POLITECNICO DI TORINO
Repository ISTITUZIONALE

A novel measure of drug benefit–risk assessment based on Scale Loss Score

Original
A novel measure of drug benefit–risk assessment based on Scale Loss Score / Saint-Hilary, Gaelle; Robert, Veronique; Gasparini, Mauro; Jaki, Thomas; Mozgunov, Pavel. - In: STATISTICAL METHODS IN MEDICAL RESEARCH. - ISSN 0962-2802. - (2019), p. 962280218786526. [10.1177/0962280218786526]

Availability:
This version is available at: 11583/2711581 since: 2018-07-31T12:04:47Z

Publisher:
SAGE Publications Ltd

Published
DOI:10.1177/0962280218786526

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
Sage postprint/Author's Accepted Manuscript

(Article begins on next page)

12 October 2024

Check for updates

Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
0(0) 1–16
© The Author(s) 2018
Reprints and permissions:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/0962280218786526
journals.sagepub.com/home/smm
SAGE

A novel measure of drug benefit–risk assessment based on Scale Loss Score

Gaelle Saint-Hilary,[1] Veronique Robert,[2] Mauro Gasparini,[1] Thomas Jaki[3] and Pavel Mozgunov[3]

Abstract

Quantitative methods have been proposed to assess and compare the benefit-risk balance of treatments. Among them, multicriteria decision analysis (MCDA) is a popular decision tool as it permits to summarise the benefits and the risks of a drug in a single utility score, accounting for the preferences of the decision-makers. However, the utility score is often derived using a linear model which might lead to counter-intuitive conclusions; for example, drugs with no benefit or extreme risk could be recommended. Moreover, it assumes that the relative importance of benefits against risks is constant for all levels of benefit or risk, which might not hold for all drugs. We propose Scale Loss Score (SLoS) as a new tool for the benefit–risk assessment, which offers the same advantages as the linear multicriteria decision analysis utility score but has, in addition, desirable properties permitting to avoid recommendations of non-effective or extremely unsafe treatments, and to tolerate larger increases in risk for a given increase in benefit when the amount of benefit is small than when it is high. We present an application to a real case study on telithromycin in Community Acquired Pneumonia and Acute Bacterial Sinusitis, and we investigated the patterns of behaviour of Scale Loss Score, as compared to the linear multicriteria decision analysis, in a comprehensive simulation study.

Keywords

Benefit–risk, bounds penalisation, decision-making, loss score, multicriteria decision analysis

1 Introduction

A drug benefit–risk assessment consists of balancing its favourable therapeutic effects versus adverse reactions it may induce.[1] The benefit–risk balance is a strong predictor of the therapy's long-term viability and a key element for decision-making during the drug's development, the regulatory approval process, and the post-marketing follow-up.[2–4] For many years, a qualitative description of evidences had been the main approach to establish a drug's profile.[5,6] This approach, however, tends to lack transparency since the decision of taking (dropping) a drug is based on a large amount of data coming from different sources and on criteria which can vary for different experts. Structured frameworks and quantitative methodologies have been recently proposed to make a benefit–risk assessment more comprehensive and consistent.[7–12]

According to the European Medicine Agency Benefit–Risk Methodology Project,[6,13–15] one of the most comprehensive quantitative approaches is multicriteria decision analysis (MCDA).[16–19] It has also been recommended by several highly profiled expert groups, e.g. see IMI PROTECT Work package 5.[20] The main idea of MCDA is to calculate a single *utility score* using multiple criteria and taking into account the importance of each criterion. While non-linear forms of the utility score are recognised in various application areas of MCDA,[21,22] a linear aggregation of treatment's effects on benefits and risks remains the

[1]Dipartimento di Scienze Matematiche (DISMA) Giuseppe Luigi Lagrange, Politecnico di Torino, Torino, Italy
[2]Department of Biostatistics, Institut de Recherches Internationales Servier (IRIS), Suresnes, France
[3]Medical and Pharmaceutical Statistics Research Unit, Department of Mathematics and Statistics, Lancaster University, Lancaster, UK

Corresponding author:
Gaelle Saint-Hilary, Politecnico di Torino, Dipartimento di Scienze Matematiche (DISMA) Giuseppe Luigi Lagrange, Corso Duca degli Abruzzi 24, Torino 10129, Italy.
Email: gaelle.sainthilary@polito.it

most common choice for the drug development.[17,23–26] The major advantage of the linear model is its intuitive interpretation: a poor efficacy can be compensated by a good safety, and vice-versa. However, the linear utility score can result in the recommendation of highly unsafe or poorly effective drugs[27,28] and, consequently, in a counter-intuitive conclusion. Moreover, the linearity implies that the relative tolerance in the toxicity increase is constant for all levels of benefit. This leads to implicit assumptions on decision-makers' preferences which might not hold for all drugs. Avoiding these pitfalls is possible with, first, adopting good practices to ensure that the modeling approach makes sense,[27] and, second, by using non-additive and non-linear models.[12,29] The main objectives of this work are to explicitly illustrate the issues of the additive linear MCDA model through a comprehensive simulation study, and to provide an alternative approach, namely, Scale Loss Score, for aggregation of treatment's effect overcoming these issues. The proposed approach is based on recent developments in the theory of estimation in restricted parameter spaces[30,31] and is shown to have soundness in the context of drug evaluation.

The case study of telithromycin (Ketek®) raises questions regarding the suitability of a linear MCDA utility score for the drug benefit–risk assessment. Telithromycin was approved for the treatment of infections in several indications in 2001 by the EMA[32] and in 2005 by the FDA.[33] It was (qualitatively) reassessed in 2006–2007 by both agencies based on updated safety data. In particular, some serious visual adverse reactions, syncopes and acute liver failures have been reported. The terms of the marketing authorisations were varied in order to better describe the drug safety profile, and two indications were removed from the labeling by the FDA, among them Acute Bacterial Sinusitis (ABS). More recently, the IMI PROTECT Benefit–Risk Group[34] applied MCDA to this clinical example. Even if this assessment was performed solely for the purpose of testing the methodology, the main results indicated a fairly strong superiority of telithromycin versus the comparators in ABS, which is not consistent with the concerns expressed by the health authorities. Consequently, alternative methods more accurately reflecting decision-makers' preferences are of great interest.

In this work, we extend the assumption of non-linearity of preferences, which is well established in other fields such as microeconomics or ecology,[21,35,36] to the drug development context. We advocate two properties that a desirable measure of drug benefit–risk assessment should have:

(1) Decreasing level of risk tolerance relative to benefits: an increase in risk could be more tolerated when benefit improves from 'very low' to 'moderate', compared to from 'moderate' to 'very high'.
(2) Non-effective or/and extremely unsafe treatments should never be recommended.

Motivated by recent developments in the theory of the weighted information measures[37,38] and in the theory of estimation in restricted parameter spaces,[31] we propose Scale Loss Score (SLoS) as a novel measure for the benefit–risk assessment which shares both of these properties. The first property is achieved through convex preferences between efficacy and safety and the second one by a strong penalisation of extremely low benefit and high risk values.

We perform a comprehensive simulation study investigating the performances of SLoS and MCDA in many different scenarios. Note that this is, to our knowledge, the first time the properties of MCDA are systematically explored by simulations in the medical context. We also apply the new measure to the motivating clinical context of telithromycin. The elicitation of criterion weights for linear MCDA utility scores is widely discussed in the literature.[11,12,39–45] Therefore, we provide an algorithm of mapping MCDA weights to SLoS weights so that the same elicitation process could be followed while preserving the weight interpretation.

The rest of the paper is organised as follows. The MCDA utility score and the novel measure are detailed in Section 2. Section 3 describes the application of both measures in the real case study (telithromycin). We present a simulation study in Section 4 and conclude with discussion in Section 5. Additional information including source code to reproduce the results may be found in the Supplemental Material.

2 Methods

2.1 MCDA utility score

The original proposal of MCDA[16,17] ignores the uncertainty of parameter estimates. As this uncertainty can bare crucial information, an extension of MCDA taking into account the variability of estimates was proposed by Waddingham et al.[46] This approach is often called Probabilistic MCDA (or Stochastic MCDA) and is described below.

2.1.1 Utility score

Consider m treatments (indexed by i) which are assessed on n criteria (indexed by j). We adopt the following notation:

(i) ξ_{ij} is the performance of treatment i on criterion j, so that treatment i is characterised by the vector $\boldsymbol{\xi}_i = (\xi_{i1}, \ldots, \xi_{in})$.

(ii) The monotonically increasing partial value functions $0 \leq u_j(\cdot) \leq 1$ are used to normalise the criterion performances. Let ξ_j' and ξ_j'' be the most and the least preferable values, then $u_j(\xi_j'') = 0$ and $u_j(\xi_j') = 1$. The inequality $u_j(\xi_{ij}) > u_j(\xi_{hj})$ indicates that the performance of the treatment i is preferred to the performance of the treatment h on criterion j. In this work, we focus on linear partial value functions, one of the most common choice in drug benefit–risk assessment.[11,16,23,25,46] They can be written as

$$u_j(\xi_{ij}) = \frac{\xi_{ij} - \xi_j''}{\xi_j' - \xi_j''} \tag{1}$$

(iii) The weights indicating the relative importance of the criteria are known constants denoted by w_j such that $\sum_{j=1}^{n} w_j = 1$. The vector of weights used for the analysis is denoted by $\mathbf{w} = (w_1, \ldots, w_n)$.

The MCDA linear utility score is obtained as

$$u(\boldsymbol{\xi}_i, \mathbf{w}) := \sum_{j=1}^{n} w_j u_j(\xi_{ij}) \tag{2}$$

The higher the utility score, the more preferable the benefit–risk ratio. Then, the comparison of treatments i and h is based on

$$\Delta u(\boldsymbol{\xi}_i, \boldsymbol{\xi}_h, \mathbf{w}) := u(\boldsymbol{\xi}_i, \mathbf{w}) - u(\boldsymbol{\xi}_h, \mathbf{w})$$

While maximising utility is common in economics,[36] the concept of a loss function is usually preferred in statistical decision theory and Bayesian analysis for parameter estimation.[35] The complement of the MCDA utility score, $\bar{u}(\boldsymbol{\xi}_i, \mathbf{w}) = 1 - u(\boldsymbol{\xi}_i, \mathbf{w})$, could be considered as a *linear loss score* to be minimised, and it can be used equivalently as a measure of discrimination.

Although the term 'MCDA' outside of the health domain refers to the general methodology to summarise several characteristics in a single aggregated score, in this work we adopt the notation 'MCDA' for the additive utility score with linear partial values functions corresponding to the conventional model adopted so far in the drug benefit–risk assessment.[12]

2.1.2 Estimation

Within a Bayesian approach, the utility score $u(\boldsymbol{\xi}_i, \mathbf{w})$ is a random variable having a prior distribution. Given observed outcomes $\mathbf{x_i} = (x_{i1}, \ldots, x_{in})$ and $\mathbf{x_h} = (x_{h1}, \ldots, x_{hn})$ (corresponding to treatment performances $\boldsymbol{\xi}_i$ and $\boldsymbol{\xi}_h$, respectively) for i and h, one can obtain the posterior distribution of $\Delta u(\boldsymbol{\xi}_i, \boldsymbol{\xi}_h, \mathbf{w})$. The inference is based on the complete posterior distribution and the conclusion on the benefit–risk balance is supported by the probability of treatment i to have a greater utility score than treatment h:

$$\mathcal{P}_u^{ih} = \mathbb{P}(\Delta u(\boldsymbol{\xi}_i, \boldsymbol{\xi}_h, \mathbf{w}) > 0 | \mathbf{x_i}, \mathbf{x_h}) \tag{3}$$

The probability (3) is used to guide a decision on taking/dropping a drug. A possible way to formalise the decision based on this probability is to compare it to a threshold confidence level $0.5 \leq \psi \leq 1$. Then, $\mathcal{P}_u^{ih} > \psi$ would mean that one has enough evidence to say that treatment i has a better benefit–risk balance than h with a level of confidence ψ. Note that $\mathcal{P}_u^{ih} = 0.5$ corresponds to the case where the benefit–risk profiles of i and h are equal according to MCDA.

2.1.3 Weight elicitation

Weighting is a structured way to capture the stakeholders' preferences between the criteria. It is recognised as a complex problem since it involves both clinical and societal value judgments.[41] Methods for quantifying subjective

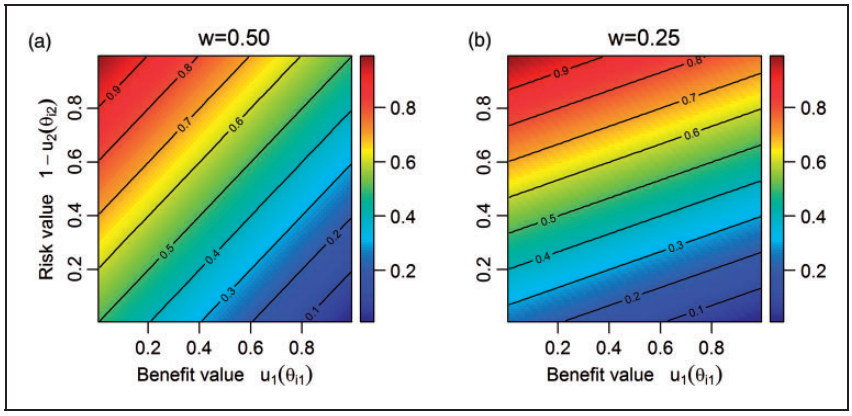

Figure 1. Left panel: contours of equal linear loss score $\bar{u}(\theta_{i1}, \theta_{i2}, w = 0.5)$. Right panel: contours of equal linear loss score $\bar{u}(\theta_{i1}, \theta_{i2}, w = 0.25)$.

preferences have been widely studied in the literature,[11,12,39,40,42,43,45] among which Discrete Choice Experiment and Swing-Weighting appeared to be appropriate in terms of theoretical foundations, cognitive burden, feasibility and robustness.[16,44,47,48] In the MCDA framework, the weight assigned to one criterion is interpreted as a scaling factor which relates one increment on this criterion to increments on all other criteria.

2.1.4 MCDA illustration: two criteria

Let us consider an example with two criteria (one benefit indexed by 1, one risk indexed by 2) to illustrate an insight on the linear utility score in equation (2). The utility score for treatment i at fixed parameter values θ_{i1}, θ_{i2} takes the form

$$u(\theta_{i1}, \theta_{i2}, w) := wu_1(\theta_{i1}) + (1 - w)u_2(\theta_{i2}) \tag{4}$$

As values $u_1(\theta_{i1}), u_2(\theta_{i2}) \in (0, 1)$, one can interpret $u_1(\theta_{i1})$ as a probability of benefit and $1 - u_2(\theta_{i2})$ as a probability of risk. The contours of equal linear loss score $\bar{u}(\theta_{i1}, \theta_{i2}, w) = 1 - u(\theta_{i1}, \theta_{i2}, w)$ for all values of $u_1(\theta_{i1})$ and $(1 - u_2(\theta_{i2}))$ using $w = 0.5$ (left panel) and $w = 0.25$ (right panel) are given in Figure 1.

Lower values of $\bar{u}(\theta_{i1}, \theta_{i2}, w)$ correspond to better drug benefit–risk profiles. It is minimised (right bottom corner) when the maximum possible benefit is reached ($u_1(\theta_{i1}) = 1$) with no risk ($1 - u_2(\theta_{i2}) = 0$). The contours of equation (4) are linear, with a constant slope $w/(1 - w)$. It implies that if one treatment has an increased probability of risk of x% compared to another, its benefit probability should be increased by $(1 - w)/w \times$ x% to have the same utility score. This holds for all values of benefit and risk. While the linear form of the utility score makes the interpretation simple, it might lead to some counter-intuitive conclusions. Below, we illustrate possible paradoxes for $w = 0.25$, i.e. when the importance of the risk is three times higher than the importance of the benefit.

(1) The benefit–risk trade-off is the same for all values of the risk/benefit.

Consider two cases where a drug increases the benefit probability from (a) 0.15 to 0.30 and (b) from 0.80 to 0.95 compared to another therapy. In case (a), the increase doubles the benefit probability and a higher increase in toxicity can usually be tolerated. At the same time, the same increase in case (b) is not as relatively large, therefore it can be argued that only a smaller increase in the risk probability may be tolerated. However, the linear utility score implies that the same increase in risk to match the benefit increase can be sacrificed.

(2) Drugs with 0% benefit or 100% risk can be recommended.

Consider the first example in Table 1: drug 1 that cannot treat patients and causes adverse events only would be preferred. At the same time, drug 2 that adds 11% toxicity, but 30% efficacy would not be chosen. Similarly, in the second example in Table 1, drug 1 that leads to an adverse event for 100% patients would be preferred.

Table 1. Examples of MCDA linear utility scores with two criteria and $w = 0.25$.

	Example 1		Example 2	
	Drug 1	Drug 2	Drug 1	Drug 2
Benefit: $u_1(\theta_{i1})$	0.00	0.30	0.96	0.50
Risk: $1 - u_2(\theta_{i2})$	0.09	0.20	1.00	0.85
Utility score: $u(\theta_{i1}, \theta_{i2}, w = 0.25)$	0.6825	0.6750	0.2400	0.2375

Even if none of those drugs are likely to be taken to the market, the goal of MCDA is to rank treatments and these examples reveal some counter-intuitive conclusions to which MCDA can lead. Note that decreasing values of w would help to solve the paradox in Example 1, but would worsen it in Example 2.

We advocate two properties of a benefit–risk analysis measure:

(i) for a given increase in benefit, one can tolerate a larger increase in risk if the amount of benefit is small than if it is high, and
(ii) one is not interested in the level of risk (benefit) if the drug does not treat (harm all) patients.

Formally, these properties correspond to (i) the concavity of equal loss score contours (or, equivalently, the convexity of equal utility score contours) and to (ii) a strong penalisation of extreme low benefit values and extreme high risk values. We would like to stress that the convexity of utility (concavity of loss) is widely advocated in microeconomics and is believed to reflect preferences in a more adequate way than linear ones in many applications.[21,36]

One can check that none of these properties are satisfied for MCDA due to its linearity. There are two forms of linearity in MCDA: in the partial value functions and in the utility score. Note that property (i) of decreasing level of risk tolerance relative to benefits can be achieved by varying the shape of the partial value functions (for instance, using concave functions for benefit and linear functions for risk). However, the explicit elicitation of non-linear forms for the partial value functions may be challenging. As the linear partial function remains a common choice in drug benefit–risk assessment, we propose a novel measure of aggregation which allows for both properties to be achieved even under linear partial value functions.

2.2 Scale loss score

2.2.1 Derivation

As an alternative to the linear MCDA utility score (equation (2)), we define SLoS for aggregation of treatment's performances as

$$l(\boldsymbol{\xi}_i, \tilde{\mathbf{w}}) := \sum_{j=1}^{n} \left(\frac{1}{u_j(\xi_{ij})} \right)^{\tilde{w}_j} \tag{5}$$

where $\tilde{w}_j$ is the weight indicating the average relative importance of criterion j compared to the others and $u_j(\cdot)$ is a linear partial value function (equation (1)). The form of SLoS is motivated by the scale symmetric loss function[31,49] and the precautionary loss function.[30] These functions allow to stay away from 'boundary' values $u_j(\cdot) = 0$. In the context of the benefit–risk assessment, they correspond to an extremely undesirable performance of a drug: low benefit or high risk. SLoS can be interpreted as a divergence between drug i characteristics $\boldsymbol{\xi}_i$ and the 'perfect' benefit–risk characteristics $(1, \ldots, 1)_{n\times 1}$. As a *loss score* is used rather than a utility score, lower values of $l(\boldsymbol{\xi}_i, \tilde{\mathbf{w}})$ correspond to more desirable performances of the drug.

Clearly, $l(\boldsymbol{\xi}', \tilde{\mathbf{w}})$ is minimised for $\boldsymbol{\xi}'$ such that $u_j(\xi'_j) = 1$ for all $j = 1, \ldots, n$, i.e. at the point of the ideal benefit–risk profile. Additionally, $l(\boldsymbol{\xi}''_{(k)}, \tilde{\mathbf{w}}) = +\infty$ for $\boldsymbol{\xi}''_{(k)}$ a vector of parameters containing ξ''_k such that $u_k(\xi''_k) = 0$, for at least one $k \in \{1, \ldots, n\}$, so the loss score for a treatment with at least one extreme negative performance is equal to infinity. The lower bounds are determined by the least preferred values ξ''_k used in the partial value functions, and correspond to *unacceptable* levels of benefit or risk. It should be noted that SLoS is intentionally sensitive to these unacceptable values, therefore their choice could have a non-negligible impact on the results. While unacceptable values of 0 (for benefit) or 1 (for risk) may be obvious choices for probabilities of a binary outcome, the unacceptable value for a continuous outcome may be more subjective and requires a careful investigation.

SLoS is a measure of the benefit–risk balance permitting to discriminate treatments according to their performances and according to the weights attributed to the criteria. The lower the SLoS, the more preferable the benefit–risk ratio, and the comparison of treatments i and h is based on

$$\Delta l(\xi_i, \xi_h, \tilde{\mathbf{w}}) := l(\xi_i, \tilde{\mathbf{w}}) - l(\xi_h, \tilde{\mathbf{w}})$$

2.2.2 Estimation

Similarly to MCDA, we consider a Bayesian model and assign a prior probability distribution to ξ_{ij}. Given the observed outcomes $\mathbf{x_i} = (x_{i1}, \ldots, x_{in})$ and $\mathbf{x_h} = (x_{h1}, \ldots, x_{hn})$ for the treatments i and h, one can obtain a posterior distribution of $\Delta l(\xi_i, \xi_h, \tilde{\mathbf{w}})$. Again, the inference is based on the complete posterior distribution and the conclusion on the benefit–risk balance is supported by the probability of treatment i to have a smaller SLoS than treatment h:

$$\mathcal{P}_l^{ih} = \mathbb{P}(\Delta l(\xi_i, \xi_h, \mathbf{w}) < 0 | \mathbf{x_i}, \mathbf{x_h})$$

This probability can be compared to a fixed confidence threshold ψ as in the MCDA approach.

2.2.3 SLoS illustration: two criteria

To illustrate the properties of SLoS, consider the example presented in Section 2.1.4 with one benefit and one risk. The SLoS for treatment i in the point of fixed parameter values takes the form

$$l(\theta_{i1}, \theta_{i2}, w) := \left(\frac{1}{u_1(\theta_{i1})}\right)^{\tilde{w}} + \left(\frac{1}{u_2(\theta_{i2})}\right)^{1-\tilde{w}} \tag{6}$$

Figure 2 presents the contours of SLoS (equation (6)) for all pairs of $u_1(\theta_{i1})$ and $1 - u_2(\theta_{i2})$ using $\tilde{w} = 0.5$ (left panel) and $\tilde{w} = 0.25$ (right panel). The tangents of the contours at the point (0.5,0.5) are presented on the graph for the purpose of the weight mapping detailed in the next section.

SLoS is minimised when the benefit–risk balance of the drug is maximised, at the point (1,0) (right bottom corner), where the maximum possible benefit is reached with no risk. The loss score is infinite for extreme low benefit values and extreme high risk values, thus non-effective or extremely unsafe treatments could never be recommended. Considering the cases presented in Table 1, the drug 2 had a SLoS equal to 2.53 for the first example and of 5.34 for the second example, and it is preferred to drug 1 which SLoS is infinite in both cases.

The contour lines of equal loss are concave, which is equivalent to having convex preferences between additional benefit and avoided risk, and have the form

$$1 - u_2(\theta_{i2}) \;=\; 1 - (z - u_1(\theta_{i1})^{-\tilde{w}})^{-\frac{1}{1-\tilde{w}}}, \quad \text{for } u_1(\theta_{i1}) > z^{-1/\tilde{w}}$$

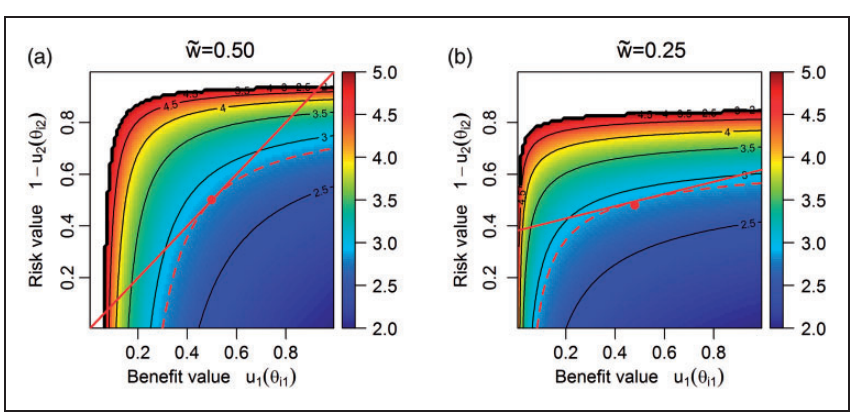

Figure 2. Left panel: contours of $l(\theta_{i1}, \theta_{i2}, \tilde{w} = 0.5)$. Right panel: contours of $l(\theta_{i1}, \theta_{i2}, \tilde{w} = 0.25)$. Red lines correspond to tangents at the point (0.5,0.5).

for a fixed value $l(\theta_{i1}, \theta_{i2}, \tilde{\mathbf{w}}) = z$. The slope of the tangent of the contour at a given $u_1(\theta_{i1})$ for the loss score value z is

$$\frac{\tilde{w}}{1-\tilde{w}}\left(z - u_1(\theta_{i1})^{-\tilde{w}}\right)^{\frac{\tilde{w}-2}{1-\tilde{w}}} . \, u_1(\theta_{i1})^{-(\tilde{w}+1)} \tag{7}$$

The slope decreases as benefit increases. It follows that the relative importance of the benefit criterion over the risk criterion decreases with the amount of benefit itself. In other words, an increase in toxicity is more tolerated if, in parallel, efficacy improves from 'very low' to 'moderate', compared to from 'moderate' to 'very high'.

2.2.4 Weight elicitation

Since comprehensive work has been published and is currently being continued on the weight elicitation for MCDA, we present a way to map MCDA weights w_j to SLoS weights $\tilde{w}_j$. Note that the slope of MCDA contour tangents is *constant* for all values of parameters and defined by the weights w_j only, while the slope of SLoS contour tangents is *non-constant* and defined by both $\tilde{w}_j$ and values of the criteria. To map weights, we would interpret $\tilde{w}_j$ as an *average* relative importance of each criterion over the others. With two criteria, the weight $\tilde{w}_j$ corresponding to the MCDA weight w_j can be found from the equality of the slopes of the tangents of MCDA and SLoS contours in the middle point $u_1(\theta_{i1}) = u_2(\theta_{i2}) = 0.5$ of treatment i performances

$$\frac{\tilde{w}_j}{1-\tilde{w}_j} . \, 2^{2\tilde{w}_j - 1} = \frac{w_j}{1-w_j} \tag{8}$$

where the slopes of SLoS and MCDA contour tangents in this point are given on the left- and right-hand sides, respectively.

The weight mapping (equation (8)) does not have an analytical solution, but the approximate value of $\tilde{w}_j$ can be obtained by line search. The mapping of the weights is illustrated in Figure 3. The weights are the same when both criteria are considered equally important ($w = \tilde{w} = 0.5$), while $w < 0.5$ corresponds to slightly greater values of $\tilde{w}$. For instance, $\tilde{w} = 0.30$ corresponds to $w = 0.25$.

Considering an arbitrary number of criteria, the mapping (equation (8)) can be applied to each value of the MCDA weights. For instance, using four criteria with a weight vector $\mathbf{w} = (0.30, 0.15, 0.15, 0.15, 0.25)$, the vector of SLoS weights is equal to $\tilde{\mathbf{w}} = (0.35, 0.21, 0.21, 0.21, 0.30)$. It should be noted that, in this case, the weights $\tilde{w}_j$ do not necessarily sum to 1, but this does not prevent from calculating SLoS, for which formula (5) still applies.

Mapping weights to the middle point of the benefit and risk treatment performance range relies on the assumption that MCDA weights were elicited across the entire range, or that the trade-off between criteria was

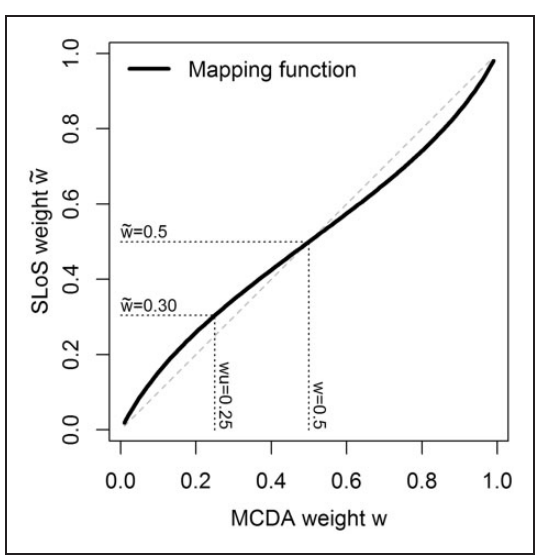

Figure 3. Weight mapping.

anchored on average at the middle point. However, in practice, MCDA weights could have been elicited at any other point and extrapolated. In this case, the mapping procedure above could be performed accordingly by finding the SLoS weight satisfying the equality of the slopes of MCDA and SLoS contour tangents in any other point of interest.

3 Case study: telithromycin

We illustrate the use of SLoS and MCDA in a real clinical context on the case-study thelithromycin (Ketek®) reported by the IMI PROTECT Benefit–Risk Group.[34]

Telithromycin was approved in 2001 for several indications as an alternative when beta-lactam antibiotics are not appropriate, and we will focus on the indications Community Acquired Pneumonia (CAP) and Acute Bacterial Sinusitis (ABS) as they well illustrate similarities and differences between the two methods. A Probabilistic MCDA model was considered in the IMI PROTECT report[34] (called Stochastic Multicriteria Acceptability Analysis with fixed weights), and MCDA utility scores presented here are derived from the original report.

Telithromycin is compared to a single alternative called 'comparator', which comprises amoxicillin-clavulanic acid, cefuroxime and clarithromycin, used as comparators in clinical studies and pooled together. The probabilities of five binary criteria, one benefit and four adverse events (AE), were transformed using linear partial value functions (equation (1)) with the following most and least preferred probabilities of occurrence ξ'_j and ξ''_j[34]

- Benefit: cure rate (CAP: $\xi'_1 = 1,\ \xi''_1 = 0.4$; ABS: $\xi'_1 = 0.86,\ \xi''_1 = 0.71$),
- Risks:
  - Hepatic AE (CAP: $\xi'_2 = 0,\ \xi''_2 = 0.1$; ABS: $\xi'_2 = 0,\ \xi''_2 = 0.02$),
  - Cardiac AE (CAP: $\xi'_3 = 0,\ \xi''_3 = 0.1$; ABS: $\xi'_3 = 0,\ \xi''_3 = 0.01$),
  - Visual AE (CAP: $\xi'_4 = 0,\ \xi''_4 = 0.1$; ABS: $\xi'_4 = 0,\ \xi''_4 = 0.02$),
  - Syncope (CAP: $\xi'_5 = 0,\ \xi''_5 = 0.1$; ABS: $\xi'_5 = 0,\ \xi''_5 = 0.01$).

Using uniform priors and given the number of cures and AE (see Supplemental Material, Table S1), Beta posterior distributions for the event probabilities are approximated using 100,000 simulations in R,[50] and are used to compute the corresponding distributions of the partial value functions. Means and 95% Credibility Interval (CrI) of the probabilities and of the partial value functions, and the MCDA weights, are summarised in Table 2.

This information was used to approximate the posterior distributions of MCDA linear utility score and SLoS. The mapped SLoS weights corresponding to the MCDA weights are $\tilde{\boldsymbol{w}} = (0.35,\ 0.21,\ 0.21,\ 0.21,\ 0.30)$.

For the CAP indication, MCDA and SLoS provide similar results, with probabilities that telithromycin is better than the comparator equal to 59% and 51%, respectively. These results indicate that telithromycin has a slightly better benefit–risk profile than the comparator, but with large uncertainty.

For the ABS indication, the probability that the benefit–risk balance of telithromycin is better than the comparator is equal to 71% using MCDA and 55% using SLoS. While they both indicate results in favour of telithromycin, this advantage appears to be much more uncertain with SLoS than with MCDA. The difference between the methods can be mainly explained by a higher rate of Visual AE with telithromycin (1.3% versus 0.5%), which is close to the least preferred value for this criterion in this indication ($\xi''_4 = 2\%$). This leads to low values of the corresponding partial value function (mean (95% CrI) $u_4(\xi_{14})$: 0.36 [0.02; 0.63]), and values at the lower end of the distribution are strongly penalised by SLoS. At the same time, the mass of the corresponding partial value function distribution of the comparator (mean (95% CrI) $u_4(\xi_{24})$: 0.74 [0.46; 0.91]) is shifted further from the bound, which results in lower value of SLoS. A similar argument could be applied to Hepatic AE, and the combination of these safety issues is more penalised by SLoS than by MCDA, despite the worse cure rate of the comparator. Even if the benefit–risk assessment by IMI PROTECT[34] was performed in order to test the methodologies and may have been conducted differently in the actual regulatory context, it is worth noting that the conclusion obtained using SLoS for the ABS indication is more in line with the concerns expressed by the Committee for Medicinal Products for Human Use (CHMP) regarding the atypical safety profile of the drug[51] and the removal of this indication from the labeling by the FDA.[33] This could be an example of SLoS reflecting the decision-makers' preferences more accurately than MCDA.

A sensitivity analysis was conducted using MCDA weights to compute SLoS (omitting the weight mapping) and the conclusions are globally robust, with the probability of telithromycin being better than the comparator equal to 57% for CAP and 62% for ABS.

Table 2. Mean and 95% CrI of the Beta posterior distributions of benefit and risk parameters and of corresponding partial value functions, with their MCDA weight, for Telithromycin (Teli.) and Comparator (Comp.).

	CAP		ABS		MCDA weights
	Teli.	Comp.	Teli.	Comp.	
Cure rate					
ξ_{i1}					
Mean	0.908	0.877	0.828	0.772	30%
95% CrI	[0.896; 0.919]	[0.855; 0.897]	[0.800; 0.855]	[0.715; 0.824]	
$u_1(\xi_{i1})$					
Mean	0.846	0.795	0.787	0.414	
95% CrI	[0.827; 0.864]	[0.759; 0.829]	[0.601; 0.964]	[0.036; 0.760]	
Hepatic AE					
ξ_{i2}					
Mean	0.044	0.042	0.011	0.004	15%
95% CrI	[0.034; 0.056]	[0.031; 0.054]	[0.006; 0.017]	[0.001; 0.009]	
$u_2(\xi_{i2})$					
Mean	0.561	0.582	0.468	0.789	
95% CrI	[0.444; 0.664]	[0.457; 0.691]	**[0.158; 0.707]**	**[0.542; 0.942]**	
Cardiac AE					
ξ_{i3}					
Mean	0.005	0.004	0.002	0.002	15%
95% CrI	[0.002; 0.01]	[0.001; 0.009]	[0.000; 0.004]	[0.000; 0.006]	
$u_3(\xi_{i3})$					
Mean	0.947	0.956	0.849	0.790	
95% CrI	[0.902; 0.979]	[0.909; 0.985]	[0.579; 0.982]	[0.414; 0.974]	
Visual AE					
ξ_{i4}					
Mean	0.011	0.004	0.013	0.005	15%
95% CrI	[0.006; 0.018]	[0.001; 0.009]	[0.008; 0.020]	[0.002; 0.011]	
$u_4(\xi_{i4})$					
Mean	0.887	0.956	0.357	0.736	
95% CrI	[0.823; 0.937]	[0.909; 0.986]	**[0.016; 0.625]**	**[0.461; 0.914]**	
Syncope					
ξ_{i5}					
Mean	0.002	0.004	0.001	0.002	25%
95% CrI	[0.000; 0.005]	[0.001; 0.008]	[0.000; 0.003]	[0.000; 0.006]	
$u_5(\xi_{i5})$					
Mean	0.977	0.964	0.924	0.789	
95% CrI	[0.945; 0.995]	[0.922; 0.990]	[0.719; 0.998]	[0.414; 0.974]	

In the next section, we present a simulation study illustrating the properties of SLoS and MCDA in many different scenarios.

4 Simulation study

4.1 Setting

To investigate the performances of SLoS and MCDA, we simulated randomised controlled clinical trials with two treatments $i = 1, 2$, named T_1 and T_2, $N = 100$ patients per group, and two uncorrelated binary criteria ($j = 1$ for benefit and $j = 2$ for risk). We assume that benefit events are desirable (e.g. treatment response), while risk events should be avoided (e.g. adverse event), with the performance parameters ξ_{ij} being their probability of occurrence. The partial value functions are defined as $u_1(\xi_{i1}) = \xi_{i1}$ and $u_2(\xi_{i2}) = 1 - \xi_{i2}$. Equally important criteria with weights $w_j = \tilde{w}_j = 0.5$, $j = 1, 2$, are considered.

The investigated scenarios are summarised in Table 3, where the expected probabilities of event θ_{ij} are presented for T_1 (•) and T_2 (◇). Nine sets of T_1 characteristics are fixed. For each set, all possible combinations of T_2

Table 3. Simulation scenarios with two criteria.

		Probability of Benefit θ_{i1}								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Probability of Risk θ_{i2}	0.9	◇●	◇	◇	◇	◇	◇	◇	◇	◇●
	0.8	◇	◇	◇	◇	◇	◇	◇	◇	◇
	0.7	◇	◇	◇●	◇	◇	◇	◇●	◇	◇
	0.6	◇	◇	◇	◇	◇	◇	◇	◇	◇
	0.5	◇	◇	◇	◇	◇●	◇	◇	◇	◇
	0.4	◇	◇	◇	◇	◇	◇	◇	◇	◇
	0.3	◇	◇	◇●	◇	◇	◇	◇●	◇	◇
	0.2	◇	◇	◇	◇	◇	◇	◇	◇	◇
	0.1	◇●	◇	◇	◇	◇	◇	◇	◇	◇●

●: treatment T_1; ◇: treatment T_2.

characteristics with $\theta_{21},\ \theta_{22} \in \{0.1,\ 0.2,\ 0.3,\ 0.4,\ 0.5,\ 0.6,\ 0.7,\ 0.8,\ 0.9\}$ are considered. This results in 81 profiles for T_2 and in 729 cases in total: we explored the grid of treatment performances in order to identify under which conditions MCDA and SLoS lead to different conclusions. For example, the first scenario corresponds to fixed expected probabilities of benefit and risk for treatment T_1 $\theta_{11} = \theta_{12} = 0.5$ compared to all considered combinations of probabilities of event for T_2. In this scenario, we expect SLoS to recommend T_1 more than MCDA when T_2 is associated with an extreme risk or no benefit. In the other scenario where $\theta_{11} = \theta_{12} = 0.1$, T_1 has almost no benefit, so it should not be recommended despite its good safety profile. Indeed, even if T_1 does not harm the patients (it is similar to a placebo), administrating it to the patients implies we make the assumption that it has a positive effect, while it has not in reality. This interpretation is close to the usual type I error, and is not acceptable from a regulatory and health economics perspective. Similarly, when $\theta_{11} = \theta_{12} = 0.9$, T_1 should not be recommended despite its outstanding efficacy as it is associated with an extreme risk. All intermediate cases are considered.

Let S be the total number of simulated trials and K the number of samples generated to approximate the distributions of interest. In each trial $s = 1, \ldots, S$, the number of events for each criterion was simulated using a Binomial likelihood $x_{ij}^s \sim Bin(N, \theta_{ij})$. Then, values ξ_{ij}^{sk} are sampled from the posterior distribution of the parameters $\mathcal{B}(x_{ij}^s, N - x_{ij}^s)$ for $k = 1, \ldots, K$, assuming implicitly an improper conjugate prior $\mathcal{B}(0, 0)$. The posterior distributions of the utility score and the loss score are approximated by the samples $u(\boldsymbol{\xi}_i^{sk}, \mathbf{w})$ and $l(\boldsymbol{\xi}_i^{sk}, \mathbf{w})$.

Assuming the threshold confidence level $\psi = 0.8$, MCDA and SLoS are compared using $\mathbb{P}[\mathcal{P}_u^{1,2} > 0.8]$, $\mathbb{P}[\mathcal{P}_l^{1,2} > 0.8]$ and $\phi = \mathbb{P}[\mathcal{P}_l^{1,2} > 0.8] - \mathbb{P}[\mathcal{P}_u^{1,2} > 0.8]$. As a difference between probabilities, ϕ ranges in $(-1, 1)$. A value $-1 \leq \phi < 0$ indicates that SLoS recommends treatment T_1 more often than MCDA and $0 < \phi \leq 1$ that SLoS recommends T_1 less often than MCDA. The two approaches are in agreement when $\phi = 0$. A similar analysis for $\mathbb{P}[\mathcal{P}_l^{2,1} > 0.8]$ and $\mathbb{P}[\mathcal{P}_u^{2,1} > 0.8]$ is presented in the Supplemental Material (Figure S1). Simulations with other choices of ψ led to similar conclusions on the comparison between the two methods and are not presented here.

The analyses were conducted using R, with $S = 2500$ simulated clinical trials and $K = 2000$ simulations to estimate the parameter distributions.

4.2 Results

The results are presented in Figure 4. All nine scenarios for treatment T_1 are presented in rows and numbered 1 to 9. Each graph corresponds to fixed expected probabilities of event for treatment T_1 (●), and each cell corresponds to a combination of expected probabilities of benefit and risk for T_2. The probabilities $\mathbb{P}[\mathcal{P}_l^{1,2} > 0.8]$ are presented on the left panel, $\mathbb{P}[\mathcal{P}_u^{1,2} > 0.8]$ on the middle panel, and ϕ on the right panel, for which positive values are displayed in blue, negative values in red, and null values in white.

In scenario 1, the two measures are in agreement to recommend T_1, which has moderate benefit and risk ($\theta_{11} = \theta_{12} = 0.5$), when T_2 has less benefit and more risk. On the diagonal, SLoS favours T_1 to more effective treatments but with very high risk (respectively, to safer treatments but with very low benefit). In contrast, MCDA recommends more effective but highly unsafe treatments, or safer but no effective treatments, compared to T_1. For example, when $\theta_{21} = 0.8$ and $\theta_{22} = 0.9$ (large benefit but high risk), SLoS favours T_1 in 100% of the trials while MCDA recommends it in 62% only, resulting in $\phi = 0.38$. Also, when $\theta_{21} = 0.6$ and $\theta_{22} = 0.7$ (increased

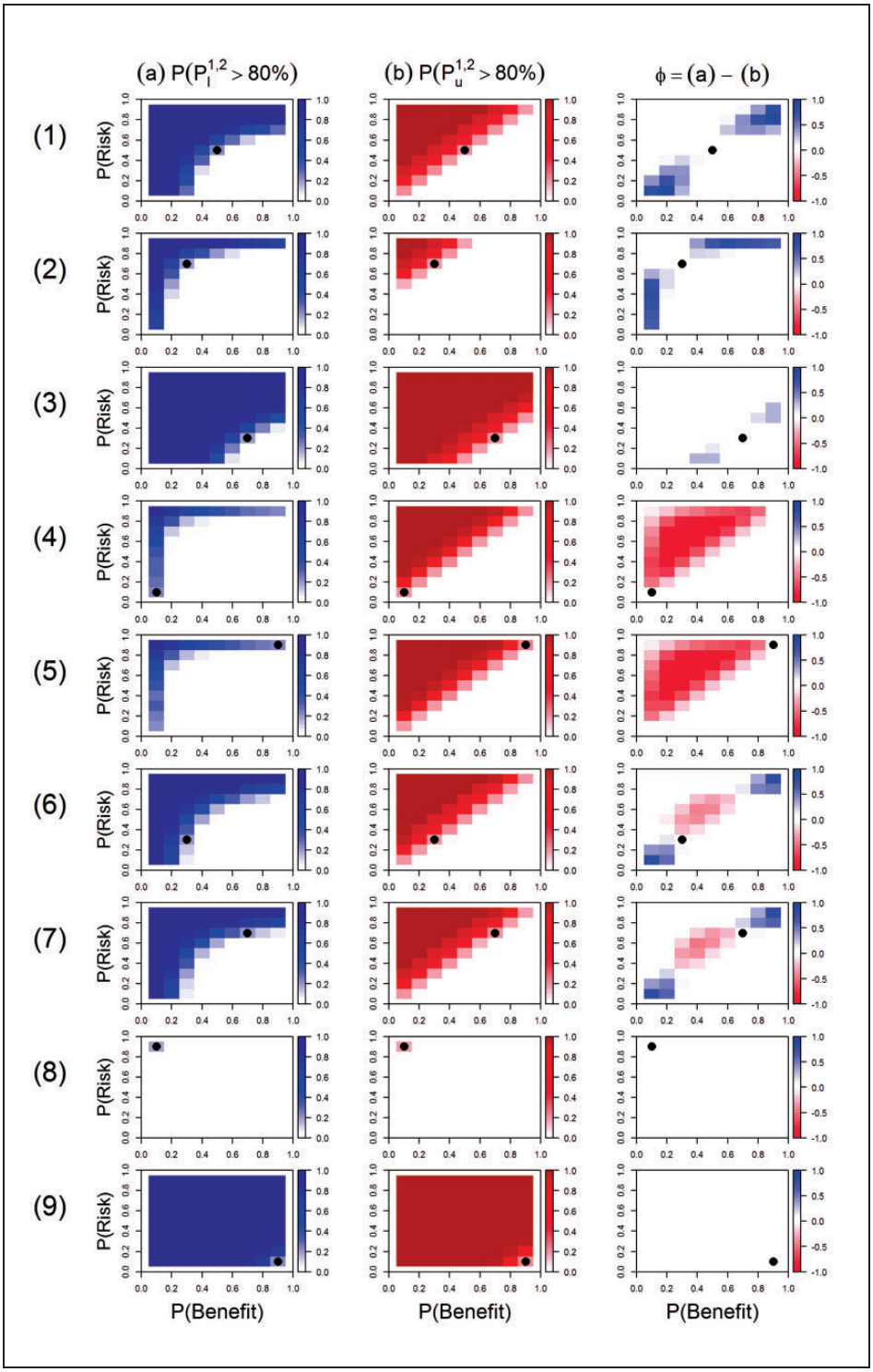

Figure 4. Results of MCDA and SLoS performances in all simulation scenarios for two equally important criteria ($w_j = \tilde{w}_j = 0.5$ for $j = 1,2$). • = T1. Left panel: $P[P^{1,2}\ l > 0.8]$. Middle panel: $P[P^{1,2}\ u > 0.8]$. Right panel: $\Phi = P[P^{1,2}\ l > 0.8] - P[P^{1,2}\ u > 0.8]$, for which blue cells (resp., red cells) indicate that SLoS recommends T1 more often (resp., less often) than MCDA.

benefit by 0.1 and risk by 0.2 compared to T_1), SLoS favours T_1 in 80% and MCDA in 57% of the cases ($\phi = 0.23$). This reflects the property of SLoS that increases in risk are less tolerated when the amount of benefit is large enough. Similar patterns are observed in scenarios 2 and 3 where treatment T_1 has either a low benefit and a large risk, or a large benefit and a low risk, but not extreme probabilities of event.

In scenario 4, T_1 has almost no benefit nor risk, with $\theta_{11} = \theta_{12} = 0.1$. As expected, it is almost never recommended by SLoS, but it could be recommended by MCDA in scenarios where the alternative T_2 has some benefit but a higher increase in risk. For example, when $\theta_{21} = 0.2$ and $\theta_{22} = 0.3$ (increased benefit by 0.1 and risk by 0.2 compared to T_1), MCDA recommends T_1 in 70% of the cases while it is never recommended by SLoS ($\phi = -0.70$). This is consistent with the stated desirable property that we are not interested in the level of risk if the drug does not treat the patients. On the other hand, when $\theta_{21} = 0.3$ and $\theta_{22} = 0.2$ (increased benefit by 0.2 and risk by 0.1 compared to T_1), SLoS discriminates better the treatments and recommends T_2 in 100% of the cases while MCDA recommends it in only 68% (Supplemental Material, Figure S1). Similar conclusions are obtained in scenario 5, where T_1 has both extreme efficacy and risk ($\theta_{11} = \theta_{12} = 0.9$): SLoS never recommends the unsafe treatment T_1 if alternative treatments T_2 have lower risk and at least some small benefit, while MCDA recommends T_1 as compared to treatments with a larger decrease in benefit than in risk. This is the case for instance when $\theta_{21} = 0.6$ and $\theta_{22} = 0.7$ (decreased benefit by 0.3 and risk by 0.2 compared to T_1), where T_1 is recommended in 65% of the cases by MCDA and never recommended by SLoS ($\phi = -0.65$). In contrast, when $\theta_{21} = 0.7$ and $\theta_{22} = 0.6$ (decreased benefit by 0.2 and risk by 0.3 compared to T_1), SLoS favours T_2 in 100% and MCDA in 67% of the cases (Supplemental Material, Figure S1).

Scenarios 6 and 7 correspond to treatment T_1 with either both low benefit and risk ($\theta_{11} = \theta_{12} = 0.3$) or large benefit and risk ($\theta_{11} = \theta_{12} = 0.7$) but where the probabilities of event are not extreme. The measures are in agreement to recommend T_1 when T_2 is indisputably worse. On the diagonal, T_1 is more often recommended by SLoS when T_2 has no benefit nor risk ($\theta_{21} = \theta_{22} = 0.1$) or very large benefit and risk ($\theta_{21} = \theta_{22} = 0.9$). On the other hand, SLoS favours more treatments with benefit and risk probabilities closer to 50%. For example, in scenario 6, when $\theta_{21} = 0.4$ and $\theta_{22} = 0.5$ (increased benefit by 0.1 and risk by 0.2 compared to T_1), SLoS recommends T_1 in only 17% of the cases, but MCDA in 59% ($\phi = -0.42$). Similarly, when $\theta_{21} = 0.5$ and $\theta_{22} = 0.4$ (increased benefit by 0.2 and risk by 0.1 compared to T_1), T_1 is not favoured by any of the methods, but SLoS recommends the alternative T_2 in 88% of the cases and MCDA in only 59% (Supplemental Material, Figure S1). Similar results are observed in scenario 7 for the same examples.

In all scenarios, both methods are in agreement to recommend T_1 when it is indisputably better than T_2, i.e. more effective and safer (or to recommend T_2 when T_1 is indisputably worse, i.e. less effective and more toxic, see Supplemental Material Figure S1). This is well illustrated in scenarios 8 ($\theta_{11} = 0.1$ and $\theta_{12} = 0.9$) and 9 ($\theta_{11} = 0.9$ and $\theta_{12} = 0.1$). In scenario 8, MCDA discriminates slightly better treatments with no efficacy or high risk between themselves, while SLoS penalises them equally, as they should not be recommended anyway.

Overall, both MCDA and SLoS have good performances to discriminate the benefit–risk balance of the treatments. They provide similar conclusions in many situations, and the cases where they differ highlight the two desirable properties of SLoS. Over all possible scenarios, SLoS recommends safer treatments than MCDA in half of the scenarios, and less safe treatments in the other half.

4.3 Sensitivity analyses

While the case of equally important and uncorrelated criteria is considered above, we investigated the robustness of the results in cases of:

- Equally important criteria $w_j = \tilde{w}_j = 0.5$ for $j = 1$, 2 and strongly correlated criteria: $\rho = 0.8$ (positive correlation) and $\rho = -0.8$ (negative correlation).
- More weight on the risk criterion, using MCDA weights $(w_1, w_2) = (0.25, 0.75)$ and mapped SLoS weights $(\tilde{w}_1, \tilde{w}_2) = (0.30, 0.70)$, no correlation between the criteria.
- More weight on the risk criterion, with $(w_1, w_2) = (\tilde{w}_1, \tilde{w}_2) = (0.25, 0.75)$ (no mapping), no correlation between the criteria. This scenario aims at evaluating the impact of the weight mapping on the results, by comparing its results to those of the previous case.

The results of the sensitivity analyses are given in the Supplemental Material.

Both measures are robust to positive and negative correlations between the outcomes, with very similar results (Supplemental Material, Figures S2–S5). When an MCDA weight of 25% is given to the benefit, both measures penalise more the risk, but analogous differences and similarities as before could be observed between them (Supplemental Material, Figures S6–S7). Since the mapping is not far from an identity transformation, omitting it does not have a major impact on the results (Supplemental Material, Figures S8–S9).

A simulation study was also conducted with four criteria (two benefits $j = 1, 2$ and two risks $j = 3, 4$), for which the investigated scenarios are summarised in the Supplemental Material, Table S1, under the following assumptions:

- Equally important criteria with weights $w_j = \tilde{w}_j = 0.25$ for $j = 1, \ldots, 4$, no correlation between the criteria.
- Equally important criteria with weights $w_j = \tilde{w}_j = 0.25$ for $j = 1, \ldots, 4$, correlated criteria (see correlation matrices in the Supplemental Material).
- More weight on the risk criteria, with MCDA weights $(w_1, w_2, w_3, w_4) = (0.10, 0.10, 0.40, 0.40)$ and mapped SLoS weights $(\tilde{w}_1, \tilde{w}_2, \tilde{w}_3, \tilde{w}_4) = (0.15, 0.15, 0.43, 0.43)$, no correlation between the criteria.
- More weight on the risk criteria, with $(w_1, w_2, w_3, w_4) = (\tilde{w}_1, \tilde{w}_2, \tilde{w}_3, \tilde{w}_4) = (0.10, 0.10, 0.40, 0.40)$ (no mapping), no correlation between the criteria.

Similar conclusions could be drawn when comparing MCDA and SLoS using four criteria, even if the interpretation of the simulation scenarios is somewhat less straightforward as the amount of possible situations (low/moderate/high benefits and risks) increases (Supplemental Material, Figures S10-S19).

Overall, the conclusions are robust to correlations, number of criteria, weighting and weight mapping for both measures.

5 Discussion

In this paper, we propose SLoS as a new tool for drug benefit–risk assessment. It offers the same advantages as MCDA to summarise the benefit–risk balance of the treatments in a single measure, but it has additional desirable properties permitting to avoid recommendations of non-effective or extremely unsafe treatments, and to tolerate larger increases in risk for a given increase in benefit when the amount of benefit is small than when it is high. In contrast, we have shown that the linear form of the MCDA utility score involves implicit assumptions of the decision-makers, such as a constant benefit–risk trade-off for all values of benefit or risk, and might lead to counter-intuitive conclusions. It is worth noting that these additive and linear properties were shown to be inadequate in other application areas of MCDA,[21,22] and its limitations in the health domain have been highlighted as well.[27,28]

The independence of the benefit and risk criteria is usually assumed for the sake of simplicity. Correlations could be taken into account in the analyses; however, our simulation study shows that both measures are robust to correlations between outcomes.

Importantly, SLoS penalises drugs with no efficacy, which is sensible for comparisons between active treatments. Indeed, a 'no treatment/placebo' option, in the absence of placebo effect, will most likely be strongly penalised by SLoS due to its lack of efficacy, although it may be preferable to any active treatment with a small amount of efficacy but that causes more harm overall. Therefore, MCDA's recommendations may be more reliable in such cases and this should be carefully considered before choosing the method and when interpreting the results. However, the area of application of SLoS remains large, as many drug comparisons involve a standard of care, or a placebo with expected effects that are non-negligible.[52]

The MCDA weights of the criteria should be elicited according to the preferences of the decision-makers (regulators, experts, patients, etc.) and methods have been proposed in the literature for this purpose.[11,12,16,42,44,45,47,48] We propose a simple mapping to obtain SLoS weights from MCDA weights, so that the same elicitation process could be followed while preserving the weight interpretation. It should be noted that the mapping is not far from an identity transformation, and omitting it does not strongly affect the results. We considered in this paper fixed weights, but extended models have been proposed where the weights are treated as random variables to account for an uncertainty in their assignments.[23,26]

As an aggregation method involving multiple criteria, SLoS could be included within the family of non-linear MCDA models. It was shown that SLoS has the desirable properties even under the linear partial value functions on which this work has focused only. An alternative approach between linear MCDA and SLoS could be to handle the decreasing level of risk tolerance relative to benefits by varying the shape of the partial value functions. For instance, one can derive linearly-weighted partial value functions used in the linear utility score which exhibit the same degree of decreasing risk tolerance as SLoS. This, however, seems to be non-trivial and requires extensive attention. Furthermore, as stated above, the explicit elicitation of non-linear forms for partial value functions may be difficult for project teams. The weight elicitation and their interpretation appear also more challenging,

in particular if the shapes of the partial value functions are different from one criterion to another. Meanwhile, an exploration of the use of non-linear partial value functions both in the framework of the additive utility score and SLoS is of great practical interest and is to be investigated.

In many cases, SLoS and MCDA provide similar conclusions, but SLoS shows clear advantages when treatments have no benefit or extreme risk. In general, this situation may occur in early stage drug developments, or at least before the time of marketing authorisation application, since treatments with no evidence of efficacy or high toxicities usually do not reach this point and are stopped before. Until now, benefit–risk assessments were mainly conducted in late stage by the sponsor and/or regulatory agencies, but it is recommended to initiate the benefit–risk assessment earlier in order to better support internal decisions and discussions with health authorities about the development strategy.[41] Therefore, SLoS could be used in early development, and then updated during the following phases and the regulatory process until post-marketing surveillance, in order to ensure a transparent and consistent benefit–risk assessment throughout the drug life-cycle.

Acknowledgements

The authors would like to thank both reviewers and the editor for insightful comments and suggestions that helped to improve the original work significantly. Veronique Robert is an employee of Institut de Recherches Internationales Servier (IRIS). The views expressed in this publication are those of the authors and not necessarily those of the National Health Service, the National Institute for Health Research or the Department of Health.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Gaelle Saint-Hilary's research is supported by the Institut de Recherches Internationales Servier (IRIS). Pavel Mozgunov, Thomas Jaki and Mauro Gasparini have received funding from the European Union Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 633567. Thomas Jaki's contribution arises in part from his Senior Research Fellowship (NIHR-SRF-2015-08-001) supported by the National Institute for Health Research.

Supplemental material

Additional information including the results of the sensitivity analyses and the source code to reproduce the results may be found in the Supplemental Material.

ORCID iD

Gaelle Saint-Hilary http://orcid.org/0000-0003-1643-3348

References

1. Chuang-Stein C, Entsuah R and Pritchett Y. Measures for conducting comparative benefit: risk assessment. *Drug Inform J* 2008; **42**: 223–233.
2. EMA. Guidance document on the content of the<co->rapporteur day 80 critical assessment report, http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2009/10/WC500004856.pdf (2013, accessed 5 July 2018).
3. EMA. ICH guideline E2C (R2) on periodic benefit-risk evaluation report (PBRER), http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2012/12/WC500136402.pdf (2013, accessed 5 July 2018).
4. US FDA. Providing postmarket periodic safety reports in the ICH E2C (R2) format (periodic benefit-risk evaluation report), http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm346564.pdf (2013 accessed 5 July 2018).
5. Hughes DA, Bayoumi AM and Pirmohamed M. Current assessment of risk-benefit by regulators: is it time to introduce decision analyses? *Clin Pharmacol Therapeut* 2007; **82**: 123–127.

6. EMA. Benefit-risk methodology project. Work package 1 Report: description of the current practice of benefit-risk assessment for centralised procedure products in the EU regulatory network, http://www.ema.europa.eu/docs/en_GB/document_library/Report/2011/07/WC500109478.pdf (2011, accessed 5 July 2018).
7. Mt-Isa S, Peters R, Phillips LD, et al. Review of visualisation methods for the representation of benefit-risk assessment of medication: Stage 1 of 2, www.imi-protect.eu/documents/ShahruletalReviewofvisualisationmethodsfortherepresentationofBRassessmentofmedicationStage1F.pdf (2013, accessed 5 July 2018).
8. Mt-Isa S, Hallgreen CE, Asiimwe A, et al. Review of visualisation methods for the representation of benefit-risk assessment of medication: Stage 2 of 2, http://www.imi-protect.eu/documents/ShahruletalReviewofvisualisationmethodsfortherepresentationofBRassessmentofmedicationStage2A.pdf (2013, accessed 5 July 2018).
9. Leong J, Salek S and Walker S. *Benefit-risk assessment of medicines*. Switzerland: Springer International Publishing, 2015.
10. Mt-Isa S, Ouwens M, Robert V, et al. Structured benefit-risk assessment: a review of key publications and initiatives on frameworks and methodologies. *Pharmaceut Stat* 2015; **15**: 324–332.
11. Thokaka P, Devlin N, Marsh K, et al. Multiple criteria decision analysis for health care decision making - an introduction: Report 1 of the ISPOR MCDA emerging good practices task force. *Value Health* 2016; **19**: 1–13.
12. Marsh K, IJzerman M, Thokala P, et al. Multiple criteria decision analysis for health care decision making – emerging good practices: Report 2 of the ISPOR MCDA Emerging Good Practices Task Force. *Value Health* 2016; **19**: 125–137.
13. EMA. Benefit-risk methodology project. Work package 2 Report: applicability of current tools and processes for regulatory benefit-risk assessment, http://www.ema.europa.eu/docs/en_GB/document_library/Report/2010/10/WC500097750.pdf (2010, accessed 5 July 2018).
14. EMA. Benefit-risk methodology project. Work package 3 report: Field tests, http://www.ema.europa.eu/docs/en_GB/document_library/Report/2011/09/WC500112088.pdf (2011, accessed 5 July 2018).
15. EMA. Benefit-risk methodology project. Work package 4 Report: benefit-risk tools and processes, http://www.ema.europa.eu/docs/en_GB/document_library/Report/2012/03/WC500123819.pdf (2012, accessed 5 July 2018).
16. Mussen F, Salek S and Walker S. A quantitative approach to benefit-risk assessment of medicines – Part 1: the development of a new model using multi-criteria decision analysis. *Pharmacoepidemiol Drug Safety* 2007; **16**: S2–S15.
17. Mussen F, Salek S and Walker S. Development and application of a benefit-risk assessment model based on multi-criteria decision analysis. In: Mussen F, Salek S and Walker S (eds.) *Benefit-risk appraisal of medicines: a systematic approach to decision-making*.. Chichester, UK: John Wiley & Sons Ltd, 2008, pp. 111–119.
18. NICE Decision Support Unit. Multi-criteria decision analysis (MCDA), http://scharr.dept.shef.ac.uk/nicedsu/wp-content/uploads/sites/7/2016/03/MCDA-for-HTA-DSU.pdf (2011, accessed 5 July 2018).
19. Marsh K, Lanitis T, Neasham D, et al. Assessing the value of healthcare interventions using multi-criteria decision analysis: a review of the literature. *PharmacoEconomics* 2014; **32**: 345–365.
20. IMI PROTECT. Work package 5: Benefit-risk integration and representation, http://www.imi-protect.eu/wp5.shtml (2013, accessed 5 July 2018).
21. Langhans SD and Lienert J. Four common simplifications of multi-criteria decision analysis do not hold for river rehabilitation. *PLOS One* 2016; **11**: 1–27.
22. Martin DM and Mazzotta M. Non-monetary valuation using multi-criteria decision analysis: sensitivity of additive aggregation methods to scaling and compensation assumptions. *Ecosystem Services* 2018; **29**: 13–22.
23. Tervonen T, Van Valkenhoef G, Buskens E, et al. A stochastic multicriteria model for evidence-based decision making in drug benefit-risk analysis. *Stat Med* 2007; **30**: 1419–1428.
24. Nixon R, Dierig C, Mt-Isa S, Stöckert I, et al. A case study using the PrOACT-URL and BRAT frameworks for structured benefit risk assessment. *Biometric J* 2016; **58**: 8–27.
25. Marcelon L, Verstraeten T, Dominiak-Felden G, et al. Quantitative benefit-risk assessment by MCDA of the quadrivalent HPV vaccine for preventing anal cancer in males. *Expert Rev Vaccines* 2016; **15**: 139–148.
26. Saint-Hilary G, Cadour S, Robert V, et al. A simple way to unify multicriteria decision analysis (MCDA) and stochastic multicriteria acceptability analysis (SMAA) using a dirichlet distribution in benefit-risk assessment. *Biometric J* 2017; **59**: 567–578.
27. Morton A. Treacle and smallpox: two tests for multicriteria decision analysis models in health technology assessment. *Value Health* 2017; **20**: 512–515.
28. Marsh KD, Sculpher M, Caro J, et al. The use of MCDA in HTA: great potential, but more effort needed. *Value Health* 2017; **21**: 394–397.
29. Keeney RL and von Winterfeldt D. Practical value models. *Published Articles Papers* 2009; 36, https://research.create.usc.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=1054&context=published_papers (accessed 5 July 2018).
30. Norstrom JG. The use of precautionary loss functions in risk analysis. *IEEE Transactions on reliability* 1996; **45**: 400–403.
31. Mozgunov P, Jaki T and Gasparini M. *Loss functions in restricted parameter spaces and their Bayesian applications*. Preprint arXiv:1706.02104 (2017). Department of Mathematics and Statistics, Lancaster University, Lancaster.
32. EMA. EMEA/H/C/000354-N/0068 – Ketek: EPAR, http://www.ema.europa.eu/ema/index.jsp?curl=pages/medicines/human/medicines/000354/human_med_000873.jsp&mid=WC0b01ac058001d124 (2017, accessed 5 July 2018).

33. US FDA. Approval date(s) and history, letters, labels, reviews for NDA 021144 – Ketek, https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm?event=overview.process&ApplNo=021144 (2017, accessed 5 July 2018).
34. IMI PROTECT. Work package 5: Report 1:b:ii Benefit Risk Wave 1 Case Study Report: Ketek® (telithromycin), http://www.imi-protect.eu/documents/QuarteyetalBenefitRiskWave1CasestudyReportTelithromycinFeb2012.pdf (2012, accessed 5 July 2018).
35. Berger JO. *Statistical decision theory and Bayesian Analysis*, 2nd ed. New York, NY: Springer-Verlag, 1985.
36. Varian HR. *Intermediate microeconomics – a modern approach*, 8th ed. New York, NY: WW Norton & Co, 2010.
37. Kelbert M and Mozgunov P. Generalization of Cramér-Rao and Bhattacharyya inequalities for the weighted covariance matrix. *Math Commun* 2017; **22**: 25–40.
38. Mozgunov P and Jaki T. *An information-theoretic approach for selecting arms in clinical trials*. Preprint arXiv:1708.02426 (2017). Department of Mathematics and Statistics, Lancaster University, Lancaster.
39. Guo JJ, Pandey S, Doyle J, et al. 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; **13**: 657–666.
40. Riabacke M, Danielson M and Ekenberg L. State-of-the-art prescriptive criteria weight elicitation. *Adv Decision Sci* 2012; 1–24. DOI:10.1155/2012/276584.
41. Sashegyi A, Felli J and Noel R. *Benefit-risk assessment in pharmaceutical research and development*. USA: Chapman and Hall/CRC, 2013.
42. Broekhuizen H, Groothuis-Oudshoorn CGM, Hauber AB, et al. Estimating the value of medical treatments to patients using probabilistic multi criteria decision analysis. *BMC Med Inform Decis Making* 2015; **15**: 102.
43. Hughes D, Waddingham E, Mt-Isa S, et al. Recommendations for benefit-risk assessment methodologies and visual representations. *Pharmacoepidemiol Drug Safe* 2016; **25**: 251–262.
44. Tervonen T, Gelhorn H, Sri Bhashyam S, et al. MCDA swing weighting and discrete choice experiments for elicitation of patient benefit-risk preferences: a critical assessment. *Pharmacoepidemiol Drug Safety* 2017; **26**: 1483–1491. DOI:10.1002/pds.4255.
45. Broekhuizen H, IJzerman MJ, Hauber AB, et al. Weighing clinical evidence using patient preferences: An application of probabilistic multi-criteria decision analysis. *PharmacoEconomics* 2017; **35**: 259–269.
46. Waddingham E, Mt-Isa S, Nixon R, et al. A Bayesian approach to probabilistic sensitivity analysis in structured benefit-risk assessment. *Biometric J* 2016; **58**: 28–42.
47. Johnson FR, Lancsar E, Marshall D, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. *Value Health* 2013; **16**: 3–13.
48. Hauber AB, Gonzalez JM, Groothuis-Oudshoorn CGM, et al. Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR Conjoint Analysis Good Research Practices Task Force. *Value Health* 2016; **19**: 300–315.
49. Mateu-Figueras G, Vera Pawlowsky-Glahn and Egozcue JJ. The normal distribution in some constrained sample spaces. *SORT: Statist Operat Res Transact* 2013; **37**: 29–56.
50. R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2015, http://www.R-project.org/ (accessed 5 July 2018).
51. EMA. EMEA/H/C/354/A22/41 – Ketek: EPAR – Scientific discussion – variation; http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Scientific_Discussion_-_Variation/human/000354/WC500100697.pdf (2007, accessed 5 July 2018).
52. Gupta U and Verma M. Placebo in clinical trials. *Perspect Clin Res* 2013; **4**: 49–52.