disease at baseline.

3 Guidelines for MCDA of comparative benefit-risk assessment

MCDA can potentially be useful for supporting different decisions. These include regulatory decisions at the market entry level, development of clinical practice guidelines when it is imperative to recommend a specific treatment option to initiate prescription drug therapy, and prescribing decisions in clinical practice. Although different in nature, all of these problems concern choosing among multiple treatment alternatives, and therefore the evidence concerning the beneficial and harmful effects of the available drug options need to be evaluated. The general guidelines we describe are therefore relevant for a broad spectrum of decision problems. An overview of the guidelines is presented in Table 1.

<< TABLE 1: guidelines approx here >>

3.1 Define the decision problem

The first step in the benefit-risk assessment of a prescription drug is to explicitly define the decision problem. This involves specifying, amongst others, the indication for which the assessment is conducted, the alternative treatments under consideration, the criteria on which the different treatments are to be evaluated, and the decision maker(s). The indication considered for our case study was the use of statins to reduce cardiovascular disease risk in patients who have one or more elevated cardiovascular risk factors, who are free of diabetes, and who do not have a history of cardiovascular disease, i.e., primary prevention. The alternatives under consideration were the six different statins included in our dataset, and the criteria were defined based on the clinical endpoints contained in this dataset. Our decision maker is a clinical expert from the domain of cardiovascular diseases.

We specifically focus on the comparative benefit-risk assessment of statins and the difficulty in choosing a first-line treatment out of the six currently available options. Our case study on statins is important in a number of ways. First, statins are among the most widely prescribed and used medications around the world. The recent clinical practice guidelines in the United States and the United Kingdom considerably expanded the scope and intensity of statin therapy for a broader population of individuals with or at risk of developing coronary heart disease. Second, the six statins currently on the market differ in terms of their benefit and harm profiles. So far, clinical practice guidelines have not considered the important differences among the six statins. Third, statins constitute a case in which the selection among alternative treatments is primarily a clinical one: by 2016, all six statins will be available in generic formulations, making cost considerations largely irrelevant.

3.2 Choose a set of non-overlapping evaluation criteria

When selecting the evaluation criteria, it is important to avoid overlaps as much as possible. In general, one can therefore not simply conduct the benefit-risk assessment based on all the available study endpoints as this is prone to result in an over representation of certain health effects. For example, while the change in HbA1c and the change in fasting plasma glucose are two well-established endpoints in clinical trials related to treatment with glucose-lowering products, they both serve as surrogate endpoints that measure how well the patients have, on average, responded to the investigated treatments. Only one of these endpoints should therefore be included in a multi-criteria decision model, especially if this model is of an additive structure. Similarly, it may happen that the same clinical events get counted multiple times. In such situations, the data set is ideally restructured such that overlap in the definition of the endpoints is avoided. Where this is not possible, the decision maker should select a non-overlapping subset of these endpoints.

When there are multiple decision makers, they might disagree on which set of non-overlapping endpoints is most relevant for the benefit-risk assessment. In such cases, either a consensus on the model structure should be reached through discussions, or if no consensus can be reached, multiple analyses with different endpoints must be performed. Yet another option is to keep all (overlapping) endpoints, but this would require using a more complex, non-additive multi-criteria decision model, which includes additional preference parameters whose elicitation is out of scope of this paper.

Most of the discontinuation events in our data set are likely to be due to myalgia or transaminase elevation. Therefore, when applied to our statins case study, the principle of removing overlapping criteria implies that either discontinuation or myalgia and transaminase should be excluded from the set of decision criteria. Our decision maker chose to include

myalgia and transaminase elevation and to exclude discontinuation. Additionally, if we were to simultaneously consider all-cause mortality, risk of stroke, and risk of myocardial infarction, the occurrence of a fatal stroke or a fatal myocardial infarction would be counted towards two of these criteria. The main reason for initiating statin treatment is to reduce an individual's risk of experiencing a cardiovascular event. One way to resolve this problem would therefore be to include the risks of stroke and myocardial infarction and exclude all-cause mortality. A downside of this approach is that any beneficial or harmful effects that statin treatment may have on non-cardiovascular related mortality would then no longer be captured. Alternatively, one could further refine the available data by decomposing stroke and myocardial infarction into fatal and non-fatal and then including all-cause mortality, non-fatal stroke, and non-fatal myocardial infarction as the beneficial effects. The advantage of this approach is that it captures all the relevant clinical endpoints. A potential drawback is however that those clinical trials in which fatal and non-fatal events are not clearly differentiated can no longer be included. Because we had sufficient data available, we chose this latter approach in our statin selection example. The following six criteria were therefore included in our analysis: all-cause mortality, non-fatal stroke, non-fatal myocardial infarction, myalgia, transaminase, and CK elevation.

Criteria that represent the same health effects must necessarily be highly correlated. However, this does not mean that including two correlated criteria always results in overlap. For example, the effect of a blood-thinning agent on the prevention of thrombosis is strongly correlated to the risk of it causing bleeding events. However, despite this correlation, both are separate events and should be represented as distinct criteria in the MCDA model. Data permitting, this correlation can be taken into account in the analysis (see 10, 11).

3.3 Synthesize and summarize the available data

The next step is to numerically assess the performance of the treatments on the selected benefit and risk criteria. For some criteria, there could be only one clinical study available from which the estimates of absolute treatment effects (e.g. incidence rates) can be obtained directly (2). For other criteria, there may be multiple studies available, meaning that some form of evidence synthesis is required before one is able to express a treatment's performance on these criteria in terms of a single numerical value (with associated credible intervals). For criteria measurements that are derived from randomized controlled trials, the use of network meta-analysis is now commonplace (12-15). As this approach utilizes both direct and indirect comparisons when estimating differences in the performance between the considered treatments, it is not required to restrict the analysis to only those studies that have a chosen common comparator, which would be required when applying traditional pairwise meta-analyses.

The clinical trials included in our dataset cover multiple indications, including primary prevention, secondary prevention, diabetes management (among individuals with or at risk of developing coronary heart disease), and treatment of acute coronary syndrome. As both the relative and absolute reduction in cardiovascular disease risk associated with the use of statins can vary depending on the patient population considered, it is important that all the studies included in the benefit-risk assessment fit the indication for which a decision has to be taken. Previous network meta-analyses indicated no significant differences in the estimated relative effects with different subgroups (7-9, 16). For this reason, and to increase accuracy of the estimates, we chose to estimate relative effects for the four chosen decision criteria using data from all available 184 studies. Figure 1 presents the network structure for studies that measured all-cause mortality events. Odds ratios estimated for the chosen six decision criteria are presented in Figure 2.

<< FIGURE 1: network approx here >>

<< FIGURE 2: odds approx here >>

3.4 Translate relative measures obtained in evidence synthesis to absolute scales that permit comparisons between the criteria

Network meta-analyses produce relative effect estimates, which are unsuitable for specifying value trade-offs, as they do not contain information on the baseline effect. To illustrate why, consider two alternatives that are evaluated in terms of two criteria. Suppose that the relative risk of alternative 1 against alternative 2 is 1.5 for criterion A and 0.8 for criterion B. While on a relative scale, the difference between the two treatments is much larger for criterion A than for criterion B, these differences are impossible to interpret without considering the baseline effect of treatment 2. For example, suppose that the baseline effect of alternative 2 is 2% on criterion A and 50% on criterion B. For alternative 1, the previously reported relative risks then translate into an absolute risk of 3% on criterion A and an absolute risk of 40% on criterion B. Depending on what these criteria entail, a 10% difference in absolute risk on criterion B may be far more important than a 1% difference in absolute risk on criterion A, showing that the results from a network meta-analysis first need to be translated to values measured on absolute scales before one is able to make value trade-offs in a meaningful way.

The results of our previously conducted network meta-analyses on statins were presented on the odds ratio scale, which suffers from the same problem illustrated above for the risk ratio. In both cases, the solution is the same: translate the relative effects to an absolute scale using an estimate of the absolute effect for a suitably selected baseline treatment (6, 17). The absolute effect estimates can be from randomized trials or observational studies, and the baseline treatment can either be a placebo or an active treatment. What matters is that the studies included in the estimation of the baseline effects are representative of the target population and that the mean follow-up of each of these studies is similar so that the event rates observed in these studies are comparable. If no suitable data are available, estimations of the baseline effects may need to be elicited from expert clinicians.

We estimated baseline effects in our case study for both the hard clinical outcomes (nonfatal strokes, nonfatal MIs, and all-cause mortality) and side effects (myalgia, transaminase, CK elevation). For each side effect, we estimated the baseline effect using a Bayesian random-effects pooling of the event rates in the placebo arms across all trials. We specified an informative prior (a half-normal: $|N(0, 0.25)|$) for the heterogeneity standard deviation to ensure it could be estimated. The absolute effect of the placebo intervention was then defined as the resulting predictive distribution. The predictive distribution incorporates both the uncertainty around the mean and the between-studies heterogeneity, and thereby fully accounts for heterogeneity in the observed effects. Baseline risks of the hard clinical outcomes were estimated using a single large study representative of the target population. The study results were then used to model the Beta distributed baseline effects using a Bayesian approach with a flat Beta(1,1) prior, following Tervonen et al. (2). We chose ALLHAT-LLT as the largest non-industry sponsored study corresponding to the primary prevention population (18). According to the American College of Cardiology risk calculator, the trial population of ALLHAT-LLT has on average a 21.40% 10-year risk for atherosclerotic cardiovascular disease, defined as coronary death or nonfatal myocardial infarction, or fatal or nonfatal stroke (http://tools.cardiosource.org/ASCVD-Risk-Estimator/).

Estimates of the absolute effects for the 6 statins were subsequently obtained by combining the absolute effect of placebo with the relative effects obtained from the network meta-analyses. This is achieved by sampling from both the estimated distribution for the baseline risk and the distribution for the log odds ratios. For a given baseline risk p_A and a

log-odds ratio for treatment B of d_{AB}, the absolute risk for treatment B is given by $p_B = \text{logit}^{-1}(\text{logit}(p_A) + d_{AB})$. The absolute effects are illustrated for the chosen 6 criteria in Figure 3. When compared to Figure 2, the different rate of side effects appears to be much less pronounced because all absolute effect estimates incorporate uncertainty around the highly uncertain baseline effect. However, the occurence of side effects is strongly correlated through this common baseline, and this is taken into account in the decision analysis, so that the true treatment differences are preserved.

<< FIGURE 3: absolute risks approx here >>

3.5 Define suitable scale ranges

In practical applications of MCDA, the constructed multi-criteria model is often assumed to be of an additive structure, which is illustrated in Figure 4. Then the problem of formally representing the decision maker's preferences reduces to the problem of specifying a set of partial value functions that reflect the relative desirability of decision criteria levels (e.g. increase of a side effect from 0% to 5% vs. from 5% to 10%), and a set of weights that reflect the relative importance of worst-best scale increases across the criteria ranges. In order to contextualize the decision and make subsequent weight elicitation meaningful, the criteria scale ranges should be defined with respect to plausible outcome ranges. That is, if mortality over the set of considered treatments varies only within 2–6%, the partial value functions should be defined for this range instead of e.g. 0–100%. Although the partial values can always be interpolated within the range, using ranges irrelevant for the decision context causes the preferences to be captured with a lower accuracy.

<< FIGURE 4: additive model approx here >>

The challenge in defining suitable scale ranges relates to uncertainty of the measurements. We have previously (2) suggested to define the worst-best scaling based on interval hulls of the per-criterion 95% credible intervals from the absolute scale joint distribution. For example, in case of all-cause mortality, this is 0.05-0.15, as defined by the distributions of fluvastatin and lovastatin (Figure 3, mid-left panel). Although such ranges capture most of the variance, they might be inappropriate with long-tailed measurement distributions. Table 2 illustrates this with respect to the statins case; the span of 95% empirical credible intervals for transaminase (0.01-0.31) is considerably smaller than the full sample range (0.00-0.86). However, as using too large scale ranges causes imprecision for the preference elicitation, we do not currently have a better recommendation than using the 95% ranges.

<< TABLE 2: quantiles approx here >>

3.6 Elicit preference information

After the scale ranges have been defined, additive value models incorporate decision maker preferences by eliciting partial value functions and their scaling factors (weights).

Partial value functions

Partial value functions reflect the desirability of scale values within individual criteria. Eliciting preferences over scale values on continuous criteria can be done with the bisection method (19). For example, if an outcome of interest is the risk of stroke and its worst and best levels are 6% and 2%, respectively, then the first step in the bisection method would be to ask for the value of x such that a decrease from 6% to x% is as important as a decrease from x% to 2%. If the decision maker replies by stating that x equals 4, the partial value function for stroke is likely to be linear. Linear

partial value functions are often appropriate for criteria that measure event rates in a defined patient population, as in these cases equal size ranges in percentages (e.g. 5-7% or 90-92%) reflect the same number of affected patients. Non-linear partial value functions, in contrast, reflect situations where the value associated with a fixed performance increment depends on the level of achievement on a criterion. For example, suppose that the amount of toxicity that a decision maker is willing to accept for a 6-month increase in mean survival time is larger for an increase from 6 months to 1 year (e.g., patients who underwent prior treatment) than for an increase from 2.5 years to 3 years (e.g., treatment-naïve patients). The partial value function for overall survival would then be concave (i.e., a function whose slope decreases as the level of performance increases). Partial value functions that are convex or S-shaped are also possible.

The lower value function in Figure 4 illustrates a hypothetical case of bisection elicitation where the criterion measures weight loss and the observed losses vary from 0% to 15%. Let us suppose that the first answer of the clinical expert is that half of the benefits are obtained at 10% weight loss, which is normally used as a threshold for clinically significant effect. Then, the analyst could ask again what is the half-point of effects between 0% and 10%, and that the expert states 7%. This leads to the partial value function given in the figure, which can then be used for calculating the alternatives' partial values, and once the weights are known, to rank the alternative treatments. The bisection procedure in principle is continued infinitely, but usually a few answers provide a good approximation of the “true” partial value function.

Weight information

Weights of the additive model express accepted trade-offs over the criteria scale swings. For example, if the scale of mortality is [6%, 2%] and the scale of discontinuation is [50%, 10%], then if mortality has weight 100 and discontinuation weight 1, the increase of mortality from 2% to 6% is considered one hundred times worse than the increase of discontinuation from 10% to 50%. Note that this is the only meaning of the weights - they do not express any kind of absolute importance. Therefore, elicitation questions such as 'what is more important: all-cause mortality or myalgia?' are meaningless when dealing with additive value models.

In the above example, the criteria measurements are expressed as incidence rates and the partial value functions are taken to be linear, and therefore it may seem reasonable to trade off one event against another, rather than compare the incidence ranges. However, doing so depends heavily on the linearity assumption - i.e. that a change on a single criterion from 2% to 1% is equally valuable to a change from 100% to 99%. This may decrease the precision as well as value of the elicited weight information because it would become less specific to the problem at hand. The importance of taking scale ranges into account is clearer when the decision involves outcomes such as blood pressure lowering in the treatment of hypertension. For example, the decision maker may face the dilemma of whether the difference in blood pressure lowering on the scale [-5, -15] mmHg outweighs the occurrence of serious adverse events on the scale [2%, 0%].

Weights can be elicited with the swing method, in which the decision maker is asked to judge the relative importance of the worst-best scale swings (as described above). However, all elicitation techniques resulting in exact weights are subject to behavioral biases (20). Therefore many modern MCDA approaches allow incorporating weight information in imprecise or incomplete formats. Imprecise information can be modeled, instead of point estimates, as intervals for the trade-off ratios. For example, instead of trade-off ratio of 2, the decision maker could express imprecision with the ratio belonging to the interval [1.5, 2.5]. Incomplete information expresses exact but poor information, similarly to pair-

wise choices in conjoint analysis. The simplest form of incomplete weight information is (partial) ranking of scale swings (ordinal information).

Both incomplete and imprecise statements define linear constraints on the set of feasible weights, where the weights are always non-negative and normalized to sum to a constant, usually to unity. For example, if the decision maker provides exact weight information in a two-criterion problem, stating that the scale swing of the first criterion is twice as important as that of the second criterion, this would result in a single normalized weight vector [2/3, 1/3]. If instead the decision maker provides imprecise information stating that the trade-off ratio belongs to the interval [1, 2], this results in a feasible weight space where the first weight is bound within interval [1/(1+1), 2/(2+1)] = [0.5, 2/3], and the second weight is one minus the first weight (and bound within [1/3, 0.5]). If the decision maker provides only incomplete (ordinal) information stating that the scale swing of the first criterion is more important than that of the second criterion, this restricts feasible weights so, that $w_1 > w_2$, and the resulting range for the first weight is (0.5, 1.0].

3.7 Incorporate uncertainty in the analysis

The two main sources of uncertainty in benefit-risk assessment are the uncertain outcome estimates due to limited sample sizes in clinical trials, and imprecise or incomplete weight information. To propagate the uncertainty in these inputs into uncertainty in the ranking of the treatments, we have previously proposed (2, 6) to apply Stochastic Multicriteria Acceptability Analysis (SMAA) (21, 22). This entails sampling a sufficient amount of observations (23) from the measurement distributions, and for each of these, sampling a weight vector from a uniform distribution within the feasible weight space (24, 25). In each of the Monte Carlo iterations, the alternatives are ranked from best to worst according to their total value, which is computed using the realized values of the weights and measurement outcomes (see Figure 4). Different realizations of the weights and measurement outcomes may translate into a different ranking of the treatments. In SMAA, this uncertainty is captured by computing the rank acceptability indices, which describe, for all possible combinations of ranks and treatments, the fraction of Monte Carlo iterations for which a treatment is ranked at a certain position.

Sampling weights uniformly from the feasible weight space specified with trade-off intervals is not trivial, especially if there are many decision criteria. We recommend using the Hit-And-Run sampler, which is an efficient Markov Chain Monte Carlo technique (24, 25). There exists an open source R package 'hitandrun' that implements the sampling method and another package 'smaa' for computing the SMAA rank acceptability indices. Both are freely available at the CRAN repository.

The choice of criterion scales has an effect on the simulation technique. For example, in our statins case, the 95% credible interval hull for transaminase is 1-31%, but samples from the pooled distribution span the range 0-86%. Therefore, if the partial value function is defined for 1-31%, some of the samples will be outside this range. A simple approach to solve the problem is to extrapolate using extreme points of the function range. This will cause the simulation results to contain higher variance. However, as the rank acceptabilities are computed based on rank counts from the individual simulations, the out-of-range samples do not introduce excessive variance due to the ordinal nature of the computations.

By allowing imprecise weight information, the SMAA approach enables to analyze the treatment benefit-risk profiles with increasingly precise weight information: missing, ordinal, trade-off intervals, and exact trade-offs. To illustrate the

effects of increasing precision of weight information, we elicited weight information from an expert in cardiovascular medicine. We first asked for a ranking of the criteria scale swings and then elicited exact trade-off statements. Afterwards we asked the expert to assess uncertainty of his exact statements to obtain trade-off intervals. The elicitation protocol is presented in Appendix A.

4 Results

The results of the weight elicitation are presented in Table 3. For exact and interval trade-offs, transaminase and CK elevation had weights much smaller than the sampling error (<< 0.01), and therefore in those analyses we included only the other 4 criteria. We computed the SMAA rank acceptability indices using the four sets of weight information: missing, ordinal, interval trade-offs, and exact trade-off statements. The rank acceptabilities are illustrated in Figure 5. The analysis code and the full data set are freely available online (26).

<< TABLE 3: weight elicitation results approx here >>

<< FIGURE 5: rank acceptabilities approx here >>

The results indicate that without any weight information (Figure 5, top-left panel), all treatments apart from the no-treatment alternative, control, have a possibility to be the preferred one, although atorvastatin, fluvastatin, and simvastatin have the highest first rank probabilities. When ordinal information is incorporated in the analysis, fluvastatin becomes clearly the likely candidate for being the preferred treatment (76% first rank acceptability), and pravastatin and rosuvastatin obtain approximately zero first rank probabilities. When more precise weight information (trade-off intervals) is added into the analysis, the results remain approximately the same. The precise weight statements indicate that the hard clinical endpoints are very important in the analysis. For example, the scale swing of all-cause mortality was considered by our decision maker to be approximately 440 times more important than the scale swing of transaminase, which could be then excluded from the analysis altogether. The imprecise ratio bounds are quite tight, and therefore results from the analysis with exact trade-off ratios (weights) are very similar to the ones from the analysis with interval trade-offs. The remaining uncertainty is due to the imprecise relative effects that were obtained as pooled estimates in the network meta-analysis.

5 Discussion

Considering the benefits and harms of multiple treatment options has clear appeal for a variety of decisions in healthcare. Such considerations are an essential component of prescription decisions in clinical practice, development of clinical practice guidelines by expert committees, and benefit-risk assessments for market entry decisions in regulatory settings (27-31). MCDA offers a framework to explicitly compare and contrast, and transparently trade-off the benefits and harms of multiple healthcare interventions.

Our case on statins illustrated the key challenges in applying MCDA to comparative benefit-risk assessment. First, the set of criteria should be defined to capture all aspects relevant for the decision, and to avoid double counting. Second, evidence from clinical trials should be synthesized through network meta-analysis to obtain relative treatment estimates, and these should be transformed to absolute effect estimates. The transformation requires a baseline estimate, which should be obtained taking into account the target population. Third, suitable scale ranges for the absolute scales should be defined for the weight elicitation. We recommend using the 95% credible interval hull from the absolute scale joint

distribution. Fourth, weight information should be elicited at different levels of precision to understand the effect of imprecise and incomplete weights on the analysis results. And finally, uncertainty on both the effect estimates and the weights should be incorporated into the analysis by applying a simulation-based technique, resulting in rank probabilities for all the different precisions of weight information.

Beyond the methodological considerations outlined in this paper, operationalizing the use of MCDA for comparative benefit-risk assessment has a number of limitations. While MCDA is gaining momentum in regulatory settings for evaluating the benefit and risk assessment of single agents (28), drug licensing agencies such as the European Medicines Agency and the Food and Drug Administration are still reluctant to consider comparative evidence to evaluate the relative benefit-risk profiles of new drugs (32, 33). When regulatory agencies adopt relative effectiveness as a criterion for licensing decisions in the future, network meta-analysis, and its combination with MCDA, would serve as a valuable tool to inform decision-making in the regulatory setting.

Comparative benefit-risk assessment using MCDA has clear implications for routine clinical practice. However, using MCDA for making prescribing decisions in clinical practice faces a number of practical challenges. Incorporating MCDA models into evidence-based computerized decision aids would necessitate pre-specifying, automating, and making available large parts of the MCDA model ahead of a clinical encounter. Decision aids such as SHARE-IT are already capable of automating and packaging key aspects of existing evidence into accessible summaries for patients (34). Future efforts should focus on integrating MCDA capabilities to such decision aids. Patients could use these either prior to a clinic visit or during the patient-clinician encounter. This would allow the patient-provider interaction to focus primarily on patient preferences on various benefit and harm outcomes, which would inform the prescribing decision. As we argued previously (35), we envision a future where computer decision aids are informed by systematic reviews and syntheses of all relevant clinical evidence on clinically meaningful benefit and harm outcomes. Combining evidence syntheses with MCDA would make feasible evidence-based decisions that are informed by provider expertise and knowledge, and tailored according to patient preferences.

The approach presented in this paper is purely illustrative and is not intended to dictate prescribing decisions in clinical practice. To the contrary, our case study highlights the importance of carefully accounting for decision maker preferences when considering both benefit and harm outcomes, and their trade-offs. Irrespective of how the benefit-risk assessment is conducted, however, we appreciate that there are external factors that influence the final decision taken. For example, consider a new treatment that is compared against the standard of care. Even when this new treatment is considered to have a better benefit-risk balance, the decision may still fall in favor of the established treatment because of concerns with certain identified risks for which the available evidence was highly uncertain. Depending on the nature of the problem, the final benefit-risk decision may also be affected by various ethical, social, and economic aspects that do not directly influence a treatment's benefit-risk balance but could nevertheless still have a profound impact on the acceptance of the decision by the public and other stakeholders. As such MCDA approaches present an opportunity to guide and inform decisions, rather than to dictate them.

References

1. Mt-Isa S, Hallgreen CE, Wang N, Callréus T, Genov G, Hirsch I, et al.. Balancing benefit and risk of medicines: a systematic review and classification of available methodologies. Pharmacoepid Dr S. 2014 May;23:667–78.

2. Tervonen T, van Valkenhoef G, Buskens E, Hillege HL, Postmus D. A stochastic multi-criteria model for evidence-based decision making in drug benefit-risk analysis. Stat Med. 2011 May 30;30:1419–28.

3. Holden WL. Benefit-risk analysis: a brief review and proposed quantitative approaches. Drug Saf. 2003 Oct;26:853–62.

4. Mussen F, Salek S, Walker S. A quantitative approach to benefit-risk assessment of medicines—part 1: the development of a new model using multi-criteria decision analysis. Pharmacoepid Dr S. 2007 Jul;16(Suppl. I):S12–15.

5. Felli JC, Noel RA, Cavazzoni PA. A multiattribute model for evaluating the benefit-risk profiles of treatment alternatives. Med Decis Making. 2009 Jan/Feb;29(1):104–15.

6. van Valkenhoef G, Tervonen T, Zhao J, de Brock B, Hillege HL, Postmus D. Multi-criteria benefit-risk assessment using network meta-analysis. J Clin Epidemiol. 2012 Apr;65(4):394–403.

7. Naci H, Brugts JJ, Ades AE. Comparative Tolerability and Harms of Individual Statins: A Study-Level Network Meta-Analysis of 234,550 Participants from 133 Randomized Controlled Trials. Circ Cardiovas Qual Outcomes. 2013 Jul 1;6(4):390–9.

8. Naci H, Brugts JJ, Tsoi B, Toor H, Fleurence R, Ades AE. Comparative Benefits of Statins in Primary and Secondary Prevention of Major Coronary Events and All-cause Mortality: A meta-analysis of placebo-controlled and active-comparator trials. Eur J Prev Cardiol. 2013 Aug;20(4):641–57.

9. Naci H, Brugts JJ, Fleurence R, Ades AE. Comparative Effects of Statins on Major Cerebrovascular Events: A network meta-analysis of placebo-controlled and active-comparator trials. Q J Med. 2013 Apr;106(4):299–306.

10. Lahdelma R, Makkonen S, Salminen P. Multivariate Gaussian criteria in SMAA. Eur J Oper Res. 2006 May 1;170(3):957–70.

11. Lahdelma R, Makkonen S, Salminen P. Two ways to handle dependent uncertainties in multi-criteria decision problems. Omega. 2009 Feb;37(1):79–92.

12. Ades AE. ISPOR states its position on network meta-analysis. Value Health. 2011 Jun;14(4):414–6.

13. Caldwell DM, Ades AE, Higgins JPT. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ. 2005 Oct 13;331:897–900.

14. Dias S, Sutton AJ, Ades AE, Welton NJ. Evidence synthesis for decision making 2: A generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. Med Decis Making. 2013a Jul;33(5):607–17.

15. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med. 2004 Oct 30; 23(20):3105–24.

16. Naci H. Generating comparative data on clinical benefits and harms of statins to inform prescribing decisions: evidence from network meta-analyses. PhD thesis, The London School of Economics and Political Science (LSE), 2014.

http://http://etheses.lse.ac.uk/973/1/Naci_Generating_Comparative_Data_Clinical_Benefits_Harms_Statins.pdf. Accessed 20 January 2015.

17. Dias S, Welton NJ, Sutton AJ, Ades AE. Evidence synthesis for decision making 5: The baseline natural history model. Med Decis Making. 2013 Jul;33(5):657–70.
18. Coordinators for the ALLHAT Collaborative Research Group. Major outcomes in moderately hypercholesterolemic, hypertensive patients randomized to pravastatin vs usual care: the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT-LLT). JAMA. 2002 Dec;288(23):2998–3007.
19. Keeney RL, Raiffa H. Decisions with multiple objectives: preferences and value trade-offs. Cambridge University Press;1976.
20. Weber M, Borcherding K. Behavioral influences on weight judgments in multiattribute decision making. Eur J Oper Res. 1993 May 28;67(1):1–12.
21. Lahdelma R, Salminen P. SMAA-2: Stochastic multicriteria acceptability analysis for group decision making. Oper Res. 2001 Jun 1;49(3):444–54.
22. Tervonen T, Figueira JR. A survey on stochastic multicriteria acceptability analysis methods. Journal of Multi-Criteria Decision Analysis. 2008 Jan-Apr;15(1-2):1–14.
23. Tervonen T, Lahdelma R. Implementing stochastic multicriteria acceptability analysis. Eur J Oper Res. 2007 Apr 16;178(2):500–513.
24. Tervonen T, van Valkenhoef G, Baştürk N, Postmus D. Hit-and-run enables efficient weight generation for simulation-based multiple criteria decision analysis. Eur J Oper Res. 2013 Feb 1;224(3):552–559.
25. van Valkenhoef G, Tervonen T, Postmus D. Notes on 'hit-and-run enables efficient weight generation for simulation-based multiple criteria decision analysis'. Eur J Oper Res. 2014 Dec 16;239(3):865–67.
26. Tervonen T, van Valkenhoef G, Naci H, Postmus D. Analysis code and data set for "Applying Multiple Criteria Decision Analysis (MCDA) to Comparative Benefit-Risk Assessment - Choosing Among Statins in Primary Prevention". ZENODO, 2015. http://dx.doi.org/10.5281/zenodo.16856. Accessed 17 April 2015.
27. European Medicines Agency (EMA). Report of the CHMP working group on benefit-risk assessment models and methods; 2007 Jan 19. http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2010/01/WC500069668.pdf. Accessed 20 January 2015.
28. EMA Benefit-Risk Methodology Project Team. Benefit-risk methodology project. Work package 4 report: benefit-risk tools and processes; 2012 Jun 13. http://www.ema.europa.eu/docs/en_GB/document_library/Report/2012/03/WC500123819.pdf. Accessed 20 January 2015.
29. Food and Drug Administration (FDA). PDUFA reauthorization performance goals and procedures fiscal years 2013 through 2017. http://www.fda.gov/downloads/forindustry/userfees/prescription-druguserfee/ucm270412.pdf. Accessed 30 April 2013.

30. Garrison LP. Regulatory benefit-risk assessment and comparative effectiveness research: strangers, bedfellows or strange bedfellows? Pharmacoeconomics. 2010 Oct;28(10):855–65.

31. Guo JJ, Pandey S, Doyle J, Bian B, Lis Y, Raisch DW. A review of quantitative risk-benefit methodologies for assessing drug safety and efficacy: report of the ISPOR risk-benefit management working group. Value Health. 2010 Aug;13(5):657–66.

32. Sorenson C, Naci H, Cylus J, Mossialos E. Evidence of comparative efficacy should have a formal role in European drug approvals. BMJ. 2011 Sep;343:d4849.

33. Stafford, RS, Todd HW, Lavori LW. New, but not improved? Incorporating comparative-effectiveness information into FDA labeling. NEJM. 2009 Sep;361(13):1230–1233.

34. Agoritsas T, Heen AF, Brandt L, Alonso-Coello P, Kristiansen A, Akl EA. Decision aids that really promote shared decision making: the pace quickens. BMJ. 2015 Feb 10; 350:g7624.

35. Naci H, van Valkenhoef GHM, Higgins JPT, Fleurence RL, Ades AE. Evidence-based prescribing: Combining network meta-analysis and multi-criteria decision analysis to choose among drugs. Circ Cardiovasc Qual Outcomes. 2014;To appear.

Conflict of interest/disclosure

TT, HN, AEA, AA, HLH and DP declare no conflicts of interest.

GvV has provided consulting services to Johnson & Johnson and (as a subcontractor of Deloitte) for UCB Pharma on the conduct of network meta-analyses.

Tables

Table 1: Summary of the guidelines

Decision making phase	**Main questions for the analyst**
Define the decision problem	Which indication are we assessing? What are the alternative treatments under consideration? What are the relevant evaluation criteria? Who is the decision maker?
Choose a set of non-overlapping evaluation criteria	Are some of the evaluation criteria measuring the same underlying concept? Which ones is the decision maker comfortable with removing? Are some important criteria missing?
Synthesize and summarize the available data	Is there more than a single study available? Would network meta-analyses be suitable for synthesizing the evidence?
Translate relative measures obtained in evidence synthesis to absolute scales	Are the baseline effects invariant over subgroups? Can we distinguish a high quality study with a population representative of the decision problem target population?
Define suitable scale ranges	Are the 95% credible intervals suitable for preference elicitation?
Elicit preference information	Are linear partial value functions appropriate? Does the decision maker understand the weight elicitation process?
Incorporate uncertainty in the analysis	Can we distinguish some good / bad treatment alternatives with only ordinal weight information? What amount of weight information is sufficient for discriminating some of the best alternatives? How much decision uncertainty remains with exact weights?

Table 2: Sample quantiles (100,000 draws) on absolute measurement scales[1]

Endpoint / Quantile	**0%**	**2.5%**	**97.5%**	**100%**

[1] The 0% and 100% (min/max) are highly unstable, and are shown only to illustrate that by setting the scale ranges to the 95% credible interval, we must accept a small error in the final analysis due to sampled values falling outside the defined scale range.

Nonfatal MI	0.01	0.02	0.06	0.09
Nonfatal Stroke	0.00	0.00	0.09	0.92
All-cause Mortality	0.03	0.05	0.15	0.21
Myalgia	0.00	0.00	0.13	0.74
Transaminase	0.00	0.01	0.31	0.86
CK Elevation	0.00	0.00	0.06	0.82

Table 3: Swing weight elicitation results for the four analyses with increasingly precise weight information

Analysis	**Weight information**[2]
Preference-free	Missing
Ordinal	$w_{mort} > w_{stroke} > w_{MI} > w_{myalgia} > w_{CK} > w_{trans}$
Exact trade-offs	Unnormalized weights, defined with respect to the more important outcome. For example, mortality scale swing was considered by the decision maker to be 4 times as important as the scale swing of stroke, meaning that the weight of non-fatal strokes = ¼ of the weight of all-cause mortality. $w_{mort} = 1$ $w_{stroke} = \frac{1}{4} * w_{mort}$ $w_{MI} = \frac{1}{2} * w_{stroke}$ $w_{myalgia} = \frac{1}{25} * w_{MI}$ $w_{CK} = \frac{1}{2} * w_{myalgia}$ $w_{trans} = \frac{1}{1.1} * w_{CK}$ Normalized weights. CK elevation and transaminase have weights << 0.01, and are thus irrelevant for the analysis.

[2] Weights of the different criteria: w_{MI} (nonfatal MI), w_{stroke} (nonfatal stroke), w_{mort} (all-cause mortality), $w_{myalgia}$ (myalgia), w_{trans} (transaminase), w_{CK} (CK elevation).

	$w_{mort} = 0.72,\ w_{stroke} = 0.18,\ w_{MI} = 0.09,\ w_{myalgia} = 0.003$
Interval trade-offs	$w_{mort} / w_{stroke} \in [3, 5]$ $w_{stroke} / w_{MI} \in [1.5, 2.5]$ $w_{myalgia} / w_{MI} \in [20, 30]$ $w_{MI} / w_{CK} \in [1.5, 2.5]$ $w_{CK} / w_{trans} \in [0.8, 1.4]$

Figure captions

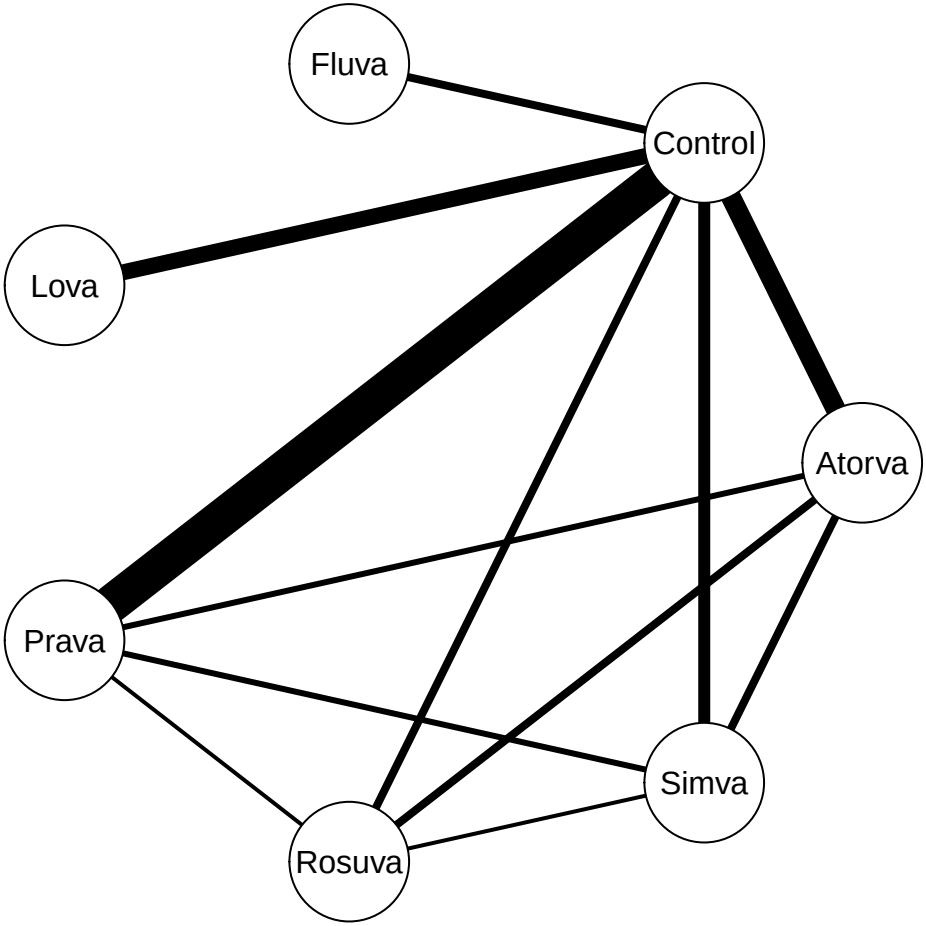
Figure 1: Network of trials used for the analysis. Edge thickness indicates the amount of studies comparing the treatments on the outcome 'All-cause Mortality' (1–22).

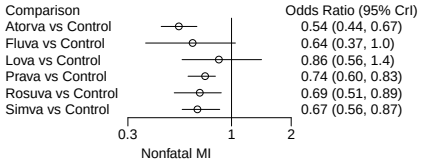
Figure 2: Odds ratios of the six treatments against control on the chosen 6 benefit-risk criteria.

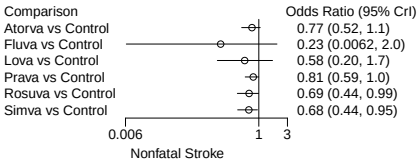
Figure 3: Absolute scale measurement ranges for the 6 treatments and control (used as baseline) on the chosen 6 benefit-risk criteria.

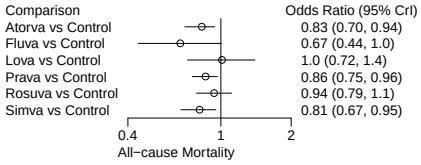
Figure 4: Example partial value functions (u_1, u_2) for a two-criterion additive value model, for a hypothetical choice of an obesity treatment. The first criterion measures the incidence rate of serious side effects (treatment risks), and the second criterion measures the observed weight loss in percentages (treatment benefit). Treatment a^1 has 0.5-incidence rate of the side effects, which translates to a value of 0.66 on criterion 1. Treatment a^2 has 0.6-incidence rate of side effects, which translates to a value of 0.33 on criterion 1. The weights 0.6 and 0.4 express that the scale swing of the first criterion [0.7, 0.4] is 50% more important than the scale swing of the second criterion [0%, 15%]. In this example, given exact measurements and preference information (shapes of partial value functions and weights), the second alternative (a^2) is preferred over the first alternative (a^1) because it has a higher value with the given value function (0.56 vs 0.44).

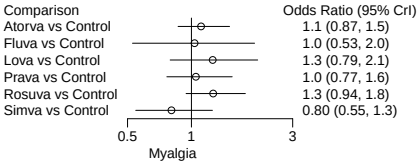
Figure 5: Rank acceptability indices for the 6 treatments and control from the 4 analyses with increasingly precise weight information.

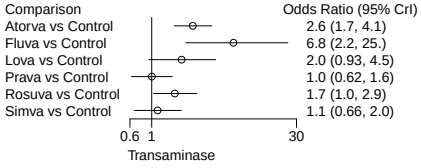

Fluva
Control
Lova
Atorva
Prava
Simva
Rosuva

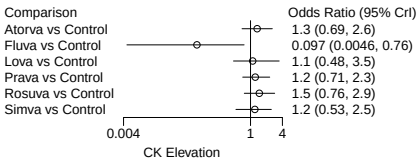

Comparison
Odds Ratio (95% CrI)
Atorva vs Control
0.54 (0.44, 0.67)
Fluva vs Control
0.64 (0.37, 1.0)
Lova vs Control
0.86 (0.56, 1.4)
Prava vs Control
0.74 (0.60, 0.83)
Rosuva vs Control
0.69 (0.51, 0.89)
Simva vs Control
0.67 (0.56, 0.87)
0.3
1
2
Nonfatal MI

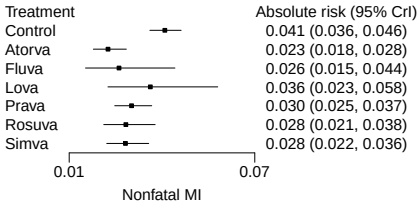

Comparison
Odds Ratio (95% CrI)
Atorva vs Control
0.77 (0.52, 1.1)
Fluva vs Control
0.23 (0.0062, 2.0)
Lova vs Control
0.58 (0.20, 1.7)
Prava vs Control
0.81 (0.59, 1.0)
Rosuva vs Control
0.69 (0.44, 0.99)
Simva vs Control
0.68 (0.44, 0.95)
0.006
1
3
Nonfatal Stroke

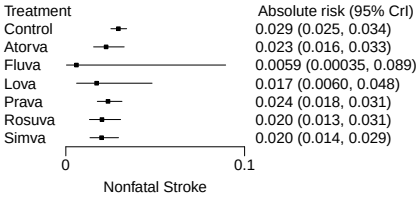

Comparison
Odds Ratio (95% CrI)
Atorva vs Control
0.83 (0.70, 0.94)
Fluva vs Control
0.67 (0.44, 1.0)
Lova vs Control
1.0 (0.72, 1.4)
Prava vs Control
0.86 (0.75, 0.96)
Rosuva vs Control
0.94 (0.79, 1.1)
Simva vs Control
0.81 (0.67, 0.95)
0.4
1
2
All−cause Mortality

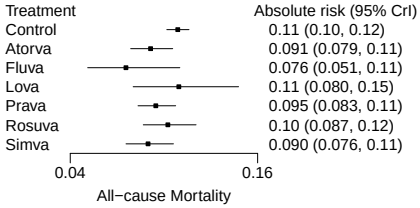

Comparison
Odds Ratio (95% CrI)
Atorva vs Control
1.1 (0.87, 1.5)
Fluva vs Control
1.0 (0.53, 2.0)
Lova vs Control
1.3 (0.79, 2.1)
Prava vs Control
1.0 (0.77, 1.6)
Rosuva vs Control
1.3 (0.94, 1.8)
Simva vs Control
0.80 (0.55, 1.3)
0.5
1
3
Myalgia

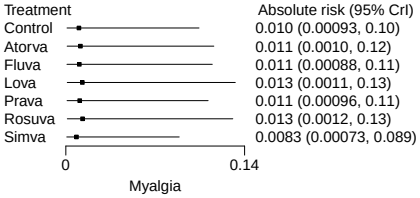

Comparison
Odds Ratio (95% CrI)
Atorva vs Control
2.6 (1.7, 4.1)
Fluva vs Control
6.8 (2.2, 25.)
Lova vs Control
2.0 (0.93, 4.5)
Prava vs Control
1.0 (0.62, 1.6)
Rosuva vs Control
1.7 (1.0, 2.9)
Simva vs Control
1.1 (0.66, 2.0)
0.6
1
30
Transaminase

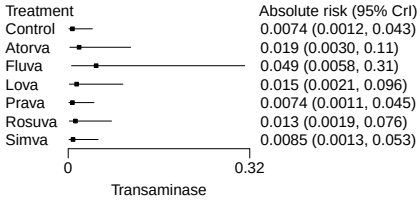

Comparison
Odds Ratio (95% CrI)
Atorva vs Control
1.3 (0.69, 2.6)
Fluva vs Control
0.097 (0.0046, 0.76)
Lova vs Control
1.1 (0.48, 3.5)
Prava vs Control
1.2 (0.71, 2.3)
Rosuva vs Control
1.5 (0.76, 2.9)
Simva vs Control
1.2 (0.53, 2.5)
0.004
1
4
CK Elevation

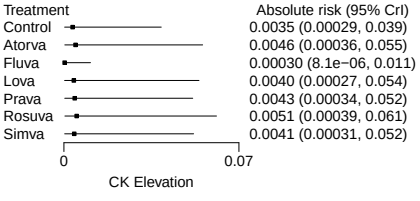

Treatment
Absolute risk (95% CrI)
Control
0.041 (0.036, 0.046)
Atorva
0.023 (0.018, 0.028)
Fluva
0.026 (0.015, 0.044)
Lova
0.036 (0.023, 0.058)
Prava
0.030 (0.025, 0.037)
Rosuva
0.028 (0.021, 0.038)
Simva
0.028 (0.022, 0.036)
0.01
0.07
Nonfatal MI

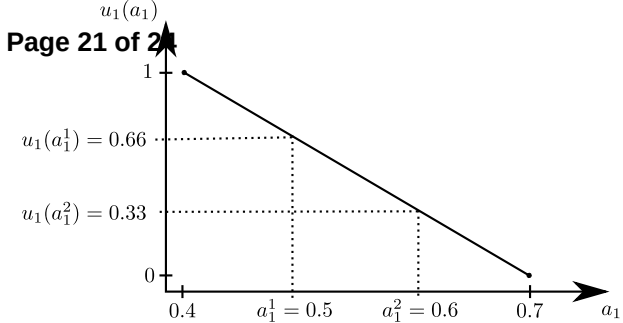

Treatment
Absolute risk (95% CrI)
Control
0.029 (0.025, 0.034)
Atorva
0.023 (0.016, 0.033)
Fluva
0.0059 (0.00035, 0.089)
Lova
0.017 (0.0060, 0.048)
Prava
0.024 (0.018, 0.031)
Rosuva
0.020 (0.013, 0.031)
Simva
0.020 (0.014, 0.029)
0
0.1
Nonfatal Stroke

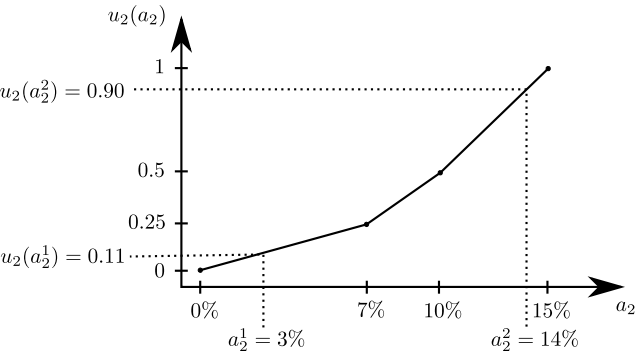

Treatment
Absolute risk (95% CrI)
Control
0.11 (0.10, 0.12)
Atorva
0.091 (0.079, 0.11)
Fluva
0.076 (0.051, 0.11)
Lova
0.11 (0.080, 0.15)
Prava
0.095 (0.083, 0.11)
Rosuva
0.10 (0.087, 0.12)
Simva
0.090 (0.076, 0.11)
0.04
0.16
All−cause Mortality


Treatment
Absolute risk (95% CrI)
Control
0.010 (0.00093, 0.10)
Atorva
0.011 (0.0010, 0.12)
Fluva
0.011 (0.00088, 0.11)
Lova
0.013 (0.0011, 0.13)
Prava
0.011 (0.00096, 0.11)
Rosuva
0.013 (0.0012, 0.13)
Simva
0.0083 (0.00073, 0.089)
0
0.14
Myalgia


Treatment
Absolute risk (95% CrI)
Control
0.0074 (0.0012, 0.043)
Atorva
0.019 (0.0030, 0.11)
Fluva
0.049 (0.0058, 0.31)
Lova
0.015 (0.0021, 0.096)
Prava
0.0074 (0.0011, 0.045)
Rosuva
0.013 (0.0019, 0.076)
Simva
0.0085 (0.0013, 0.053)
0
0.32
Transaminase


Treatment
Absolute risk (95% CrI)
Control
0.0035 (0.00029, 0.039)
Atorva
0.0046 (0.00036, 0.055)
Fluva
0.00030 (8.1e−06, 0.011)
Lova
0.0040 (0.00027, 0.054)
Prava
0.0043 (0.00034, 0.052)
Rosuva
0.0051 (0.00039, 0.061)
Simva
0.0041 (0.00031, 0.052)
0
0.07
CK Elevation





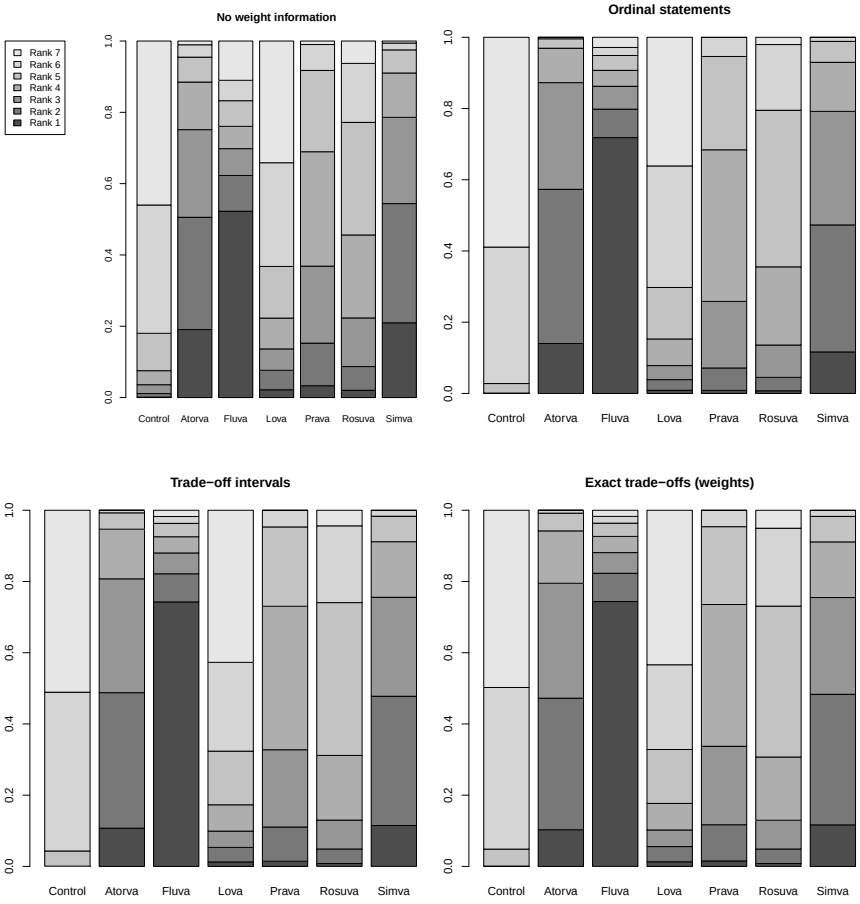
$$u(a) = w_1 \cdot u_1(a_1) + w_2 \cdot u_2(a_2)$$

$$\mathrm{w} = [0.6,\ 0.4]$$

$$a^1 = [0.5, 3\%],\ u(a^1) = 0.6 \cdot 0.66 + 0.4 \cdot 0.11 = 0.44$$

$$a^2 = [0.6, 14\%],\ u(a^2) = 0.6 \cdot 0.33 + 0.4 \cdot 0.90 = 0.56$$

$$\Rightarrow a^2 \succ a^1$$


No weight information
Ordinal statements
Trade−off intervals
Exact trade−offs (weights)
Rank 7
Rank 6
Rank 5
Rank 4
Rank 3
Rank 2
Rank 1
0.0
0.2
0.4
0.6
0.8
1.0
Control
Atorva
Fluva
Lova
Prava
Rosuva
Simva

Appendix A: weight elicitation protocol

Endpoint	Definition	Best value (%)	Worst value (%)	Rank	Exact	Ratio bounds
Nonfatal MI	5-year incidence of non-fatal heart attacks with severity ranging from mild to severe	2	6			
Nonfatal Stroke	5-year incidence of non-fatal strokes with severity ranging from mild to severe	0	9			
All-cause Mortality	5-year incidence of mortality	5	15			
Myalgia	Fraction of individuals with muscle pain	0	13			
Transaminase	Fraction of individuals with clinically meaningful (3x baseline values) elevations in either alanine aminotransferase or aspartate aminotransferase	0	31			
CK Elevation	Fraction of individuals with clinically meaningful (1.5-3x baseline values) elevations in creating kinase	0	6			

1. Ordinal elicitation

Given a treatment with outcome values of 'Worst value', give an order of importance for changing the outcome values to 'Best value' (would you rather change the risk of 'Nonfatal MI' from 6% to 2% or the risk of 'Nonfatal Stroke' from 9% to 0%?).

Mark these ranks in the column 'Rank', such that "1" is the best, "2" second best, etc.

2. Exact importance ratios

For each outcome other than the one you chose as the least important (rank n) in the previous step, give an estimate on how many times more important the **worst-best** scale swing of that outcome is compared to the swing of the next most important one.

For example, assume that 'Nonfatal MI' was your most important outcome, 'Myalgia' the second most important one, followed by Transaminase. Now, if you judge Nonfatal MI risk scale swing to be 2 times more important than that of Myalgia, then the decrease of Nonfatal MI risk from 6% to 2% is 2 times more important than the decrease of Myalgia from 13% to 0%. The following question would then be of the ratio of importance of the scale swing of Myalgia against

that of Transaminase.

Mark these importance ratios in the column 'Exact'.

3. Ratio bounds

For each of the judgments done in the previous step (exact importance ratios), give lower and upper bounds on your judgments.

For example, if you in the previous step judged the Nonfatal MI scale swing to be 2 times more important than that of Myalgia and you're quite uncertain about the exact number, the ratio bounds could be [1, 3] (i.e. the scale swing is 1-3 times as important).

Mark these bounds in the column ‘Ratio bounds’.